Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey

Weiting Zhang, Dong Yang, *Member, IEEE*, and Hongchao Wang, *Member, IEEE*

***Abstract*—With the tremendous revival of artificial intelligence, predictive maintenance (PdM) based on data-driven methods has become the most effective solution to address smart manufacturing and industrial big data, especially for performing health perception (e.g., fault diagnosis and remaining life assessment). Moreover, because the existing PdM research is still in primary experimental stage, most works are conducted utilizing several open-datasets, and the combination with specific applications such as rotating machinery is especially rare. Hence, in this paper, we focus on data-driven methods for PdM, present a comprehensive survey on its applications, and attempt to provide graduate students, companies, and institutions with the preliminary understanding of the existing works recently published. Specifically, we first briefly introduce the PdM approach, illustrate our PdM scheme for *automatic washing equipment*, and demonstrate the challenges encountered when we conduct a PdM research. Second, we classify the specific industrial applications based on six algorithms of machine learning and deep learning (DL), and compare five performance metrics for each classification. Furthermore, the accuracy (a metric to evaluate the algorithm performance) of these PdM applications is analyzed in detail. There are some important conclusions: 1) the data used in the summarized literature are mostly from public datasets, such as case western reserve university (CWRU)/intelligent maintenance systems (IMS); and 2) in recent years, researchers seem to focus more on DL algorithms for PdM research. Finally, we summarize the common features regarding our surveyed PdM applications and discuss several potential directions.**

***Index Terms*—Artificial intelligence (AI), deep learning (DL), fault diagnosis, machine learning (ML), predictive maintenance (PdM), remaining life assessment.**

I. Introduction

WITH the requirement of economic considerations and the advancement of technology, many countries, including the United States, Germany, and China, have successively put forward plans to revitalize manufacturing [1]. Specifically, the United States deployed the "smart manufacturing" strategy, which is implemented by introducing the cyber physical system (CPS, a system that links the cyber world of computing and communication with the physical world) to the manufacturing process [2]. Then, Germany presented "Industry 4.0" [3] at the Hannover Messe, in April 2013, and the core idea of its initiative is to build smart factories, smart production, and smart logistics. Moreover, the government of China proposed the "Made in China 2025" strategy [4], which is particularly a significant strategic initiative in the new round of technological revolution. Its core meaning is to realize the transformation and upgrading of manufacturing, that is, to achieve intelligence.

With the arrival of "Industry 4.0," the prognostic and health management (PHM) concept has become the inevitable tendency in the context of smart manufacturing and industrial big data [5], and it provides a reliable solution for managing the health status of equipment. In addition, the PdM strategy for industrial equipment can accurately perceive performance degradation since it was designed to achieve near-zero failures, near-zero hidden dangers, near-zero accidents, and near-zero pollution throughout the entire manufacturing process [6]. In particular, the prediction of machine health can, not only significantly reduce the unexploited downtime and expensive labor costs but also ensure safe operation and optimize the maintenance plan.

According to the review of current research results [6], the PdM methods are mainly divided into the following three categories: 1) model-based prognosis; 2) knowledge-based prognosis; and 3) data-driven prognosis. Notably, the data-driven PdM technology has attracted wide attention. Specifically, in [7], the PdM methods of mechanical equipment were comprehensively studied and sorted out from the aspect of data acquisition, data processing, and decision support. A data-driven intelligent PdM system was proposed in [8] for reaching zero-defect manufacturing.

In recent years, industrial wireless sensor networks (IWSNs) [9], [10] and industrial cyber physical systems [11] have become an emerging data acquisition technology in the complex industrial environment, and mechanical data can be collected using various types of sensors with high reliability and in a real-time manner [12]. Obviously, because of the continuous improvement of data acquisition ability [13], as well as the exponential growth of data volume [14], data-driven methods for health monitoring have achieved great success and received widespread attention regarding the PdM of industrial equipment. Therefore, since artificial intelligence (AI) algorithms have achieved important progress during the past five years, we considered the relevance of performing a summary study on guiding the selection of algorithms in specific applications, particularly the corresponding significance of AI to PdM.

Manuscript received September 17, 2018; revised January 28, 2019; accepted March 11, 2019. Date of publication May 6, 2019; date of current version August 23, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1702001, in part by the National Natural Science Foundation of China under Grant 61771040, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018YJS013. *(Corresponding author: Dong Yang.)*

The authors are with the Beijing Jiaotong University, Beijing 100044, China (e-mail: 17111018@bjtu.edu.cn; dyang@bjtu.edu.cn; hcwang@bjtu.edu.cn).

Digital Object Identifier 10.1109/JSYST.2019.2905565

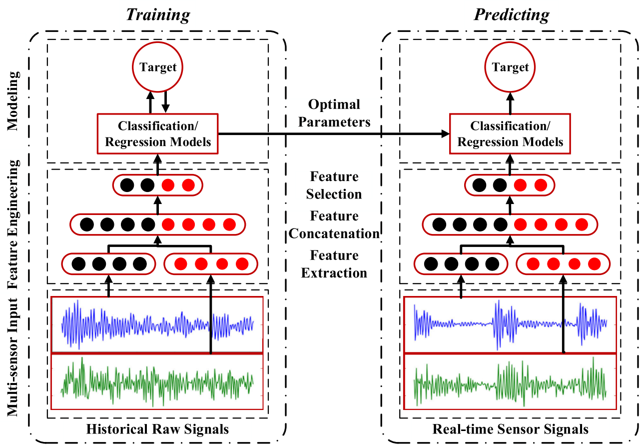

Fig 1. Flowchart of the data-driven method for predictive maintenance (PdM).

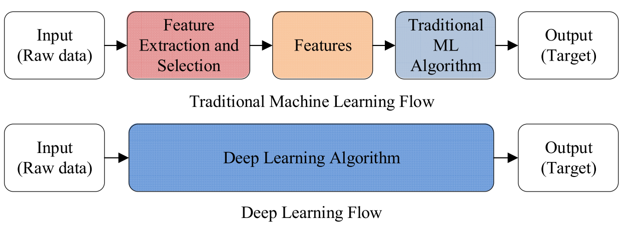

Fig. 2. Flow of ML and DL.

The increased availability of industrial data has paved the way for the development and deployment of the data-driven PdM, which utilizes cutting-edge computational methods to provide valuable information regarding the status of equipment acquired from the growing volume of operational data [15]. As shown in Fig. 1, the data-driven PdM system consists of two phases: first, a learning process (i.e., model training) is needed on the basis of historical raw sensor signals; second, the trained model is applied to predict targets and make decisions. In general, each phase consists of the following three subprocesses:

1) data acquisition and preprocessing, which can be single-sensory or multisensory;
2) feature engineering, which contains feature extraction, concatenation, and selection; and
3) model training and predicting, in which a well-trained model will be generated with the optimal parameters.

After that, the model can predict the real-time data flow. Moreover, the data-driven PdM has been extensively applied to industrial manufacturing using machine learning (ML) [16] and deep learning (DL) algorithms [17]. Specifically, as shown in Fig. 2, traditional ML algorithms, (e.g., logistic regression (LR), support vector machine (SVM), decision tree (DT), and random forest (RF)) generally require collecting large amount of data from the health conditions and various failure status scenarios for model training. Next, feature engineering is conducted from the time, frequency, and time-frequency domain [18], [19], and the representation learning of the device health is performed using the extracted features. However, DL (i.e., various neural network (NN) models) avoids the above complex feature engineering and can be learned using an end-to-end learning manner, which is implemented by adding deep layers between the raw data and the prediction result. Thus, the deep models can be deemed as a "black box," which output the prediction result from the input directly, and it is the essential difference between ML and DL. For all of these considerations, both ML and DL have been, especially, widely used in the application of PdM.

Accordingly, we concentrate on data-driven methods in the PdM and present a related comprehensive survey on their applications from the aspects of traditional ML and DL. Our objective is principally attempting to provide newly admitted graduate students and newly established companies and institutions with a preliminary understanding of the data-driven PdM, as well as the research status, methods, and effect of existing works published in the past five years. We note that there are several studies in the literature [19]–[21] related to this paper. However, in contrast to the methods we are concerned about, fault diagnosis techniques were reviewed in [19] from signal- and model-based perspectives. Moreover, Zhao *et al.* [20] and Liu *et al.* [21] not only reviewed the specific applications but also summarized the algorithm itself. Specifically, Zhao *et al.* [20] mainly focused on a deep NN (DNN), while Liu *et al.* [21] paid more attention to traditional ML algorithms. Nevertheless, in this paper, we conduct a comprehensive review on PdM applications from the aspects of ML and DL. In addition, we focus on those works with the highest and lowest accuracy in each algorithm and present our insights about the situations.

The contributions of this paper are as follows. First, a comprehensive review of the PdM is conducted from four aspects, including the definition of the PdM, significance of the data-driven PdM, specific implementation methods of the data-driven PdM, challenges in the implementation process. Second, the PdM applications of the six algorithms are compared from ML and DL perspectives, respectively, to provide graduate students, companies, and institutions with the preliminary understanding of the existing works recently published. Third, on the basis of the above comparison, accuracy, which is the most widely used evaluation metric, is analyzed in detail, and the corresponding conclusions are drawn. In addition, some potential research directions are provided consequently.

The structure of this paper is shown in Fig. 3, and the rest of this paper is organized as follows. Section II presents the background and recent research on the PdM and data-driven PdM, provides a specific case of the data-driven PdM, and summarizes some challenges in practice. Sections III and IV introduce several applications of the data-driven PdM from ML (e.g., LR, SVM, DT and RF) and DL (e.g., artificial NN (ANN), DNN, and auto-encoder (AE)) perspectives, respectively. In specific, Section III provides a review of several mainstream ML algorithms for PdM applications. Then, Section IV reviews the NN and its derived models in PdM applications. Furthermore, Section V provides several summarizations and an analysis of the accuracy. Finally, the future research directions and conclusion are given in Section VI and Section VII.

