Measurement 187 (2022) 110276

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective

Yuxin Wen [a], Md. Fashiar Rahman [b], Honglun Xu [c], Tzu-Liang Bill Tseng [b,*]

[a] Dale E. and Sarah Ann Fowler School of Engineering, Chapman University, CA 92866, USA
[b] Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at El Paso, TX 79968, USA
[c] Computational Science Program, The University of Texas at El Paso, TX 79968, USA

ARTICLE INFO

Keywords:
Machine prognostics
Predictive maintenance
Condition-based maintenance
Machine learning
Prognostics and health management
Remaining useful life

ABSTRACT

In the Engineering discipline, prognostics play an essential role in improving system safety, reliability and enabling predictive maintenance decision-making. Due to the adoption of emerging sensing techniques and big data analytics tools, data-driven prognostic approaches are gaining popularity. This paper aims to deliver an extensive review of recent advances and trends of data-driven machine prognostics, with a focus on their applications in practice. The primary purpose of this review is to categorize existing literature and report the latest research progress and directions to support researchers and practitioners in acquiring a clear comprehension of the subject area. This paper first summarizes fundamental methodologies on data-driven approaches for predictive maintenance. Then, the article further conducts a comprehensive investigation on the different fields of applications of machine prognostics. Finally, a discussion on the challenges, opportunities, and future trends of predictive maintenance is presented to conclude this paper.

1. Introduction

In the past decades, the science of prognostics and health management (PHM) of complex engineering systems attracts the research community and industrial practitioners [1]. The primary motivation of PHM is to improve system safety, increase machine reliability and availability, and reduce maintenance costs. Rapid development in digital technologies, such as 3D printing, robots, artificial intelligence (AI), digital automation, cloud computing, the Internet of Things (IoT), and more, has accelerated the Fourth Industrial Revolution, which is often known as Industry 4.0 [2]. In the context of Industry 4.0, to satisfy the escalating demand for functionality and quality, systems have become much more interconnected and complex than any other time before. Therefore, diligent monitoring of the system operations and then proposing appropriate maintenance strategies have become indispensable since an unexpected failure can result in catastrophic consequences. A proper maintenance strategy is critical to minimize unplanned downtime and ensure that facilities operate at the highest efficiency.

Corrective and preventive maintenances are the two widely used maintenance strategies since the early 90s [3]. Corrective maintenance, also called failure-driven maintenance, is carried out only after the occurrence of a malfunction or breakdown of equipment. However, it frequently results in unpredictable performance in the industry, i.e., high production cost, extensive repair time, not to mention the cost and penalties associated with machine breakdown. Preventive maintenance is carried out regularly. Most maintenance decisions are made by experts based on their experience with equipment fabricants, historic breakdowns or failure data. However, it is difficult to make a proper maintenance schedule in advance. Due to the increasing requirement of reliability, availability, maintainability, and safety of systems, preventive maintenance is becoming less effective and obsolete. Recently, predictive maintenance, also known as condition-based maintenance [4], which uses predictive tools to determine when maintenance actions are necessary, has become prevalent in the industry due to the capability of reducing maintenance cost, unexpected downtime, and while extending the life span of equipment [5]. Predictive maintenance employs non-intrusive testing techniques, such as thermodynamics, acoustics, vibration analysis, infrared analysis, etc., to monitor and evaluate equipment performance trends. The core procedures for implementing predictive maintenance include data collection, fault detection and diagnostics, and prognostics, which are later used to guide maintenance decisions such as maintenance scheduling or resource optimization. A comprehensive review of each component can be found in [6–8]. Of these procedures, diagnostics and prognostics are two

* Corresponding author.
E-mail address: btseng@utep.edu (T.-L.B. Tseng).

https://doi.org/10.1016/j.measurement.2021.110276
Received 11 July 2021; Received in revised form 28 September 2021; Accepted 2 October 2021
Available online 8 October 2021
0263-2241/Published by Elsevier Ltd.

critical aspects of predictive maintenance. Diagnostics refer to identifying the presence of operational faults and determining the root cause and effect to the functional equipment. In contrast, prognostics deal with predicting the future state or remaining useful life (RUL) based on the current and historical conditions. Accurately predicting "life-span" is the key to the success of predictive maintenance. It involves analytical computations of historical or real-time data streamed from applications, sensors, devices, etc. In general, prognostics measures the extent of deviation and degradation of any machine or system from the normal operating behavior to predicts the RUL and future performance. However, the task of prognostics is not trivial as predicting future performance depends on the analysis of failure modes, early signals of wear and aging, and the nature of faults. It also requires sound knowledge of the failure mechanisms which have a certain amount of physical randomness. Moreover, prognostics identifies the potential system parameters that are likely to cause the degradations, leading to eventual failures, which involves considerable uncertainty and complicates the prediction. Therefore, prognostics is much more challenging than diagnostics and requires effective and efficient predictive models to monitor the machine health conditions.