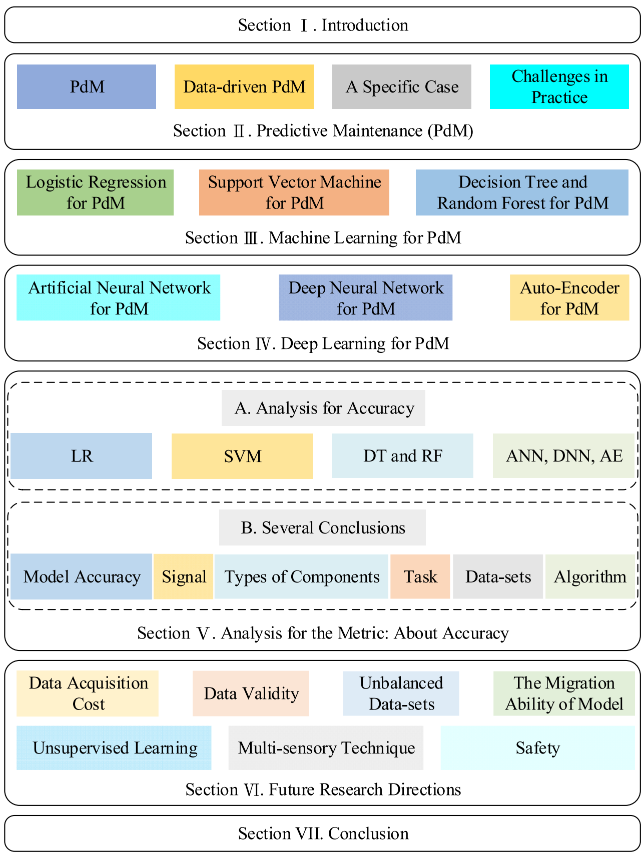

Fig. 3. Condensed structure of this survey.

TABLE I
TERMINOLOGIES AND ABBREVIATIONS

Abbr	Full Name
AI	artificial intelligence
PHM	prognostic and health management
PdM	predictive maintenance
ML	machine learning
DL	deep learning
LR	logistic regression
SVM	support vector machine
DT	decision tree
RF	random forest
RUL	remaining useful life
ANN	artificial neural network
DNN	deep neural network
AE	auto-encoder

Table I lists the terminologies and abbreviations appeared in the following sections, which will be omitted afterward.

II. PREDICTIVE MAINTENANCE

A. Predictive Maintenance

In the late 20th century, PHM was first put forward for the development of military projects [22]. However, in recent years, the PHM system has become a surefire solution for managing the health status of equipment (e.g., fault diagnosis and remaining useful life (RUL) estimation), especially industrial

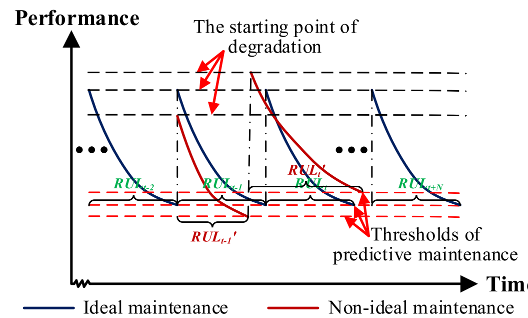

Fig. 4. PdM of degraded equipment [5], [6].

equipment, which is achieved by comprehensively utilizing the latest research results of modern information technology and AI technology [6]. In addition, the PdM is an important part of the PHM system.

According to [23], maintenance methods are mainly divided into three categories, namely, run-to-failure, preventive maintenance, and PdM. Regarding these methods, the first one is the simplest. That is, the new component runs from installation until failure occurs and then stops the machine for maintenance. However, because of the extra maintenance expenditures and unexpected downtime, the cost and efficiency of this method are also the most unreasonable. The second method refers to the failures that can be prevented, achieved by regular replacement of components, but leading to additional operating costs and an increased unexploited lifetime [24]. In addition, the third method is generally indicated based on the assessment of the health status of key components, regardless of the maintenance status, and its fundamental purpose is to make predictability achievable. That is, the incipient problems that may evolve into catastrophic failures can be predicted accurately, and effective measures can be applied to avoid these failures on the basis of the prediction results. This approach can not only minimize maintenance costs but can also extend the useful life of the equipment [25].

It is assumed that the industrial equipment can be repaired at an appropriate time (i.e., before the fault occurs), as this approach will restore the apparatus to its original condition after each maintenance is completed. As shown in Fig. 4, the curve in the figure demonstrates the performance trend of a monitoring object, such as a component, device, or system. The blue curve indicates an ideal maintenance where the repaired RUL (green) is basically identical. However, the reddish brown curve indicates a nonideal maintenance, either too early or too late, and its RUL (red) is different from that of the former. The dotted lines in black and red represent the performance threshold of the object in each operating cycle. In addition, it reflects that the PdM of industrial equipment is moving toward an extremely precise, efficient, and intelligent direction, even with slight performance degradation or safety risks, which can result in serious consequences. The health status of a component, machine, or system can be obtained at any time, and the failure can be predicted and prevented to achieve near-zero downtime performance [26]. In summary, the PdM focuses on utilizing

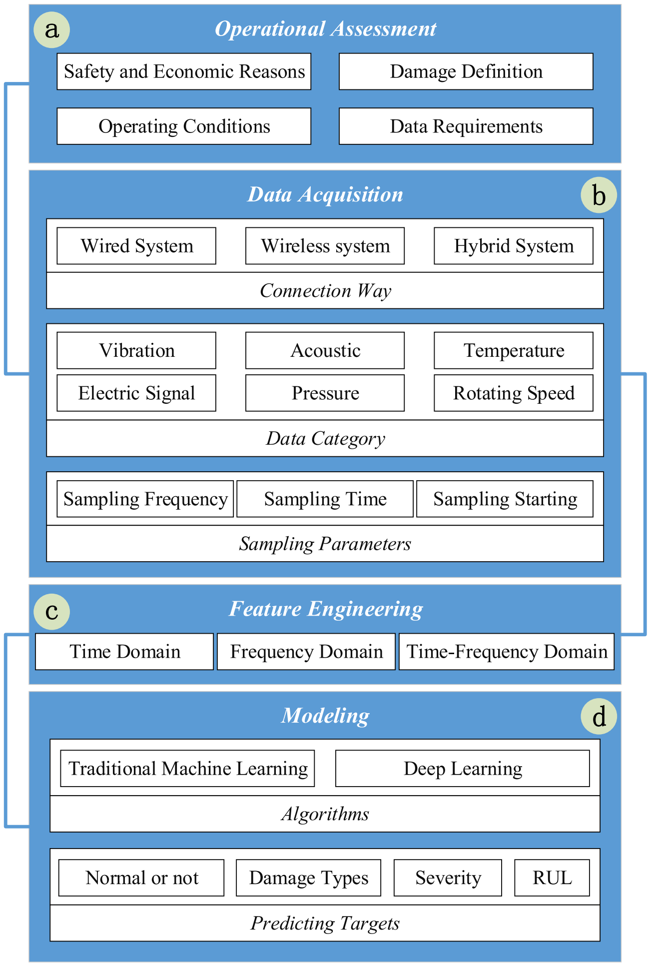

Fig. 5. Implementation process of the data-driven PdM.

predictive information to properly schedule future maintenance operations [6]. Accordingly, deploying the PdM to ensure the safety of operations is of great significance [27].

B. Data-Driven PdM

At present, the proportion of maintenance expenditure in the operating costs of enterprises is increasing, which is caused by inefficient maintenance methods (insufficient maintenance or excessive maintenance). As a result, it is necessary to formulate a reasonable maintenance strategy that not only reduces the waste caused by excessive maintenance but also ensures that the equipment is adequately maintained and remains in good working condition [6]. A new MATLAB toolbox DB-KIT was proposed in [28], which contains a series of basic and advanced algorithms and auxiliary tools for the data-driven PdM. Obviously, with the rise of AI once again, data-driven methods for fault diagnosis and RUL prediction have been the popular issues in PdM system research.

As shown in Fig. 5, the implementation process of the data-driven PdM is described in detail, and it is basically consistent with the design methodology of [29]. Overall, the process can be divided into four stages: 1) operational assessment; 2) data acquisition; 3) feature engineering; and 4) modeling. It is important to note that feature engineering can be saved when employing DL algorithms. In addition, limited by computing power, storage capacity, and time delay, the training and deployment of the model need to be completed in different platforms [30]. Specifically, model training is usually carried out on the cloud or server side because it requires strong computing power and storage space as support [31]. However, the trained-down model does not require much computing power and storage space to predict the real-time dataflow. Accordingly, to maximize the utilization of transmission and computing resources, the model will be deployed on the edge nodes, such as the edge server or edge gateway.

In summary, a considerable number of experimental studies have proven that it not only results in better effects but also simplifies the construction process of the data-driven PdM system. Therefore, this paper will review the specific PdM application from the perspectives of ML and DL.

C. Specific Case Based on Data-driven PdM

As depicted in the previous sections, to overcome the limitations of model-based methods [32], data-driven methods have been increasingly applied to the PdM [33]. As shown in Fig. 6, we divide the PdM system into five parts: 1) a physical layer; 2) perception layer; 3) signal analysis layer; 4) performance prediction layer; and 5) decision-making layer, which are specifically utilized to perform the PdM of *automatic washing equipment* (a machine used for the cleaning stage before the maintenance of a high-speed rail). It can be seen that such a system can properly implement a production and maintenance strategy, and it will lead to a longer running time and the sufficient use for key components [34], such as bearings [35], blades [36], [37], and driving motors [38], [39]. In addition, the effective degradation assessment for components is beneficial to reduce the production downtime and maintenance cost [40].

However, because of different conditions such as lubrication, load, and speed [41], the failure modes of the components are also different. Therefore, in actual applications, the fault type and degree need to be evaluated using signal analysis to determine the best time for the implementation of the maintenance strategy [42]–[44]. According to [45]–[49], in our research, four types of sensors (i.e., vibration, temperature, electrical signal, and rotating speed) are taken as input signals to represent the equipment performance. As a complementary approach, Jin *et al.* [45] and Jie *et al.* [46] considered that vibration monitoring can provide the best means for failure identification, and both studies employed vibration signal to perceive the equipment status. However, Mobley [47] deemed that signal analysis of vibration cannot provide all the necessary information for the fault diagnosis of mechanical equipment. Consequently, electric signals (e.g., current and voltage) in [48] and [49] were applied to identify the defects in equipment operation.