In general, prognostic models can be classified into two groups, physical-based and data-driven models [9]. The physical-based models capture the failure mechanisms or physical phenomena to build a mathematical representation of the degradation process. It always requires a thorough understanding of the sophisticated degradation mechanisms, making it infeasible or ineffective in practical applications due to the system complexity or unclear degrading mechanism [10,11]. On the other hand, data-driven prognostics usually deploy data mining techniques to identify the pattern and anomalies within the raw signals/data to detect any changes in system states. Due to the promising applications and data availability, data-driven models are becoming attractive in recent years. Data-driven models can be further classified into three subcategories, statistical-based models, conventional machine learning based models, and deep learning based models. In the first subcategory, the general path and stochastic process models are usually designed to track the trajectory of the degradation in a probabilistic manner. Conventional machine learning approaches, including random forest (RF), artificial neural network (ANN), support vector machine (SVM), etc., are commonly designed to extract features for machine RUL prediction. As data increases in dimensionality and volume, deep learning with automatic feature learning demonstrates outstanding performance in reliability estimation using degradation data.

Several efforts have been performed to review the topic of degradation modeling and machine prognostics in the recent decade. Si et al. provided a systematic review on data-driven models for RUL prediction. They classified the data-driven models into two broad types of models according to the criterion that if the models rely on directly observed state information or not [12]. Ye and Xie [13] classified existing degradation models into general path models, stochastic process models, and others with a focus on stochastic models. Zhang et al. [14] provided a review on degradation model-based RUL estimation approaches with an emphasis on the heterogeneity in the systems. All of the above works only focused on statistical-based models. It has been observed a proliferation of data-driven algorithms to help with prognostics in the latest few years. For example, Wang et al. [15] provided an in-depth review on health indicator construction for vibration-based bearing and gear. Khan and Yairi [16] presented a systematic review of artificial intelligence based system health management and recent trends of deep learning in the reliability field. Recently, [17 18] surveyed contemporary advancements of deep learning and its applications to machine health monitoring. All of them only survey AI techniques for system health management. In another recent work, Kordestani et al. [19] and Guo et al. [20] reviewed and summarized the emerging prognostic modeling methods, which can be classified into data-driven, physics-based/model-based, and hybrid approaches. Baur et al. [21] presented a review on diagnostics and prognostics approaches from knowledge-based, model-based, statistical-based, and data-driven concepts. This work only focuses on applying machine tools considering the feed axis, spindle speed, and hydraulic system. Table 1 highlights the contribution of some review works in the last decade.

Unlike the above-mentioned review works, this paper aims to provide an extensive and broad overview of the most recent advances and trends of data-driven machine prognostics for predictive maintenance, focusing on their applications in different industrial fields. This article provides a detailed discussion and recent advances in each of the categories of data-driven approaches for recent five years. The primary motivation of this survey is to categorize the existing literature and summarize the latest research progress and directions to assist researchers and practitioners in acquiring a clear comprehension of the subject area. The paper also discusses applied research issues when applying current technology and suggests some potentially promising directions for predictive maintenance.

The remainder of this paper is organized as follows. Section 2 provides a discussion on typical data-driven predictive models and analysis methods. The applications for each category of data-driven approaches are summarized in Section 3. Section 4 presents the current challenges and opportunities of machine prognostics for predictive maintenance. Section 5 concludes this work.

2. Data-driven prognostic algorithms

Prognostics algorithms focus on predicting when a system or a component stops to perform its intended functions. In other words, the prognostic algorithms predict the future performance or the RUL of a system or component by analyzing the extent of deviation and degradation from its expected normal operating conditions. In general, the health state of an item degrades linearly with its usage or operating cycle. However, the task of prognostics is not trivial due to the variation of operation conditions, environment, and complex nature of different parameters. Prognosis requires intensive degradation data of an item, such as lifetime data or run-to-failure histories. To make accurate prognostics, choosing a proper modeling technique is essential. There are mainly three modeling strategies for predictive maintenance based on degradation data: (1) regression models, (2) classification models, (3) survival models. A regression model seeks to model the trajectory of a degradation path and then predict when the system will fail. A classification model tries to predict if the failure occurs within a given time window. The basic idea of survival models is trying to answer how the risk of failure changes in time. To implement these strategies, data-

Table 1
Recent contribution of the review works in the field of PHM.

Reference	Year	Description
[12]	2013	The review only focused on statistical data-driven approaches
[13]	2014	Reviewed based on three categories: statistical process models (SPMs), general path models (GPMs), and other models beyond SPMs and GPMs
[22]	2014	Summarized the PHM systems for rotating machinery and provides a systematic design methodology
[14]	2015	Focused on the degradation modeling and RUL estimation for heterogeneity in the systems
[15]	2017	Reviewed on the health indicators construction for vibration-based bearing and gears using mechanical signals
[16]	2018	Provided an overview of architectures and theories of artificial intelligence-based prognostics approaches with plausible advantages and limitations
[17,18]	2019	Surveyed the deep learning based prognostics approaches and their applications
[19,20]	2019	Reviewed the recent advancement in the field of prognostics and summarized them into three categories, i.e., the data-driven, physics-based, and hybrid prognostics.
[21]	2020	Provides a review on diagnostics and prognostics approaches focusing on the application of machine tools considering the feed axis, spindle speed, and hydraulic system.

driven models can be classified into three categories: statistical-based models, conventional machine learning based models, and deep learning based models. In this section, we report a systematic overview of these three categories. The structure of this section is summarized in Fig. 1.

2.1. Statistical based models

Typically, a statistical based model for the RUL estimation is constructed via fitting a probabilistic model to data without relying on any physics or engineering principle. Two broad categories of statistical based models are general path models (GPMs) and stochastic process models (SPMs). In the following subsections, a brief review of each type of these models is provided.

2.1.1. General path model

The basic idea of a GPM, which was first introduced by Lu and Meeker in 1993 [23], is to find an appropriate parametric regression model to capture the degradation trend over time. The general path model allows the direct use of degradation data and captures the unit-wise fluctuation in degradation data. For any given time t, the degradation path of unit i is defined as

$$y_i(t) = \eta(t, \varphi, \theta_t) + \varepsilon_i \tag{1}$$

where φ is a vector of fixed (population) effects for all units, θ_t is a vector of random (individual) effects for the i^{th} unit, and ε_i $N(0, \sigma^2)$ is the normally distributed measurement errors. This model relies on three basic assumptions. Firstly, the degradation data should be captured using any appropriate failure models and mapped to a function of $\eta(\bullet)$. Secondly, the historical data should be collected under similar situations considering a reasonable variation of each individual component. Finally, there exists some defined critical level of degradation, termed as a soft failure, which indicates component failure. Due to its simplicity and ease of implementation, GPM has been well-studied, and various extensions have been developed based on the basic form in Equation (1) [12,24,25]. First, the functional form for the $\eta(\bullet)$could be linear, quadratic, exponential, etc [26]. If the degradation signals are complex and show nonlinear shapes, two or multiple phases, more generally, nonparametric regression forms can be assumed [27,28]. Second, a variety of distributions for the parameters in the $\eta(\bullet)$can be considered, such as Weibull, normal, lognormal, etc. Third, error items that capture the produce and environment noises can be assumed independently and identically distributed or correlated among different time points. To predict the RUL for working units at an individual level, the parameters in the model need to be estimated at the offline stage and updated at the online stage when new observations are available. For the offline parameter estimation, the empirical two-stage method [29], maximum likelihood estimation (MLE) and expectation–maximization (EM) algorithm provide reliable estimates. For the online parameters updating, Bayesian framework is the most natural way, where posterior distributions of the model parameters are generated based on newly collected data.

2.1.2. Stochastic process model

There are four SPMs in the literature which are commonly used for RUL prediction, namely, Wiener process, Gamma process, Gaussian process and inverse Gaussian process models. A brief description of these four models is illustrated in the following paragraphs.

2.1.2.1. Wiener process. In the stochastic process based model, for any time t, and $\Delta t > 0$, the increments $\Delta X(t) = X(t+\Delta t) - X(t)$ of degradation signal $X(t)$ in disjoint time intervals are independent. For a Wiener process, $\Delta X(t)$ is normally distributed. If we use a wiener process to describe a degradation trend, The basic form can be written as

$$X(t) = \lambda t + \sigma B(t) \tag{2}$$

where λ is a drift parameter reflecting the degradation rate, σ is a diffusion coefficient, $B(t)$ represents a standard Brownian motion. So then $\Delta X(t)$ is a normal distribution with $\Delta X(t)$ $N\big(\beta[\Lambda(t+\Delta t) - \Lambda(t)],$ $\sigma^2[\Lambda(t+\Delta t) - \Lambda(t)]\big)$ based on the property of the Wiener process. As a degradation model, the Wiener process has some unique advantages. A dominant advantage is that the distribution of the failure time can be formulated analytically by the first passage time (FPT), in which its probability density function (PDF) follows an inverse Gaussian distribution, namely

$$f_{IG}(y; a, b) = \left(\frac{b}{2\pi y^3}\right)^{\frac{1}{2}} exp\left[-\frac{b(y-a)^2}{2a^2 y}\right], y > 0 \tag{3}$$