To construct the PdM system of *automatic washing equipment*, a set of data acquisition system is developed, as shown in the submodule of Fig. 6, including three subgraphs. The left figure demonstrates a data acquisition device, which is utilized

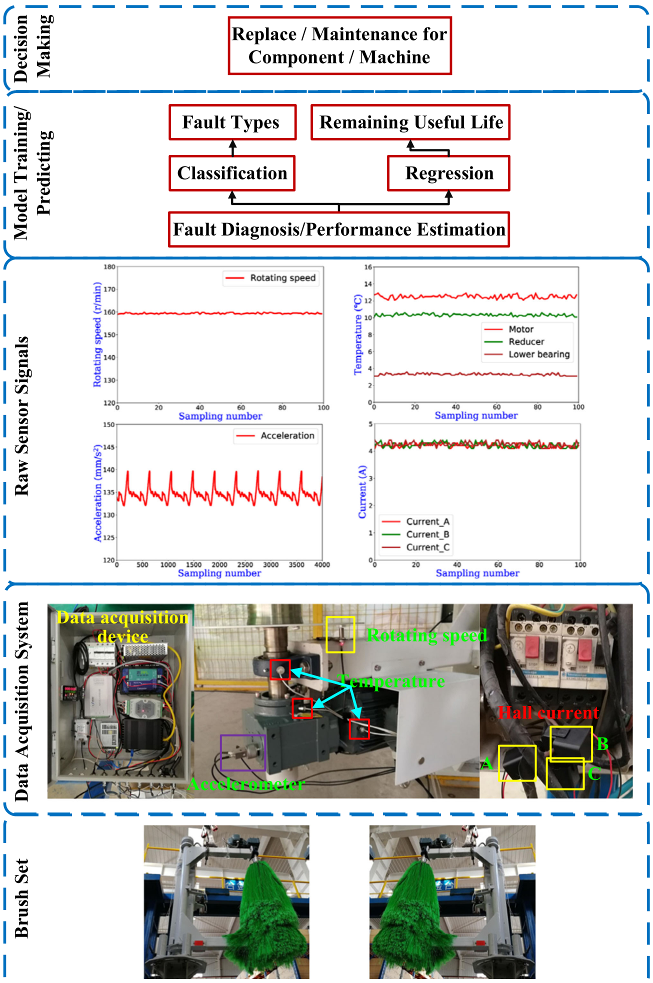

Fig. 6. Framework of the PdM. When using DL, the raw signals can be input directly into the model; when using ML, feature extraction is required.

to gather, store, and forward the collected data. The other two graphs contain four type sensors, which measure acceleration, temperature, current, and rotational speed, respectively. It can be seen that data acquisition plays a crucial role in the realization process of the PdM.

D. Challenges in Practice

As mentioned above, the advantages of the PdM are self-evident. In recent years, numerous mature methods have been formed, such as the methods in [50] and [51]. However, as far as the current literature is concerned, a large part of the literature on the data-driven PdM are based on some publicly available datasets. In practice, it needs a comprehensive analysis from the perspectives of operational assessment, data acquisition, feature engineering, and model building. As a result, there are still many challenges to overcome to reach a mature application level.

On the basis of the design experience, we summarize and present these challenges in Table II.

Operational assessment is extremely critical, and it determines the success or failure of the PdM in a sense. In

TABLE II
CHALLENGES IN THE CONSTRUCTION OF THE PDM SYSTEM

Stages	***Challenge Descriptions***
Operational Assessment	● **time cost** (e.g., system development, debugging, and deployment) ● **economic cost** (e.g., research and hardware expenditure, as well as developer cost) ● **safety** (e.g., equipment operation safety and personnel safety) ● **restriction** (e.g., environmental restrictions on data collection)
Data Acquisition	● **development mode** (e.g., integrated development or independent development) ● **budget** (e.g., design cycle) ● **conversion ability** (i.e., the ability to transform sensor data into health status information) ● **system performance** (e.g., useful life and maintenance cost)
Feature Engineering	● require **expert knowledge** and **prior experience** ● **signal analyzing** (e.g., time-, frequency-, and time frequency domain) ● feature design, extraction, selection
Modeling	● **model selection** (based on model complexity) ● **model training** (whether the data is labeled) ● **model predicting** (time efficiency and accuracy requir -ement)

general, issues such as time and economic costs need to be considered first in the implementation process. The balance between the cost of introducing a PdM system and the loss of equipment failure should be a better tradeoff. More importantly, considering safety issues, the specific implementation will be subject to various restrictions, such as the location of sensor installation.

Most notably, data-driven methods make data acquisition an indispensable part for the complete PdM design [52], providing valuable information for specific tasks. Specifically, for academic research, many groups choose to purchase existing equipment to collect data. However, for engineering applications, a personalized data acquisition system needs to be established to meet specific requirements. It has been proven that data analysis at different stages of the machine life cycle can not only process data or information more effectively but also achieve transparency of the health status of industrial equipment.

In addition, how to address the obtained data effectively (i.e., whether to use feature engineering or not) has become a major challenge for algorithm selection. Here, we make a demonstration of common statistical features in Table III. The performance difference of the algorithms can be reflected in specific applications, especially when faced with industrial sensor data. Accordingly, the PdM applications are summarized from the perspectives of ML and DL in the following.

Moreover, model building is also not to be underestimated, and the selection, training, and optimization of the model all deserve more attention. In practice, supervised learning has achieved remarkable results in many fields, but it requires data to be correctly annotated, which is also an enormous challenge for the PdM. With the quantitative tendency of industrial big data, unsupervised learning has naturally become a significant promising research direction.

TABLE III
SUMMARY OF STATISTICAL FEATURES IN TIME DOMAIN AND FREQUENCY DOMAIN ACCORDING TO [53]–[55]

No.	Name	Equation	No.	Name	Equation	No.	Name	Equation
1	Maximum	$s_1 = max(x_i)$	8	Mean square	$s_8 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$	15	Variance frequency	$F_2 = \frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})^2$
2	Minimum	$s_2 = min(x_i)$	9	Root mean square	$s_9 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	16	Standard deviation frequency	$F_3 = \sqrt{\sum_{j=1}^{N}\left((f_j - F_1)X_j\right) \Big/ \sum_{j=1}^{N} X_j}$
3	Median	$s_3 = \begin{cases} x_{(N+1)/2}, N\ is\ odd \\ \frac{x_{N/2} + x_{(N+1)/2}}{2}, N\ is\ even \end{cases}$	10	Skewness	$s_{10} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i - s_5)^3}{s_7^3}$	17	Root mean square frequency	$F_4 = \sqrt{\sum_{j=1}^{N}(f_j^2 X_j) \Big/ \sum_{j=1}^{N} X_j}$
4	Peak-to-peak	$s_4 = max\lvert x_i\rvert - min\lvert x_i\rvert$	11	Kurtosis	$s_{11} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i - s_5)^4}{s_7^4}$	18	Spectral skewness	$F_5 = \sum_{j=1}^{N}\left(\frac{f_j - F_1}{F_3}\right)^3 S(f_j)$
5	Mean	$s_5 = \frac{1}{N}\sum_{i=1}^{N} x_i$	12	Skewness factor	$s_{12} = \frac{1}{N}\sum_{i=1}^{N} x_i^3 \Big/ \left(\sqrt{s_8}\right)^3$	19	Spectral kurtosis	$F_6 = \sum_{j=1}^{N}\left(\frac{f_j - F_1}{F_3}\right)^4 S(f_j)$
6	Variance	$s_6 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$	13	Kurtosis factor	$s_{13} = \frac{1}{N}\sum_{i=1}^{N} x_i^4 \Big/ \left(\sqrt{s_8}\right)^4$	20	Spectral power	$F_7 = \sum_{j=1}^{N}(f_j)^3 S(f_j)$
7	Standard deviation	$s_7 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$	14	Mean frequency	$F_1 = \frac{1}{N}\sum_{j=1}^{N} f_j$	21	Wavelet energy	$F_8 = \sum_{j=1}^{N}\omega t_\phi^2(j)/N$

III. ML FOR PDM

ML is well known as the core technology of AI, and algorithms play a particularly important role in its gradual rise. At the same time, with the massively available data used to solve various problems, ML has been widely used in computer science and other fields, such as the PdM of industrial equipment, which is one of the potential application fields for data-driven methods (e.g., LR, SVM, DT, and RF) [56], [57]. As a result, although increasingly high-performance algorithms are continually developed, generally, employing simple and efficient methods is considered first.

In this section, we will mainly summarize the application of these algorithms in the PdM. The comparison and analysis of the reviewed literature are also given in Table IV.

A. LR for PdM

1) LR Model: LR is a well-known classification model in ML with the lowest algorithm complexity. It belongs to supervised learning; therefore, the collected data must have corresponding labels to be fed into the model. In addition, the LR model takes a linear combination of features as its input and applies a nonlinear function to conduct mapping, so that each output will fall within the range of (0, 1) and a probabilistic interpretation can be obtained. Consequently, when we acquire a large number of labeled features and have critical requirements on model complexity, we can consider using the LR model to solve these problems [58].

The prediction function of the LR model is as follows:

$$h_\theta(x) = g_\theta\left(\theta^T x\right) = \frac{1}{1 + e^{-\theta^T x}} \tag{1}$$

where $\theta^T x = \sum_{i=0}^{n}\theta_i x_i = \theta_0 + \theta_1 x_1 + \cdots + \theta_i x_i$, x the input feature of this model, θ the internal parameter, which consists of weights and biases, and $g(\cdot)$ a nonlinear transformation of x, named logistic function.

Then, the loss function of LR is defined as:

$$L(\theta) = \prod_{i=1}^{m}(h_\theta(x_i))^{y_i}(1 - h_\theta(x_i))^{1-y_i} \tag{2}$$

where m refers to the sample number, and x_i and y_i indicate the ith sample and the corresponding label.

For the convenience of calculation, there is the following likelihood function:

$$\begin{aligned} l(\theta) &= \ln L(\theta) = \sum_{i=1}^{m}[y_i \ln(h_\theta(x_i)) \\ &\quad + (1 - y_i)\ln(1 - h_\theta(x_i))]. \end{aligned} \tag{3}$$

The backpropagation and parameter updating can be defined as follows:

$$\frac{\partial y}{\partial \theta_j} J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[h_\theta(x_i) - y_i]\, x_i^j \tag{4}$$

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}[h_\theta(x_i) - y_i]\, x_i^j \tag{5}$$

where x_i^j indicates the jth feature of the ith sample, and α represents the hyperparameter of learning rate.