where b is the mean and a is the shape parameter. Due to its mathe-

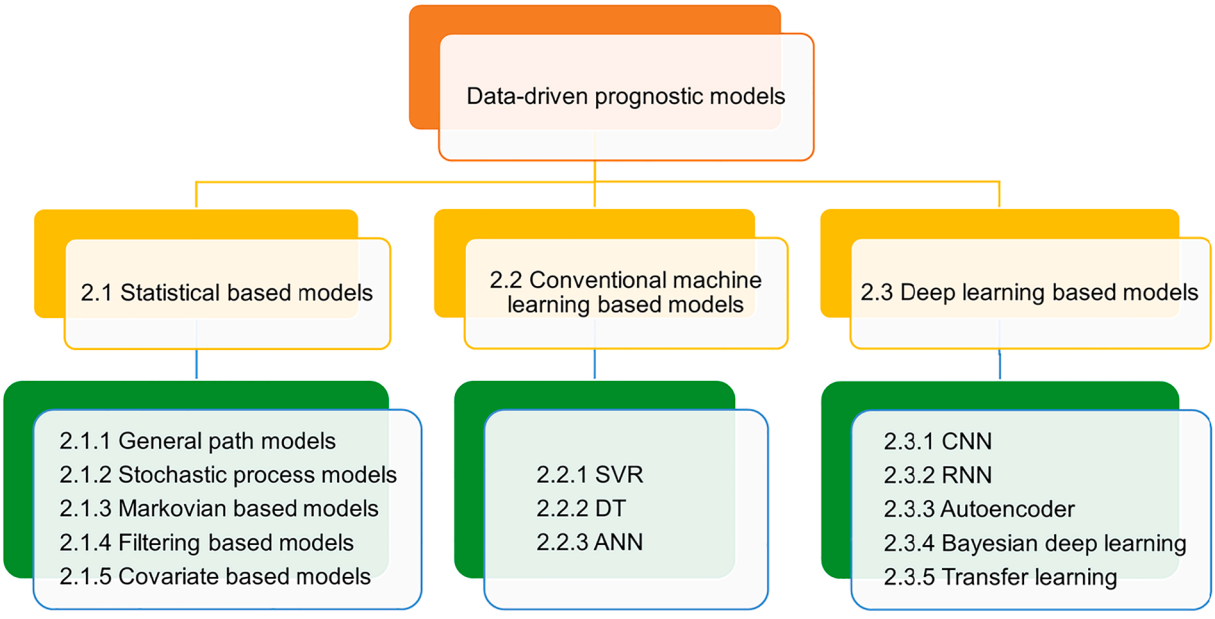

Fig. 1. Categories of data-driven prognostic models.

matical properties and physical interpretations, the Wiener process can be easily extended to satisfy different demands. One alternative is to add an error term into the basic process to capture measurement errors in degradation signals [30]; the second way is to incorporate random-effects model in dealing with unobserved heterogeneities, specifically, assume that λ or σ or both follow some certain parametric distributions, see examples [31–33] among others. The third approach is to incorporate nonlinear structure into this model to make the model more general. In particular, the more generalized model is defined as

$$X(t) = \lambda\Lambda(t;\theta) + \sigma B(\tau(t;\gamma)) \tag{4}$$

where $\Lambda(t;\theta)$ and $\tau(t;\gamma)$ are non-decreasing functions with parameter vectors of θ and γ [34]. Wiener processes have attracted significant attention in modeling several degradation trends encountered in real systems, such as bridge beams [31], fatigue crack dynamics [35], light-emitting diodes [36], thrust ball bearings [37], and micro electro mechanical systems (MEMS) [38]. Zhang [39] provided a comprehensive review of Wiener process methods with application to RUL prediction.

2.1.2.2. Gamma process. One of the distinct features of the Wiener process is that it is a non-monotone stochastic process. However, it might not be suitable in many degradation applications that show apparent monotone trends. Gamma process is an alternative in this regard. If the increment $\Delta X(t)$ follows Gamma distribution, the process $X(t)$ is called the Gamma process. The Gamma process is proved to be an efficient tool in the stochastic modeling of monotonic and gradual degradation in a sequence of small increments, such as fatigue, wear, consumption, creep, corrosion, erosion, swell, crack growth, and so forth [40]. However, Gamma process models have the following shortcomings. First, gamma process models are constrained by the assumption of Markov property. Second, gamma process models are only effective in describing the monotonic degradation processes [8]. A survey of the application of gamma processes in degradation modeling can be found in [41].

2.1.2.3. Gaussian process. The Gaussian process is another emerging approach in the field of prognostics. A Gaussian is defined mathematically as

$$f(x) = GP(m(x), k(x,x')) \tag{5}$$

where $m(x)$ and $k(x,x')$ are the mean and covariance functions respectively, denoted by

$$\begin{aligned} m(x) &= E(f(x)) \\ k(x,x') &= E\left[(f(x) - m(x))(f(x') - m(x'))^T\right] \end{aligned} \tag{6}$$

Gaussian process regression is a way to undertake non-parametric regression with Gaussian processes. The idea is that Gaussian process regression uses conditioning on Gaussian vectors to find a model that actually passes through the data points. Unlike classical regression models, Gaussian process regression does not force an analytical formula for the predictor, but a covariance structure for the outcomes. To accurately reflect the correlations presented in the data, the covariance functions need to be specified, and the hyperparameter values of the covariance function need to be optimized. Due to the probabilistic nature of the Gaussian process models, the classic model optimization approach where model parameters are optimized through the minimization of a cost function such as mean square error is not readily applicable. A probabilistic approach to the optimization of the model, such as the maximum likelihood method, is more appropriate. Some examples of Gaussian process regression applied to RUL prognostics can be found in [42–45].

2.1.2.4. Inverse Gaussian process. The inverse Gaussian (IG) process is another natural choice for degradation data which provides a monotone degradation path. If the increment $\Delta X(t)$ follows *IG* distribution, the process $X(t)$ is called inverse Gaussian process. The pdf of an IG distribution is defined as

$$f(x;a,b) = \sqrt{\frac{b}{2\pi}x^{-\frac{3}{2}}\exp\left(-\frac{b(x-a)^2}{2a^2x}\right)}, x > 0 \tag{7}$$

Let $T = \inf\{t = Y(t) = D\}$ denotes the failure time. The Failure time distribution is obtained by

$$\begin{aligned} P(T < t) &= P(Y(t) > D) \\ &= \Phi\left[\sqrt{\frac{\eta}{D}}(\Lambda(t) - D)\right] - e^{2\eta\Lambda(t)}\Phi\left[\sqrt{-\frac{\eta}{D}}(\Lambda(t) + D)\right] \end{aligned} \tag{8}$$

where Φ is the standard normal cumulative distribution function (CDF). Ye et al. [46] first justified its physical meaning by exploring the inherent relations between the *IG* process and the compound Poisson process.

To summarize, SPMs are more favorable than GPMs to account for the randomness in degradation processes caused by both inherent and environmental factors when a significant fluctuation exists in the data. However, compared to GPMs, SPMs are often complex and require a more in-depth statistical and computational ability for the model parameter estimation.

2.1.3. Markovian-based model

Define a set of states $S = \{s_1, s_2, \cdots, s_r\}$, the Markov process is a process that starts in one of these states and moves successively from one state to another. Although the Markov process still belongs to stochastic processes, this model is distinguished from the above stochastic models. Markov process assumes a finite state of the degradation and the task is to find the transition probability among those states. The main property of the Markov process is being memoryless, which states that the future degradation state only relies on the current degradation state. RUL estimation using Markovian-based models can be captured by computing the amount of time that the process will take to transit from the current state to the absorbing state for the first time [12]. In real-world applications, however, the transition probabilities may also be related to other variables, e.g., the level of degradation, the time when the product reached the current state, etc. Semi-Markovian models extend the application of Markovian-based models by incorporating the effects of these factors. In practice, the actual degradation level is not accessible due to the complexity of degradation process or the random nature of the equipment. Hidden Markov Models (HMM) and Hidden Semi-Markov models (HSMM) [47] can be used to solve this issue. In HMM, the state of the hidden process can be inferred by the observation sequences, each state is described by probability density distribution, and each observation vector is generated by the state of the corresponding probability density distribution. HSMM, as a generalization of HMM, can reflect gradual changes because of the semi-markovian assumption. Markovian-based models, which are known for exact and approximate learning and inference, have a strong statistical foundation and have been well studied. Due to their Markovian nature, they do not take into account the sequence of states leading into any given state.

2.1.4. Filtering-based model

The Kalman filtering model and Particle filters (PFs) are the most popular filtering-based degradation models. In the Kalman filtering model, the unobserved degradation x_t and the observed degradation signal y_t has the relationship that $x_t = \alpha x_{t-1} + \epsilon_t$ and $y_t = \beta x_t + \eta_t$, where ϵ_t and η_t are Gaussian noises, α and β are the parameters of the state-space model. Unlike Markovian-based models that only depend on the last degradation signal, the Kalman filtering model takes advantage of all historical data. However, the Kalman filtering model is constrained by linear assumption and Gaussian noise assumption. To overcome the drawback, PFs are particularly useful for linear/nonlinear Gaussian/non-Gaussian state-space models. The process of a PF can be expressed

by the state transition function f and the measurement function h:

$$\begin{aligned} x_k &= f(x_{k-1}, \theta_k, \nu_k) \\ z_k &= h(x_k, \omega_k) \end{aligned} \tag{9}$$

where $k$is the time step, x_k is unobserved degradation, θ_k is a vector of model parameters, z_k is observed degradation signal, ν_k and $\omega_k$are process and measurement noise, respectively. The posterior $p(x_k|z_{1:k})$ can be updated recursively using Bayesian inference based on up-to-date observations. Once the degradation model is updated, the future degradation magnitudes and RUL can be predicted based on the updated model. Comparing with the Kalman filter, PF is more elastic as it does not assume linearity and Gaussian nature of noise in data. Both filters start with a state-space representation of the stochastic processes of interest. They are robust and scale well in many applications but at the price of high computational cost.

2.1.5. Covariate based model

The risk factors that cause the degradation process are called covariates. One of the most popular covariate-based models is Cox Proportional Hazard (PH) model, which was proposed by Cox in 1972 [48]. The Cox PH model allows to describe the survival time/RUL as a function of multiple prognostic factors. The basic format of the Cox PH model is defined as,

$$h(t;\boldsymbol{z}) = h_0(t)exp(\boldsymbol{\beta z}) \tag{10}$$

where $h_0(t)$ is the baseline hazard rate function, which can be either nonparametric or parametric, so the model is often called as a semi-parametric approach. z is a vector of the corresponding covariates/prognostic factors. The covariate z is associated with the system. $\boldsymbol{\beta}$ is the unknown parameter of the model, which is called regression coefficient, defining the effects of the covariates. With the hazard function in Equation (10), the pdf of the failure time can be defined as

$$f(t) = h(t)^{\delta}\mathrm{S}(t) = \mathrm{h}(t)^{\delta}exp\left(\int_0^t -h(u)du\right) \tag{11}$$

where $S(t) = exp\left(\int_0^t -h(u)du\right)$is the survival function. δ denotes an indicator function taking value 1 if the system is failed at t, or taking value 0 if it is censored. Note that censored means that the equipment doesn't have a failure event. The Cox PH model is frequently used in medical statistics and has been extended to the manufacturing field in recent years. From the Equation (10) we can see that if z is extended to $z(t)$, the degradation signals can be easily incorporated into the equation by treating degradation data as a time-varying covariate. It is beneficial in reliability analysis for hard failure systems, where each equipment runs to fail, so that different unit has different degradation level. The functional form for $z(t)$ can be general path model [49], Wiener process [50,51], multivariate Gaussian convolution process [52]. The Cox PH model, a semi-parametric approach, is more robust than other parametric approaches as it is not vulnerable to misspecification of the baseline hazard. But the proportional hazard assumption may limit its application to accounting for complex relationships among covariates. Recently, deep learning based Cox models have been implemented to relax the proportional hazard assumption [53,54].