2) PdM Applications of LR: This section will review several PdM applications related to LR.

Li *et al.* [59] proposed a method that combines an LR model with the acoustic emission and cutting force signals; both of them were utilized to monitor the wearing process of cutting tools and determine the best maintenance time. A simple application was introduced in [60] for elevator door health monitoring by utilizing the LR model. Specifically, wavelet package energies were used as features, Fisher's criteria were used to select critical features, and then the maximum likelihood estimation was applied to determine the parameters of the LR model. Finally, a 100% diagnostic result was obtained. A data-driven modeling method was demonstrated in [15] for the health

TABLE IV
SUMMARY OF THE DATA-DRIVEN PDM METHOD USING THREE TYPES OF ML ALGORITHMS

Algorithm	Reference	Signal	Application scenario	Task	Accuracy	Datasets
Logistic Regression	H. K. Li *et al* [59]	acoustic emission	cutting tools	reliability evaluation	99.97%, 99.96%	Dongyu Machine and Tool CMV-850A Center
	J. H. Yan *et al* [60]	vibration/current	elevator door system	degradation assessment	100%	from the door control board
	J. J. A. Costello *et al* [15]	vibration	gas circulator units	health state estimation and RUL prediction		from EDF Energy
	D. H. Pandya *et al* [61]	vibration	bearings (rotating)	fault degradation estimation	100 %	from the test rig
	W. Caesarendra *et al* [62]	vibration	bearings	estimate failure degradation	99.49%	from simulation by MATLAB and test rig
	T. L. Wu *et al* [63]	vibration	the micro-piercing process	failure detection	99%	from the actual machine
	E. Skordilis *et al* [64]			RUL prediction	99%	from the simulated dataset and real industrial settings
	W. Ahmad *et al* [65]	vibration	bearings	RUL prediction	69% (average)	from the PRONOSTIA platform
	J. Phillips *et al* [66]	oil samples	the diesel engine	fault diagnosis	89%	from actual engines
	J. B. Yu *et al* [67]	acoustic and vibration	cutting tools	RUL prediction		from an milling machine
Support Vector Machine	A. Widodo *et al* [69]	acoustic and vibration	bearings	fault diagnosis	97.96%	from the test rig
	G. A. Susto *et al* [23]	pressure, voltage, and current	tungsten filaments	fault diagnosis	98.52%	from a implanter tool
	M. Baptista *et al* [70]		valves of the aircraft engine	fault prediction		from the real engine bleed valve
	U. Shafi *et al* [71]	electric	four main subsystems of vehicle	fault prediction	96.6%, 98.7%, 98%, 96.6%	from sensors on Corolla cars
	A. Soualhi *et al* [72]	vibration	bearings	fault diagnosis and RUL prediction	93.15%,46.51%, 80%	from the PRONOSTIA platform
	O. R. Seryasat *et al* [73]	vibration	bearings	fault classification	50.96%,59.09%, 71.71%	from six test bearings
	C. Li *et al* [74]	vibration	gearbox	fault diagnosis	97.08%, 88.41%	provided by the UPS
	D. L. Yang *et al* [75]	vibration	gearbox	fault diagnosis	94.67%	from a simulator
	M. Saimurugan *et al* [76]	vibration	bearings	fault diagnosis	99.33%	from a simulator
	X. Y. Zhang *et al* [77]	vibration	bearings	fault detection and classification	97.75%, 97.91%	provided by CWRU
	P. J. G. Nieto *et al* [78]	temperature, pressure, and speed	aircraft engines	RUL prediction		from the actual aircraft engine
Decision Tree and Random Forest	A. Krishnakumari [81]	vibration	spur gear	fault diagnosis	95%	from the fault simulator
	G. N. Li *et al* [82]	temperature	variable refrigerant flow system	fault diagnosis	94.44%, 80.55%	from a VRF system and real commercial buildings
	D. Z. Wu *et al* [33]	force, acoustic and vibration	tool wear in dry milling operations	performance evaluation	99.2%	from another paper
	M. Canizo *et al* [18]	status data (activations and deactivations)	wind turbines	fault prediction	82.04%	gathered from wind turbines during two years
	P. Santos *et al* [83]	vibration	gearbox	fault diagnosis		from the test bed
	R. Shrivastava *et al* [84]	flow rate, temperature	bioreactor	fault detection	98.5%	from the simulator
	C. Li *et al* [85]	acoustic and vibration	gearbox	fault diagnosis	97.68%	provided by the UPS

condition monitoring of gas circulator units by using a combination of LR and L2-SVM models. Pandya *et al.* [61] focused on fault diagnosis of rolling bearings by using multinomial LR, and the two features of energy and kurtosis were extracted using wavelet packet decomposition. In addition, by a comparison with an ANN and an SVM, the multi-LR model was considered a more effective classifier with accuracy of 100% and computational time of only 0.33 s. Similarly, Caesarendra *et al.* [62] proposed a hybrid method to estimate the performance degradation from incipient failures, which combines a relevance vector machine (RVM) with LR. However, when the selected kernel width in the simulated and experimental data is improper, the overfitting phenomenon occurs. A vibration-based monitoring approach for the online detection of punch failure in a micropiercing process was proposed in [63], which utilized the statistical overlap factor to select features and obtained more than 98.6% prediction accuracy on three datasets. To monitor the latent degradation level and track the failure progress, a model that combined a Kalman filter with LR was presented in [64]. A dynamic regression model was developed according to [65] for the health prognosis of rolling bearings, and excellent prognostic accuracy was obtained. In addition, Phillips *et al.* [66] argued that the LR model can be easily provided to industrial experts with interpretability and that its predictive performance outperformed that of the ANN and SVM in terms of predicting whether mechanical equipment or components run properly. A novel prognostic method combination of an LR with manifold regularization was proposed in [67] to assess the tool health state. Moreover, it is worth mentioning that a penalization regularization method, which has attained excellent learning performance, was designed to select effective prognostic features.

B. SVM for PdM

1) SVM Model: Typically, the SVM model is used to tackle the tasks for binary classification. In the PdM of industrial equipment, SVMs have been widely applied for identifying a specific status based on the acquired signal [68]. Moreover, because of the diversity of fault types and the ability of mapping low dimension features to hyperplanes, the SVM model can be utilized to solve multiclass tasks.

In summary, the primary purpose of SVM is to find a hyperplane and divide data points correctly on both sides of the hyperplane, and the optimization object is represented by

$$\underset{w,b}{\operatorname{argmax}} \left\{ \frac{1}{\|w\|} \min \left[y_i \left(w^T \cdot x_i + b \right) \right] \right\}$$
$$\text{s.t. } y_i \left(w^T \cdot x_i + b \right) \geq 1 \tag{6}$$

where $(x_i,\ y_i)$ refers to the sample that contains the feature and label.

Therefore, the optimization problem can be transformed to the following:

$$\max_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$$
$$\text{s.t. } y_i \left(w^T \cdot x_i + b \right) \geq 1 - \xi_i \tag{7}$$

where C is a parameter that controls the weight of the relaxation factor, and $0 \leq \alpha \leq C$.

Then, the Lagrange multiplier is applied to the linear SVM objective function and its constraints. The optimization problem of nonlinear SVM is described as follows:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \emptyset (x_i) \emptyset (x_j) - \sum_{i=1}^{N} \alpha_i$$
$$\text{s.t. } \sum_{i-1}^{N} \alpha_i y_i = 0 \tag{8}$$

where $\emptyset(\cdot)$ is a spatial transformation.

2) PdM Applications of SVM: Several PdM applications using SVMs or their variants will be reviewed in this section.

Widodo *et al.* [69] focused on fault diagnosis of bearings by using a multiclass RVM and an SVM; the data were acquired from the low-speed bearing test rig by acoustic emission sensors and accelerometers. Component analysis was also introduced to extract useful features and reduce the dimension of the original features. In addition, it was shown that an SVM outperformed an RVM for the diagnosis of low-speed bearings. A multiple classifier (MC) methodology based on an SVM was proposed in [23] for the PdM of semiconductor manufacturing (e.g., replacing tungsten filaments used in ion implantation). According to the comparison result, the MC-PdM-SVM system was able to achieve better performance than that of the kNN and RBF approaches in terms of unexpected breaks and maintenance costs. In order to extract useful information from raw sensor signals, an autoregressive moving average model with six data-driven algorithms was employed in [70] to formulate airline maintenance scheduling, and the SVM obtained the best overall results. According to [71], four classifiers were compared, namely, SVM, DT, RF, and kNN. Among them, the SVM classifier obtained the best performance on four operating systems, and the accuracy of the SVM was 96.6%, 98.7%, 98%, and 96.6%. The Hilbert-Huang transform was utilized in [72] to extract health indicators from a vibration signal; an SVM model was then used as an effective solution to assess the degradation states, and an SVR model was employed to estimate the RUL of degraded bearings. In contrast, feature extraction was conducted in [73] from the time domain of vibration signals, and a multiclass SVM was deployed to detect bearing faults. A multimodal deep support vector classification model was developed in [74] to diagnose the failure of a gearbox, and the performance was proven to outperform that of the basic SVM. However, when detecting the faults of a gearbox, the identification capacity of an SVM is enormously influenced by the kernel function and its parameters. Therefore, a parameter optimization method based on the artificial bee colony algorithm was proposed in [75] to ensure accurate identification. An SVM model with four kernel functions was applied in [76] for fault recognition, and statistical features were extracted from vibration signals under normal and faulty conditions of a rotating mechanical system. Additionally, in order to better represent the machine status, a DT algorithm was utilized to select the essential features. A hybrid SVM model optimized by the intercluster distance in the feature space was developed in [77] to extract the multiple temporal scale features from a vibration signal and was used to identify the fault types and their severity with respect to motor bearings. The above works are all about the research of SVMs for fault classification and linear regression. However, SVMs can also be used for nonlinear regression. For instance, Nieto *et al.* [78] applied an SVM in a nonlinear model to predict the remaining life of aircraft engines.