2.2. Conventional machine learning based models

Though machine learning has been around for several decades, it has seen a revival in recent years due to the dominance of data stemming from the information explosion. In the following subsection, the various machine learning algorithms have been reviewed and discussed from the predictive maintenance perspective.

2.2.1. Support vector machine

Support Vector Machine (SVM) was initially established as a methodology to be used in the binary classification problem and then is applied to solve the regression problem. When it is applied to a regression problem it is termed as Support Vector Regression (SVR). The key purpose for SVR is to get a functional relationship between input and output under the hypothesis that the joint distribution of the input and output is not defined and unknown. SVR uses a complex penalty function that a penalty cannot be enacted if the predicted value is farther away from the real value. The restricted region is called an insensitive tube [55]. Then support vectors are then fitted to regression models and apply to predict the degradation level and calculate the corresponding RUL values [56]. Benkedjouh et al. [57] used SVR and the isometric feature mapping reduction technique to predict the RUL for rotating machines. Hu et al. [58] built an RUL prediction method based on fuzzy C-mean clustering and wavelet SVM. Shen et al. [59] designed an SVR based on a generic multi-class solver to recognize the different faults pattern of rotating machinery. Liu et al. [60] proposed an improved probabilistic SVM regression technique to predict the condition of Nuclear Power Plant elements. A comprehensive review on SVM-based estimation of RUL can be found in reference [61]. SVM is very effective in high dimensional spaces and works well in cases where the number of dimensions is greater than the number of samples. More importantly, SVM is relatively memory efficient, which is a great advantage for online modeling. However, when the noise level is high, the performance may decrease significantly.

2.2.2. Decision tree

The Decision Tree (DT) is a non-parametric supervised technique based on a tree-like model for regression and classification. The key purpose of DT is to predict the value of an objective variable by establishing a hierarchical structure composed of nodes extracted from a training dataset. A DT generally consists of one root, several branches, and many interval nodes. Every path is from the root node to a leaf node through the internal nodes. This path denotes a classification with the different conditions of the components or systems. Every leaf node represents a response for regression or a class label for classification. To extend the power of DT, some variants have been developed, such as gradient boosting decision tree (GBDT), random forest (RF). The RF is a term for a collaborative approach of DT, which consists of numerous trees. Unlike classical methods that build a single tree on a whole dataset, RF randomly chooses the features and instances to build multiple trees. Each DT then votes for a particular target class and a class having the bulk votes is the model's prediction with a certain probability. In contrast to a traditional DT, RF demonstrates good predictive performance with considerable noise made by random selection of instances and features. Furthermore, it can deal with large datasets having numerous features with diverse data types, e.g., continuous or categorical values. Kundu et al. [62] presented an RF regression methodology for RUL prediction for spur gears depending on pitting failure mode. GBDT is an iteratively accumulative decision tree method. The algorithm accumulates the results of multiple decision trees as the final prediction output by creating a group of weak learners. Wang et al. [63] developed a GBDT model to estimate the RUL by choosing fault features and measuring fault severity subjected to relative entropy distance in fault prediction of electronic circuits. DT requires less effort for data preparation during pre-processing, and it is very intuitive and easy to explain. But we need to be careful that a slight change in the data can cause a significant change in the structure of the decision tree, which makes the model instable. The calculation of the tree can go far more complex compared to other algorithms.

2.2.3. Back propagation neural network

Back Propagation Neural Network (BPNN) is a supervised-learning method implemented by iterative optimization to solve the classification or regression problem. Usually, it takes a vector as the input, and outputs is a label representing the information of corresponding classes or function value. It firstly calculates the model results through the

forward propagation step and then tunes the network's weights through the back propagation step. The two steps above can be executed iteratively until the errors between the model results and the label reduce to a desired threshold. The ultimate target of the BPNN is to get the network parameters representing the relation between the input and output by minimizing a corresponding loss function. BPNN usually used the squared error sum (SES) for the network as an objective function and applied the gradient descent technique to get the objective function's minimum value. The BPNN, just like other NNs, is flexible and powerful to find the nonlinear mapping between inputs and outputs, and it doesn't require prior knowledge about the network. But back propagation is notorious for the easily getting stuck in "local minima".

2.3. Deep learning based model

In recent years, deep learning approaches have shown excellent performance in various applications ranging from feature extraction, defect detection, segmentation, medical imaging, additive manufacturing, and many more [64–70]. Realizing the promising ability, researchers have experimented on various deep architectures to develop the solution approach in remaining useful life prediction. In the following subsection, we discussed the architecture of deep Convolutional Neural Networks (CNNs) and their variants from the predictive maintenance point of view.

2.3.1. Convolutional neural network

Due to the ability to generalize the local and global features, CNNs turn out to be the most popular deep learning methods. CNNs are exceptionally successful in extracting features from input data and using them to make a trustworthy prediction. A basic CNN structure mainly has an input layer, convolution layer, pooling layer, and fully connected layer as shown in Fig. 2.

The input data could be either two-dimensional or one-dimensional such as time–frequency spectrum or time series data, respectively. The convolution layer uses a set of weights and convolutes at each layer to form the layer-wise features, which are called a feature map. The output of the convolutional layer is calculated as: $Y_n = f(X^* W_n + b_n)$, where * represents an operator of the convolution, n denotes the number of convolution filters, W_n is the weight matrix, b_n is the filter kernel bias. Following the convolution, the model parameters are reduced by subsampling, named as pooling process. After the pooling layer, multiple fully connected layers are used to convert the matrix to a row or a column. Finally, a classification or regression layer is added to get the predictions or results. To predict the RUL, CNN can be used to extract useful and robust features from data. The number of processing units or the CNN structure greatly depends on the nature of problems and datasets. To expand the power of CNN, several variants of CNN-based models have been introduced in the literature as reported in Table 2. The key benefit of using CNNs is to extract complex, non-linear, non-

Table 2
Variants of CNNs and their use for RUL prediction.

Variants	Distinctions	References
Deep CNN	• It consists of different processing units at multiple layers (usually have 5 to 10 layers, even more) • Effective in capturing the salient patterns in the signals	• Babu et al. [72] combined a regressor with a deep CNN architecture to estimate the RUL from multivariate time series data. • Ren et al. [73] fused a smoothing method with a CNN built a CNN for predicting the bearing RUL.
Deep Multi CNN	• Multiple CNN architectures are stacked together • The output of the previous CNN becomes input of other CNNs • Effective in dealing with raw signals, instead of depending on the feature extractor	• Yang et al. [74] proposed a double-CNN model architecture to predict RUL using original vibration signals without resorting to any feature extractor.
Deep Multi-scale CNN (MSCNN)	• MSCNN framework has three sequential stages: transformation, local convolution, and full convolution. • The transformation stage applies transformations on the input time series. • In the local convolution stage, extract the features for each branch. • The full convolution stage concatenates all extracted features and applies several more convolutional layers to generate the final output. • Effective to keep the multiple levels of abstraction for the prediction	• Kiranyaz et al. [75] utilized the MSCNN for fault detection and identification for a circuit monitoring system • Zhu et al. [76] used the wavelet transform to propose Time Frequency Representation (TFR), then applied this TFR to MSCNN to perform RUL estimation. • Li et al. [77] used MSCNN for RUL prediction, the model has three multi-scale blocks, where three different sizes of convolution operations are put on each block in parallel.
Hybrid CNN (HCNN)	• This CNN architecture is mainly the combination of above-mentioned CNN along with additional supporting layers. • Incorporates the advantages of different methodologies by their integration to improve the prediction performance.	• Wen et. al [85] proposed a new residual CNN (ResCNN) by adding a skip connection between convolution blocks

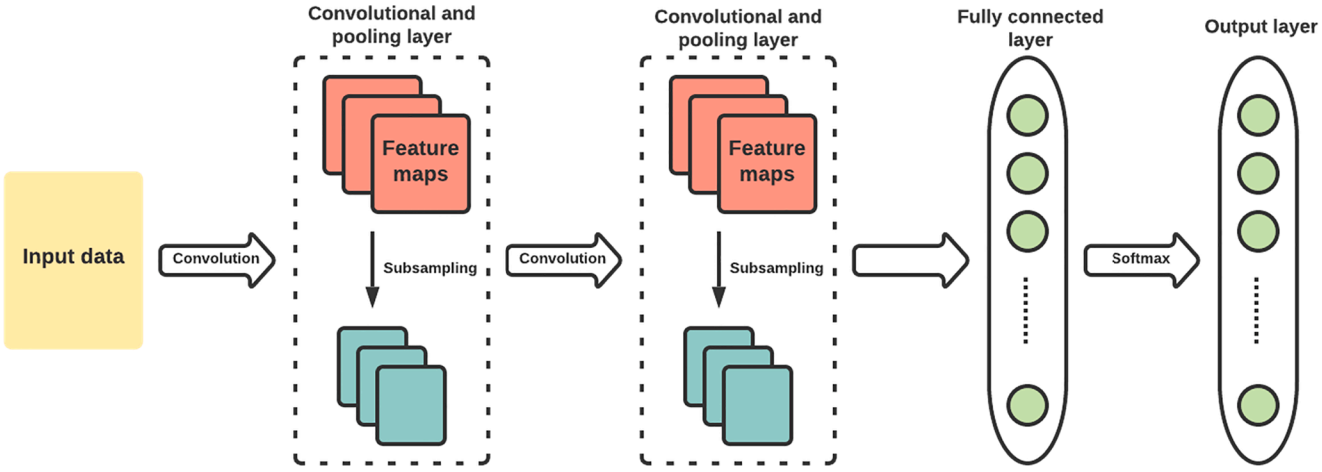

Fig. 2. Architecture of CNN.

handcrafted features because of the superior feature extraction and object recognition performances of CNN. Based on our observations, only a few research works focus on pure CNNs for RUL prediction as listed in Table 2. The reason is that, CNNs may not sufficiently model the temporal characteristics of time series data. Moreover, CNN is significantly slower due to a convolutional operation and requires a lot of data to train effectively. Also, the tuning to find the proper learning rate for the CNN methods on real-world applications is difficult [71].