C. DT and RF for PdM

1) DT and RF Model: DT classifiers have achieved great success in various fields such as character recognition, medical diagnosis, and speech recognition. Most significantly, a DT model has the ability to decompose a complex decision-making process into a collection of simpler decisions by recursively partitioning the covariate space into subspaces, thus providing a solution that is prone to interpretation [79]. Furthermore, RF is an ensemble learning algorithm composed of multiple DT classifiers, and the category of its output is determined jointly by these individual trees [80]. The RF is provided with many significant advantages. For instance, it can handle high dimensional data without feature selection; trees are independent of each other during training process, and the implementation is relatively simple; in addition, the training speed is usually fast, and at the same time the generalization ability is strong enough.

To construct a DT, the first problem is to determine which feature should be divided first, that is, which feature on the current dataset plays a decisive role in classification. Using the information theory to measure information is an effective method. Information entropy is actually the mathematical expectation of the amount of information in a random variable. To calculate entropy, we need to calculate the expected value of the information contained in all possible values of all categories, which can be obtained by the following formula:

$$H = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{9}$$

where $p(x_i)$ is the probability of choosing x_i, and n indicates the feature number.

In addition, the purity of Gini is also an indicator to measure the purity of information. The formula is as follows:

$$\mathrm{Gini}\,(D) = 1 - \sum_{i=1}^{n} p_i^2 \tag{10}$$

where D refers to the dataset.

2) PdM Applications of DT and RF: In this section, we will review several PdM applications of DT and RF models.

According to [81], a DT algorithm was deployed to perform condition monitoring of a spur gear. The RMS value, sum, skewness, minimum value, and variance were selected from the entire feature set as health indicators, which minimized the involvement and demand of human expertise. An improved diagnosis method was employed in [82] for a practical variable refrigerant flow system, which was implemented by combining a DT model with virtual sensor-based fault indicators, and it outperformed the CART, RF, and generalized boosted regression. A parallel RF algorithm was developed in [33] by using the MapReduce framework to predict tool wear in dry milling operations. An RF-based model was employed in [18] to carry out real-time PdM for wind turbines. According to the obtained results, the depth of the trees had a significant effect on the diagnostic accuracy. In addition, Santos *et al.* [83] conducted a comparison about an intelligent diagnostic model of wind turbines, and a cost-sensitive classifier using a rotation forest ensemble with the C4.4 algorithm was the best model in this research. An RF-based method was utilized in [84] to monitor the health condition of bioreactors from a penicillin production process. In terms of diagnostic accuracy and generalization ability, the overall performance was certified to be better than that of the SVM and ANN. Moreover, RF is easy to employ and can generalize to unseen data. However, in order to obtain optimal performance, the SVM and ANN require extra time and computing resources to tune various parameters. A deep RF fusion technique was utilized in [85] to improve the diagnostic performance of gearboxes.

IV. DL FOR PDM

Because of the lack of expressive ability and the defect of dimensionality [86], traditional ML algorithms require prior experience and expertise to perform fault representation [87]. The theory of DL was first proposed by Hinton and Salakhutdinov in 2006 [88]. In the past decade, DL, a method of extracting structured information from massive datasets using layered ML algorithms, has made almost unimaginable advances in AI and technology industries [20]. Moreover, it has become an indispensable technology in the data-driven PdM. Compared with the traditional shallow model, DL can automatically extract features from the original data and accurately identify the health status [89], [90]. Accordingly, this paper will review the most classic and widely used DL models, including ANN, DNN, and AE. In addition, the comparisons of these models are given in Table V.

A. ANN for PdM

1) ANN Model: Inspired by the biological neural network, ANN is designed to address nonlinear problems. It is a massively parallel computing system consisting of an extremely large number of simple processors with many interconnections. Instead of following a set of rules prescribed by human experts, ANNs appear to learn underlying rules from the given collection of representative examples [91]. The ability of ANN models to automatically learn from examples makes them attractive and extensively applied. Moreover, compared with traditional data-driven methods, NNs have obvious advantages in addressing fuzzy data, random data, and nonlinear data. They are especially suitable for systems with a large scale, a complex structure, and unclear information. Here, ANN refers to an NN with a single hidden layer. Since the DNN was also gradually derived from it, we therefore summarize ANNs in the DL section.

An ANN model consists of two crucial elements: 1) linear summation and 2) nonlinear activation, which can be described as:

$$X_j^{[l]} = f\left(\sum_{i=1}^{M} W_{ij}^{[l]} \cdot X_i^{[l-1]} + B_j^{[l]}\right) \tag{11}$$

where $l = 1, 2$, indicates the hidden layer and output layer, respectively, i the index of input feature maps, and j the index of the output feature maps. Furthermore, the "·" denotes a multiplication operation that is applied to the *i*th feature map $X_i^{[l-1]}$ from the $(l-1)$-th layer, using the *i*th filter $W_{ij}^{[l]}$. In addition, a nonlinear activation function f is applied to each element of all the feature maps. The rectified linear unit (ReLU) $f(z) = \max(0, z)$, the hyperbolic tangent function $f(z) = [\exp(z) - \exp(-z)]/[\exp(z) + \exp(-z)]$, and the logistic function $f(z) = 1/[1 + \exp(-z)]$ are the favored options.

2) PdM Applications of ANN: Next, we will introduce some PdM applications based on ANNs.

The simple yet widely used ANN model was adopted in [45] for diesel engine fault classification. Since the data were collected from the same machine, the selected features can effectively represent the fault information. Consequently, all the training, validation, and test results in this paper had a diagnostic accuracy of 100%. A feedforward NN prediction model was built in [17] to ensure precision estimating on laser welding status. Additionally, an ANN-based model was presented in [92] to improve RUL prediction of bearing failures. A hierarchical NN structure was used in [93] to perform the fault classification of bearings. It was composed of two steps, and each step consisted of a two-layer NN to assure the optimum recognition performance. Similarly, a two-stage learning approach was proposed in [94] to construct a fault diagnosis framework. First, an unsupervised two-layer NN, called sparse filtering, was utilized to learn representative features extracted from mechanical vibration signals. Second, health status identification was performed using a two-layer network, called softmax regression. An ANN-based model was adopted in [95] to perform condition monitoring and fault diagnosis of rolling bearings, and features were extracted on the basis of empirical mode decomposition

TABLE V
SUMMARY OF THE DATA-DRIVEN PDM METHOD USING THREE TYPES OF DL ALGORITHMS

Algorithm	Reference	Signal	Application scenario	Task	Accuracy	Datasets
Artificial Neural Network	C. Jin *et al* [45]	vibration, pressure, and speed	diesel engine	fault diagnosis	100%	from the test bed
	D. Y. You *et al* [17]	optical and visual	laser welding	process monitoring and fault diagnosis	98.2%	from a high-power disk laser welding system
	A. K. Mahamad *et al* [92]	vibration	bearings (rotating)	RUL prediction	100%	provided by the Center of IMS
	M. D. Prieto *et al* [93]	vibration	bearings	fault diagnosis	95%	from an induction motor
	Y. G. Lei *et al* [94]	vibration	bearings	fault diagnosis	99.66%	from a test-rig and CWRU
	J. B. Ali *et al* [95]	vibration	bearings	fault diagnosis	93%	from the IMS
	R. Ahmed *et al* [96]	vibration	a four-stroke gasoline engine	fault detection and classification	97%	from Ford's Powertrain Engineering Research and Development Center
	J. B. Ali *et al* [97]	vibration	bearings	RUL prediction	79.37%	from the test rig
Deep Neural Network	H. X. Hu [98]	vibration	high-speed train system	fault diagnosis	100%	from the multi-body dynamic software SIMPACK
	L. Wang [99]	lubricant, pressure	wind turbine gearbox	identify the impending failures	92.39%	provided by six commercial wind farms in China
	F. Jia *et al* [100]	vibration	bearings and planetary gearboxes	fault diagnosis	99.79%, 99.73%	provided by CWRU and the test rig
	F. N. Zhou *et al* [101]	vibration	bearings	fault and severity recognition	99.79%, 99.52%	from the CWRU
	F. Cipollini *et al* [102]	speed, thrust, and torque	gas turbine	Condition -based maintenance		from the University of California in Irvine
	R. Zhang *et al* [103]	vibration	bearings	fault diagnosis	100%	provided by IMS and CWRU
	M. Heydarzadeh *et al* [104]	vibration, acoustic, and torque	gearbox	fault diagnosis	97.31%, 93.24%, 95.31%	from the test rig
Auto-Encoder	Z. Q. Chen *et al* [105]	vibration	bearings	fault diagnosis	99%	from the UPS
	H. D. Shao *et al* [106]	vibration	gearbox and bearings	fault diagnosis	94.05%, 87.8%	from a test rig
	J. B. Tan *et al* [107]	vibration	bearings	fault diagnosis	99.25% , 99.62%	from the CWRU
	H. D. Shao [108]	vibration	rotor and bearings	fault diagnosis	97.10%, 95.23%, 100%	from the rotor fault test rig, CWRU, and NASA
	C. Lu *et al* [109]	vibration	bearings	fault diagnosis	99.83%, 93.54%	from the CWRU and a test-rig
	N. K. Verma *et al* [110]	acoustic	motors	fault diagnosis	97.22%	from an actual air compressor
	Z. Y. Chen *et al* [111]	vibration	bearings	fault diagnosis	97.82%	from an experimental platform
	W. J. Sun *et al* [112]	vibration	induction motor	fault classification	97.61%	from a machine fault simulator
	H. D. Shao *et al* [113]	vibration	bearings	fault diagnosis	97.18%	from the CWRU

energy entropy. Finally, the proposed method showed that an ANN can be utilized as an effective tool for bearing degradation assessment without human intervention. An engine fault diagnosis method using vibration signal was developed in [96] based on multilayer feedforward ANNs. In order to simplify the prognosis task, the data-driven prognostic method based on the ANN combined with the Weibull distribution was developed for achieving more accurate bearing RUL prediction [97].

B. DNN for PdM

Literally, a DNN is a deeper neural network. Here, it refers to an NN with many hidden layers, and all of the layers are fully connected to each other. Thus, the constraint condition l in formula (11) is also extended to $l' \in \{1, 2, \ldots m\}$. Most importantly, the ability of hierarchical nonlinear mapping learning is a replacement for manual feature acquisition with unsupervised or semisupervised feature learning. Compared with traditional data-driven methods, DNNs can self-adaptively extract fault features to effectively represent crucial information and realize intelligent diagnosis; they can also improve identification accuracy and are extremely effective in reducing defects in manual design features [98].

According to [99], a DNN was employed to develop the prediction model for monitoring the wind turbine gearbox health and identifying impending failures. A five-layer DNN was presented in [100] for fault diagnosis of rolling bearings and planetary gearboxes; it has obtained superior diagnostic accuracy compared with other ANN-based models. In addition, a hierarchical DNN (i.e., three level) was proposed in [101]. The layers of these models were 5, 4, and 3. In detail, the first hierarchy was specially designed for fault identification, the second for locating the fault source, and the third for recognizing the fault severity. As stated in [102], the most recent statistical techniques (i.e., DNN) were applied to a naval propulsion system and have shown the best performance. In addition, the proposed method resulted in an optimal PdM strategy with reduction in the operational and maintenance cost. A DNN-based model was also employed in [103] to perform health monitoring of bearings. Notably, the proposed approach considered temporal coherence with a former time-series signal. A discrete wavelet transform was employed in [104] to provide discriminative features from three common condition monitoring signals including acoustic, vibration, and torque for a diagnostic model. Then, the DNN was utilized to recognize five types of faults of a gearbox and obtained an excellent result.

C. AE for PdM

In the PdM of industrial equipment, labeled raw sensor signals are generally difficult to obtain, requiring specific and detailed experimental settings. Obviously, unsupervised feature learning is especially suitable for handling unlabeled data; it can provide a feasible solution for fault identification and RUL estimation. In addition, the AE model belongs to the type of unsupervised learning that only requires unlabeled measurement data, and it has received extensive attention and application. Accordingly, this section will give a specific presentation about the AE model.

As stated in [105], a multimodel combination method was applied to identify the fault condition of rolling bearings. Among three models, the stacked-AE is slightly better than the deep belief network (DBN), but their diagnostic accuracy is more than 99%. According to [106], a deep AE method was applied to the fault diagnosis of a gearbox and an electrical locomotive bearing, and the corresponding diagnostic accuracy was 94.05% and 87.8%, respectively. Because of the existence of noise disturbance, the composition of a fault vibration signal is relatively complicated. Furthermore, the nonlinear and nonstationary characteristics of a fault vibration signal in a rolling bearing system are considerations. As a result, an intelligent method was proposed in [107] based on wavelet transform and a stacked-AE, where the model was employed to extract features of the denoised signal. Moreover, an enhancement feature fusion method [108] based on deep AE was developed for rotating machinery and was constructed with a denoising AE and a contractive AE for improving the feature learning ability. Similarly, a stacked denoising AE was deployed in [109], which was able to adaptively extract significant fault features and effectively identify health conditions with the best diagnostic performance. In addition, the sparse AE possesses the ability to learn effective feature representation from raw signals in an unsupervised manner. Accordingly, a sparse AE-based architecture was deployed in [110] for motors and an air compressor, and the final diagnostic accuracy achieved was 97.22%. A multisensor data fusion model was developed in [111], which used a two-layer sparse AE and combined it with a DBN to assess health conditions of rotating machinery. An effective approach utilizing sparse AE was developed in [112] for the fault diagnosis of induction motor. Moreover, an ensemble deep AE method was proposed in [113].

V. Analysis of the Metric: About Accuracy

A. Analysis of Accuracy

Both the optimization and innovation of methods are aimed at improving the effectiveness of the industrial PdM. Specifically, the purpose of these applications is to improve the accuracy of fault diagnosis or RUL estimation. Accordingly, we provide several viewpoints regarding the metric of accuracy and combine the statistics in Tables III and IV. Specifically, we focus on those works with the highest and lowest accuracy in each type of algorithm and present our opinions as follows.

1) *LR:* It is well known that the complexity of the LR model is the lowest among various ML algorithms. However, it still achieves 100% accuracy in some applications, such as in [60] and [61]. There were three sensor signals (i.e., displacement, vibration, and current) in [60] used for the fault representation learning, and 19-dimension features were selected as input vectors of an LR model using wavelet packet decomposition and Fisher's criterion. Similarly, Pandya *et al.* [61] paid more attention to feature engineering, and there were five real wavelets and two wavelet selection criteria utilized to perform fault diagnosis. Perhaps, most importantly, the experimental data were acquired from a test rig (e.g., a motor), which resulted in lower data acquisition cost. In addition, the experimental data can be filtered from a large amount of collected data, which is more helpful for model training. Therefore, we consider it is reasonable to achieve 100% accuracy on the basis of such adequate data acquisition and feature engineering. In contrast, the accuracy does not seem optimistic in [65], at just 69%. However, it is worth mentioning that unlike fault diagnosis (i.e., classification), the task faced in this paper is RUL estimation (i.e., regression), and the accuracy here is the average of 11 bearings. In addition, compared with other works, this paper introduced less regarding feature engineering, and perhaps it is the reason for the low accuracy.
2) *SVM:* In reality, the SVM is not only good at classification tasks but also performs well in fault diagnosis. Data-driven algorithms mainly depend on the ability to carry out fault representation via learning from sensor signals. Moreover, Saimurugan *et al.* [76] conducted a multiclass bearing fault diagnosis task. The difference between the two types of fault signals may be especially slight, which will lead to the misclassification between classes, just as a7, a9, a10, and a12 are misclassified in the experiments. Accordingly, we consider that the diagnostic accuracy of 99.33% obtained in this paper is already considerable. In contrast, a six-class diagnosis task was conducted in [73]. According to the three confusion matrixes given in this paper, we calculate the accuracy of the proposed algorithm under three different speed and load conditions, as only 50.96%, 59.09%, 71.71%. In addition, the dataset scale and the data processing flow are not explained in detail in this paper, so it is not easy for us to provide an in-depth evaluation of the results.
3) *DT and RF:* It is well known that the DT model inherits a well interpretable and suitable approach for processing high dimensional data but is prone to overfitting. However, RF solves the problem of weak generalization ability for DT. As a result, Wu *et al.* [33] obtained excellent performance in the industrial PdM. Specifically, the authors collected the signals from three sensors (i.e., cutting force, vibration, and acoustic emission), which formed a seven-channel data acquisition system, and then four features were extracted from each channel. In summary, a detailed feature engineering task has been conducted in this paper, which is in accordance with the characteristic of the proposed model. Therefore, we deem that the tool wear prediction accuracy of 99.2% achieved by using 28-dimension features is an evidence-based result. In contrast, Canizo *et al.* [18] only obtained accuracy of 82.04%. Nevertheless, it is worth noting that this paper was based on another existing work, and although the obtained accuracy does not seem as good, it still achieved an increase of 5.54%.

4) *ANN, DNN, and AE:* It can be seen that among the three models of DL, in 80% of the literature, they have achieved more than 97% accuracy, and five papers have even reached 100%, including [45], [92], [98], [103], and [108]. These papers deployed an NN to arm models with the ability to represent the equipment health condition by using hierarchical learning. A simple NN is constructed in [44] and [90] with a single-hidden layer, and the information learned from sensor signals is transmitted forward through linear weighting and a nonlinear activation function. Especially for [44], 20 features are extracted from the time, frequency, and time-frequency domain using three sensors (i.e., vibration, pressure, and speed), and the approach is sufficient to support the model training for a four-class task. In particular, a large dataset was built for model training, and the data they used were all from the same equipment. Moreover, more hidden layers were deployed in [98] and [103] to extend the network depth, ensuring that the network has a powerful learning ability. Especially for [103], a six-layer DNN was built to conduct fault diagnosis. In addition, the temporal coherence, which makes full use of the time characteristic for the sensor sequence, is taken into consideration. A four-layer AE model that belongs to unsupervised learning was deployed in [108]. The enhanced feature learning ability enabled the proposed model to more accurately distinguish different conditions. In summary, DL has demonstrated its powerful capabilities in the industrial PdM. Specifically, building a deeper model allows for a more strong representation, which will be extremely helpful in improving prediction accuracy.

In summary, for traditional ML algorithms, sufficient data acquisition and feature engineering are helpful in optimizing algorithm performance, while for DL algorithms, the deeper network architecture and the higher dimensional feature vectors are more significant in improving the task metric.

B. Several Conclusions

In this section, some conclusions are drawn from Tables III and IV and are shown as follows.

1) Both ML and DL can remarkably complete the PdM task. The average prediction accuracy of the reviewed literature can reach 95.06%, and the highest accuracy is 100%.
2) Signals that can be used for fault diagnosis include acoustic emission, electrical signature parameters (current and voltage), temperature, pressure, rotation speed, and vibration. However, vibration signals are exploited most frequently in terms of the literature surveyed herein.
3) By using comparison, the monitored key components such as bearings and cutting blades are broadly divided into rotating machinery and reciprocating machinery.
4) In terms of the tasks completed, they are mainly divided into two aspects: fault diagnosis and RUL prediction. In some cases, it is also referred to the performance degradation monitoring of equipment.
5) For these applications, a noteworthy point is the source of the dataset. As seen from our summaries, most datasets come from public data centers, such as case western reserve university/intelligent maintenance systems, or a lab-level experimental platform (i.e., test rig). Only a few papers utilize datasets that are collected from the actual operating equipment.
6) With the great improvement of computational power and the rapid growth of data volume, AI algorithms and their variant models have increasingly demonstrated a superior advantage of performance. With the continuous innovation of algorithms, researchers will continue to focus on the data-driven methods in PdM applications.
7) In terms of model performance, these algorithms are applicable to most industrial applications. But, in the aspect of revealing the essential reasons, the existing algorithms not only lack interpretability but also lack the ability to explain specific phenomena.

VI. Future Research Directions

Although data-driven methods have achieved excellent performance in PdM applications, there is still a large potential for improvement and optimization, especially for practical applications. Accordingly, some research trends and potential directions are given as follows.

1) *Data validity:* As is known, the significance of data to an algorithm's performance is self-evident. However, a large number of studies are currently only using public datasets provided by some platforms, but fewer datasets originate from the actual operating equipment. In addition, the construction of the data acquisition system will be expensive, and the sensor itself also has the potential to fail. When the cost of the diagnosis is higher than that of the maintenance, the diagnosis will lose its original meaning in a sense. For these reasons, reliable CPS and IoT (e.g., IWSNs) technologies are required to provide low acquisition cost and high utilization value for a PdM study. This approach facilitates researchers' capability to make full use of data from the actual operating equipment, rather than from the experimental platform, and to better solve the practical problems in industrial manufacturing.
2) *Dataset construction and the unbalanced feature of datasets:* There are many challenges that need to be resolved when handling mechanical data, such as the accuracy and the actual meaning of the label, as well as the importance of the label for the PdM. In most instances, the observed data volume for normal operations far exceeds the observed data volume for an anomalous operation, which will cause great difficulties in conducting model training. As a result, it is especially necessary to acquire more data under fault status scenarios, which will make the entire dataset more balanced.
3) *Unsupervised learning:* As mentioned above, the current achievements of AI are mainly focused on supervised learning, which means that datasets need to be uniquely annotated. The challenges of this task are described in the previous item. Accordingly, learning with unlabeled data, namely unsupervised learning, is a significant research direction in the future.

4) *Model migration and generalization ability:* For practical applications, the model migration ability is identically important. Specifically, a trained model will ensure more applicability when it can be applied to other equipment, which can not only save unnecessary time consumption but also simplify the PdM design process.
5) *Multisensory technique:* The first wave of DL is using massive amounts of data to train very large neural networks, and many of the breakthroughs were based on this manner. However, not all tasks are liable to obtain large volume of effective data, so it is urgent to develop robust algorithm on small scale dataset, and multisensory fusion technique provides possibilities.
6) *Safety:* Here, safety refers to the loss caused by system misdiagnosis. It can be divided into three categories: positive misdiagnosis (i.e., no damage but damage is diagnosed), intermediate misdiagnosis (i.e., one fault type is diagnosed as another fault type), and negative misdiagnosis (i.e., the damage exists, but no damage is diagnosed). Obviously, any misdiagnosis will cause a certain degree of loss to the system and may even cause system breakdown. Consequently, how to design a safe and reliable PdM system has become another major challenge.

VII. Conclusion

In this paper, we conduct a comprehensive survey of the PdM of industrial equipment. First, we provide a systematic overview of the PdM, propose an industrial PdM scheme for *automatic washing equipment*, and demonstrate the challenges faced when we conduct a PdM research study. Second, we review the industrial applications of PdM in the recent five-year timeframe both from ML and DL perspectives, and then five indicators (e.g., signal type, application scenario, target, accuracy, and data source) are selected to summarize the application of these algorithms. Third, we summarize some conclusions from the entire PdM perspective, and the accuracy (a metric to estimate the algorithm performance) of each algorithm is analyzed in detail. Finally, we discuss several research trends and future directions.

In summary, although PdM-related works are emerging endlessly, the challenges that the PdM is facing are self-evident concerns. Accordingly, we will pay more attention to addressing these challenges in practical applications. We hope this paper can provide graduate students, companies, and institutions with a preliminary understanding of PdM research.

References

[1] H. S. Kang *et al.*, “Smart manufacturing: Past research, present findings, and future directions,” *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 3, no. 1, pp. 111–128, 2016.

[2] J. Lee, B. Bagheri, and H. A. Kao, “A cyber-physical systems architecture for industry 4.0-based manufacturing systems,” *Manuf. Lett.*, vol. 3, pp. 18–23, 2015.

[3] H. Lasi *et al.*, “Industry 4.0. business & information systems engineering,” *Bus. Inf. Syst. Eng.*, vol. 6, no. 4, pp. 239–242, 2014.

[4] L. Ling, “China’s manufacturing locus in 2025: With a comparison of “made-in-china 2025” and “industry 4.0”,” *Technol. Forecasting Social Change*, vol. 135, pp. 66–74, 2017.

[5] J. Lee, “*Industrial Big Data*, China: Mechanical Industry Press, Beijing, 2015.

[6] D. H. Zhou, M. Y. Chen, and Z. G. Xu, *Reliability Prediction and Optimal Maintenance Technology*, Univ. Sci. Technol. China Press, Hefei City, Anhui Province, 2013.

[7] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[8] K. Wang, “Intelligent predictive maintenance (IPdM) system—Industry 4.0 scenario,” *WIT Trans. Eng. Sci.*, vol. 113, pp. 259–268, 2016.

[9] S. Cheng, K. Tom, L. Thomas, and M. Pecht, “A wireless sensor system for prognostics and health management,” *IEEE Sensors J.*, vol. 10, no. 4, pp. 856–862, Apr. 2010.

[10] K. Krishnamoorthy, P. Ashok, and D. Tesar, “Simultaneous sensor and process fault detection and isolation in multiple-input–multiple-output systems,” *IEEE Syst. J.*, vol. 9, no. 2, pp. 335–349, Jun. 2015.

[11] Y. Jiang, S. Yin, and O. Kaynak, “Data-driven monitoring and safety control of industrial cyber-physical systems: Basics and beyond,” *IEEE Access*, vol. 6, pp. 47374–47384, 2018.

[12] D. Yang et al., “Assignment of segmented slots enabling reliable real-time transmission in industrial wireless sensor networks,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3966–3977, Jun. 2015.

[13] L. D. Xu, W. He, and S. Li, “Internet of Things in industries: A survey,” *IEEE Trans. Ind. Inform.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[14] X. Li *et al.*, “A review of industrial wireless networks in the context of industry 4.0,” *Wireless Netw.*, vol. 23, no. 1, pp. 23–41, 2017.

[15] J. J. A. Costello, G. M. West, and S. D. J. Mcarthur, “Machine learning model for event-based prognostics in gas circulator condition monitoring,” *IEEE Trans. Rel.*, vol. 66, no. 4, pp. 1048–1057, Dec. 2017.

[16] Y. Jiang and S. Yin, “Recursive total principle component regression based fault detection and its application to vehicular cyber-physical systems,” *IEEE Trans. Ind. Inform.*, vol. 14, no. 4, pp. 1415–1423, Apr. 2018.

[17] D. You, X. Gao, and S. Katayama, “WPD-PCA-based laser welding process monitoring and defects diagnosis by using FNN and SVM,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 628–636, Jan. 2015.

[18] M. Canizo, E. Onieva, A. Conde, S. Charramendieta, and S. Trujillo, “Real-time predictive maintenance for wind turbines using Big Data frameworks,” in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2017, pp. 70–77.

[19] Z. Gao, C. Cecati, and S. X. Ding, “A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[20] R. Zhao *et al.*, “Deep learning and its applications to machine health monitoring,” *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, 2019.

[21] R. Liu *et al.*, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, 2018.

[22] G. Smith, J. B. Schroeder, S. Navarro, and D. Haldeman, “Development of a prognostics and health management capability for the Joint Strike Fighter,” in *Proc. IEEE Autotestcon* , 1997, pp. 676–682.

[23] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, “Machine learning for predictive maintenance: A multiple classifier approach,” *IEEE Trans. Ind. Inform.*, vol. 11, no. 3, pp. 812–820, Jun. 2015.

[24] J. Wan *et al.*, “A manufacturing big data solution for active preventive maintenance,” *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 2039–2047, Aug. 2017.

[25] W. Qiao and D. Lu, “A survey on wind turbine condition monitoring and fault diagnosis—Part I: Components and subsystems,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6536–6545, Oct. 2015.

[26] J. Lee *et al.*, “Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications,” *Mech. Syst. Signal Process.*, vol. 42, no. 1-2, pp. 314–334, 2014.

[27] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song , “Industrial big data for fault diagnosis: Taxonomy, review, and applications,” *IEEE Access*, vol. 5, no. 99, pp. 17368–17380, 2017.

[28] Y. Jiang and S. Yin, “Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: DB-KIT,” *IEEE Trans. Ind. Inform.*, vol. 15, no. 5, pp. 2849–2858, May 2019.

[29] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*. Hoboken, NJ, USA: Wiley, 2012.

[30] Z. Chang, L. Lei, Z. Y. Zhou, S. Mao, and T. Ristaniemi, “Learn to cache: Machine learning for network edge caching in the Big Data era,” *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 28–35, Jun. 2018.

[31] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," *IEEE Netw.*, vol. 32, no. 1, pp. 96–101, Jan./Feb. 2018.

[32] Z. Jiang, Z. Li, N. Wu, and M. Zhou, "A Petri net approach to fault diagnosis and restoration for power transmission systems to avoid the output interruption of substations," *IEEE Syst. J.*, vol. 12, no. 3, pp. 2566–2576, Sep. 2017.

[33] D. Wu, C. Jennings, J. Terpenny, and S. Kumara, "Cloud-based machine learning for predictive analytics: Tool wear prediction in milling," in *Proc. IEEE Int. Conf. Big Data*, 2017, pp. 2062–2069.

[34] G. A. Susto *et al.*, "A predictive maintenance system based on regularization methods for ion-implantation," *Adv. Semicond. Manuf. Conf.*, vol. 497, pp. 175–180, 2012.

[35] W. Zhou, T. G. Habetler, and R. G. Harley, "Bearing condition monitoring methods for electric machines: A general review," in *Proc. IEEE Int. Symp. Diagnostics Elect. Mach., Power Electron. Drives*, 2007, pp. 3–6.

[36] V. T. Tran, F. Althobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, 2014.

[37] M. J. Schulz and M. J. Sundaresan, "Smart sensor system for structural condition monitoring of wind turbines," in *Proc. Energy Convers. Congr. Expo.*, 2006, pp. 2417–2421.

[38] G. K. Singh and S. A. S. Al Kazzaz, "Development of an intelligent diagnostic system for induction machine health monitoring," *IEEE Syst. J.*, vol. 2, no. 2, pp. 273–288, Jun. 2008.

[39] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.

[40] G. W. Cai, *Big Data Analytics in Structural Health Monitoring*, Nashville, TN, USA: Vanderbilt Univ., 2017.

[41] S. Hong and Z. Zhou, "Application of Gaussian process regression for bearing degradation assessment," *Inf. Sci. Service Sci. Data Mining*, vol. 42, pp. 644–648, 2013.

[42] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis—Part II: Signals and signal processing methods," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6546–6557, Oct. 2015.

[43] Y. Qu *et al.*, "A new acoustic emission sensor based gear fault detection approach," *Int. J. Prognostics Health Manage.*, vol. 4, pp. 32–45, 2013.

[44] F. Immovilli, A. Bellini, R. Rubini, and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison," *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1350–1359, Jul.–Aug. 2010.

[45] C. Jin, W. Zhao, Z. Liu, J. Lee, and X. He, "A vibration-based approach for diesel engine fault diagnosis," in *Proc. Int. Conf. Prognostics Health Manage.*, pp. 1–9, 2014.

[46] B. Jie, L. Hou, and M. Yongguang, "Machine fault diagnosis using industrial wireless sensor networks and support vector machine," in *Proc. 12th IEEE Int. Conf. Electron. Meas. Instrum.*, 2015, vol. 1, pp. 153–158.

[47] R. K. Mobley, *An Introduction to Predictive Maintenance*. New York, NY, USA: Elsevier, 2002.

[48] D. Lu, X. Gong, and W. Qiao, "Current-based diagnosis for gear tooth breaks in wind turbine gearboxes," in *Proc. Energy Convers. Congr. Expo.*, 2012, vol. 11, pp. 3780–3786.

[49] P. Mendonca *et al.*, "Development of a reduced-model laboratory for testing predictive fault system in internal combustion engines," in *Proc. IEEE 10th Int. Symp. Diagnostics Elect. Mach., Power Electron. Drives*, 2015, pp. 428–434.

[50] C. Chen, B. Zhang, and G. Vachtsevanos, "Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 2, pp. 297–306, Feb. 2012.

[51] X. Liu, P. Song, C. Yang, C. Hao, and W. Peng, "Prognostics and health management of bearings based on logarithmic linear recursive least-squares and recursive maximum likelihood estimation," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1549–1558, Feb. 2018.

[52] B. Cline *et al.*, "Predictive maintenance applications for machine learning," in *Proc. Rel. Maintainability Symp.*, 2017, pp. 1–7.

[53] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.

[54] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814–3824, May 2019.

[55] Y. Wang, Z. Wei, and J. Yang, "Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machines," *IEEE Trans. Ind. Inform.*, vol. 15, no. 1, pp. 105–115, Jan. 2018.

[56] A. Manghai, R. Jegadeeshwaran, and V. Sugumaran, "Brake fault diagnosis through machine learning approaches—A review," *SDHM Struct. Durability Health Monit.*, vol. 12, no. 1, pp. 43–67, 2017

[57] T. Wuest *et al.*, "Machine learning in manufacturing: Advantages, challenges, and applications," *Prod. Manuf. Res. An Open Access J.*, vol. 4, no. 1, pp. 23–45, 2016.

[58] J. Hosmer *et al.*, *Applied Logistic Regression*. Hoboken, NJ, USA: Wiley, 2013, vol. 398.

[59] H. Li *et al.*, "Cutting tool operational reliability prediction based on acoustic emission and logistic regression model," *J. Intell. Manuf.*, vol. 26, no. 5, pp. 923–931, 2015.

[60] J. Yan and J. Lee, "Degradation assessment and fault modes classification using logistic regression," *J. Manuf. Sci. Eng.*, vol. 127, no. 4, pp. 912–914, 2005.

[61] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform," *Soft Comput.*, vol. 18, no. 2, pp. 255–266, 2014.

[62] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1161–1171, 2010.

[63] T.-L. Wu, D. Y. Sari, B.-T. Lin, and C.-W. Chang, "Monitoring of punch failure in micro-piercing process based on vibratory signal and logistic regression," *Int. J. Adv. Manuf. Technol.*, vol. 93, no. 5–8, pp. 2447–2458, 2017.

[64] E. Skordilis and R. Moghaddass, "A condition monitoring approach for real-time monitoring of degrading systems using Kalman filter and logistic regression," *Int. J. Prod. Res.*, vol. 55, no. 19, pp. 1–18, 2017.

[65] W. Ahmad *et al.*, "A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models," *Rel. Eng. Syst. Saf.*, 2018.

[66] J. Phillips *et al.*, "Classifying machinery condition using oil samples and binary logistic regression," *Mech. Syst. Signal Process.*, vol. 60–61, pp. 316–325, 2015.

[67] J. Yu, "Tool condition prognostics using logistic regression with penalization and manifold regularization," *Appl. Soft Comput.*, vol. 64, pp. 453–467, 2017.

[68] A. Widodo and B. S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007.

[69] A. Widodo *et al.*, "Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 7252–7261, 2009.

[70] M. Baptista *et al.*, "Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling," *Comput. Ind. Eng.*, vol. 115, pp. 41–53, 2018.

[71] U. Shafi *et al.*, "Vehicle remote health monitoring and prognostic maintenance system," *J. Adv. Transp.*, no. 1, pp. 1–10, 2018.

[72] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.

[73] O. R. Seryasat, M. A. Shoorehdeli, F. Honarvar, and A. Rahmani, "Multi-fault diagnosis of ball bearing based on features extracted from time-domain and multi-class support vector machine (MSVM)," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2010, pp. 4300–4303.

[74] C. Li *et al.*, "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis," *Neurocomputing*, vol. 168, pp. 119–127, 2015.

[75] D. Yang *et al.*, "Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm," *Mechanism Mach. Theory*, vol. 90, pp. 219–229, 2015.

[76] M. Saimurugan *et al.*, "Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3819–3826, 2011.

[77] X. Zhang, Y. Liang, and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.

[78] P. J. G. Nieto *et al.*, "Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability," *Rel. Eng. Syst. Saf.*, vol. 138, pp. 219–231, 2015.

[79] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991.
[80] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[81] A. Krishnakumari *et al.*, "Fault diagnostics of spur gear using decision tree and fuzzy classifier," *Int. J. Adv. Manuf. Technol.*, vol. 89, no. 9–12, pp. 1–8, 2017.
[82] G. Li *et al.*, "An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators," *Appl. Thermal Eng.*, vol. 129, pp. 1292–1303, 2018.
[83] P. Santos, J. Maudes, and A. Bustillo, "Identifying maximum imbalance in datasets for fault diagnosis of gearboxes," *J. Intell. Manuf.*, vol. 29, no. 2, pp. 333–351, 2018.
[84] R. Shrivastava, H. Mahalingam, and N. N. Dutta, "Application and evaluation of random forest classifier technique for fault detection in bioreactor operation," *Chem. Eng. Commun.*, vol. 204, no. 5, pp. 591–598, 2017.
[85] C. Li *et al.*, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mech. Syst. Signal Process.*, vol. 76, pp. 283–293, 2016.
[86] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[87] F. Lv, C. Wen, Z. Bao, and M. Liu, "Fault diagnosis based on deep learning," in *Proc. Amer. Control Conf.*, 2016, pp. 6851–6856.
[88] G.E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[89] G. Zhao *et al.*, "Research advances in fault diagnosis and prognostic based on deep learning," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chengdu)*, 2016, pp. 1–6.
[90] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.
[91] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996.
[92] A. K. Mahamad, S. Saon, and T. Hiyama, "Predicting remaining useful life of rotating machinery based artificial neural network," *Comput. Math. Appl.*, vol. 60, no. 4, pp. 1078–1087, 2010.
[93] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
[94] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
[95] J. B. Ali *et al.*, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, pp. 16–27, 2015.
[96] R. Ahmed, M. El Sayed, S. A. Gadsden, J. Tjong, and S. Habibi, "Automotive internal-combustion-engine fault detection and classification using artificial neural network techniques," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 21–33, Jan. 2015.
[97] J. B. Ali *et al.*, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vol. 56, pp. 150–172, 2015.
[98] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 2106–2116, Aug. 2017.
[99] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1360–1368, Jun. 2017.
[100] F. Jia *et al.*, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vol. 72, pp. 303–315, 2016.
[101] F. Zhou, Y. Gao, and C. Wen, "A novel multimode fault classification method based on deep learning," *J. Control Sci. Eng.*, vol. 2017, no. 1, pp. 1–14, 2017, Art. no. 3583610.
[102] F. Cipollini *et al.*, "Condition-based maintenance of naval propulsion systems with supervised data analysis," *Ocean Eng.*, vol. 149, pp. 268–278, 2018.
[103] R. Zhang *et al.*, "Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence," *Sensors*, vol. 17, no. 3, 2017, Art. no. E549.
[104] M. Heydarzadeh *et al.*, "Gear fault diagnosis using discrete wavelet transform and deep neural networks," in *Proc. Annu. Conf. Ind. Electron. Soc.*, 2016, pp. 1494–1500.
[105] Z. Chen *et al.*, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectron. Rel.*, vol. 75, pp. 327–333, 2017.
[106] H. Shao *et al.*, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 95, pp. 187–204, 2017.
[107] T. Junbo *et al.*, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *Proc. Control Decis. Conf.*, 2015, pp. 4608–4613.
[108] H. Shao *et al.*, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, 2016.
[109] C. Lu *et al.*, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, 2017.
[110] N. K. Verma, V. K. Gupta, M. Sharma, and R. K. Sevakula, "Intelligent condition based monitoring of rotating machines using sparse auto-encoders," in *Proc. IEEE Conf. Prognostics Health Manage.*, 2013, vol. 7789, pp. 1–7.
[111] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.
[112] W. Sun *et al.*, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
[113] H. Shao *et al.*, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, 2018.

Weiting Zhang received the M.S. degree from the Inner Mongolia University of Technology, Hohhot, China, in 2017. He is currently working toward the Ph.D. degree in communication and information systems at Beijing Jiaotong University, Beijing, China.

His specific areas of research interest mainly focus on deep learning, predictive maintenance, and wireless sensor networks.

Dong Yang (M'11) received the B.S. degree from Central South University, Changsha, China, in 2003 and the Ph.D. degree in communications and information science from Beijing Jiaotong University, Beijing, China, in 2009.

From March 2009 to June 2010, he was a Post-Doctoral Research Associate with Jonkoping University, Jonkoping, Sweden. In August 2010, he joined the School of Electronic and Information Engineering, Beijing Jiaotong University. Currently, he is working as a Professor. His research interests are network technologies, including Internet architecture, industrial Internet, and wireless sensor networks.

Hongchao Wang (M'14) received the B.S. degree in communication engineering and the Ph.D. degree in communication and information systems from Beijing Jiaotong University, Beijing, China, in 2005 and 2012, respectively.

Currently he is working with the School of Electronic and Information Engineering, Beijing Jiaotong University. His research interests include Internet architecture, network security, and wireless sensor networks.