2.3.2. Recurrent neural network

The underlying motivation behind Recurrent Neural Network (RNN) is to mine the sequential information for any given dataset. It creates memory cells that capture the past and predict the future sequence based on the previous computation. A typical RNN is shown in Fig. 3. As shown in Fig. 3, the structure of the RNN constitutes a deep network with one layer per time step and shares the parameters across the layers. The concept of parameter sharing is a useful way to capture the relationship between one input item and its neighboring context. This makes the RNNs very successful over the traditional NNs and CNNs. The network can be trained in a similar fashion of backpropagation across the time steps. However, the training process is especially challenging due to the problem of gradients vanishing or exploding. To overcome this issue, long short-term memory networks (LSTM) is constructed [78]. LSTMs are a special kind of RNN for remembering information for long periods (long-term dependencies) and are explicitly designed to avoid the problem of standard RNN. Similar to standard RNNs, LSTMs also possess chain-like structures, but they differ in the structure of the memory cell. Instead of having a single neural network layer, LSTMs have four interacting networks connected in a very tricky way to remove or add information to the cell state by regulating the structure of different gates. Gates indicate a special setup to control the information passing to the cell state and output at each repeating module. Gated Recurrent Unit (GRU) is another modified LSTM cell, which was introduced by Cho et al. [79]. Recently, this architecture showed its promising application in the field of RUL prediction [80–82]. GRU combines the input gate and the forget gate into the update gate. It also merges with cellular sate and hidden state. A comparison of the memory cell in standard RNN, LSTM and GRU LSTM is shown in Fig. 4.

The horizontal line on the top of the repeating module indicates the cell state, which passes through the entire network chain with some minor liner interactions. They are comprised of a sigmoid neural net layer and a pointwise multiplication and addition operation. The sigmoid layer maps numbers between zero and one, where zero means no information will pass through and the value of one allows all information. Recently many researchers have introduced several variations based on the original RNN and LSTM. Among them, some major variants are Gated Recurrent Unit (GRU) RNN, Bi-Directional LSTM (BiLSTM), and Bi-Directional Handshaking LSTM (BHLSTM). Table 3 summarizes the major distinction for these variants and includes some representative publications for each variant as references. Several studies have proven that RNNs and LSTMs are outstanding and perform better than many conventional machine learning approaches and even better than CNNs for the RUL prediction tasks [92]. However, as some researchers pointed out that the LSTM network may not be robust when processing raw time series data directly since the sensor data usually contains noise [93].

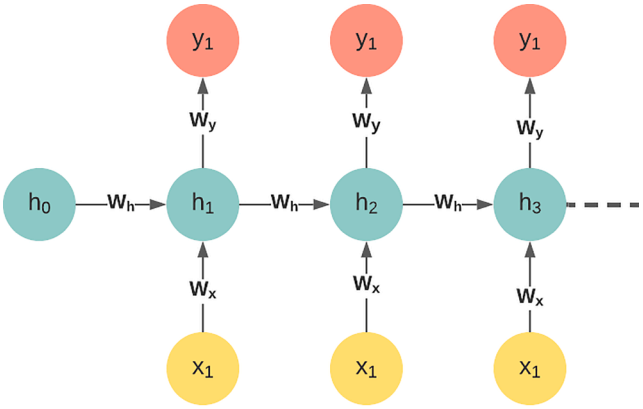

Fig. 3. Structure of a typical RNN.

2.3.3. Autoencoder

An autoencoder is an unsupervised neural network. The main idea of autoencoder is to train the model to reconstruct the original input at the output layer [94]. The autoencoder network consists of three layers: an input layer, a hidden layer for encoding, and an output layer for decoding as shown in Fig. 5. The size of the hidden neurons is usually smaller than the input. In this way, the network is forced to learn a compressed representation of the input. Define the input as $\boldsymbol{x}$, an encoder function $g(.)$ parameterized by $\boldsymbol{\phi}$, a decoder function $f(.)$ parameterized by $\boldsymbol{\theta}$, the output as $\boldsymbol{x}$', then the reconstructed input is $\boldsymbol{x}' = f_\theta(g_\phi(\boldsymbol{x}))$. The parameters $(\boldsymbol{\theta},\boldsymbol{\phi})$are trained to minimize the reconstruction error so that the output is similar to the original input, i.e., $\boldsymbol{x} \approx \boldsymbol{x}'$. Several variants have been developed to solve different issues based on the basic mode. If the number of network parameters is larger than the number of inputs, the basic autoencoder will face overfitting issues. To avoid overfitting and improve the robustness, denoising autoencoder is developed by randomly changing some of the input values to zero, then loss function is designed in a way that the output values are compared with the tampered input instead of original input. Another tactic to control the number of hidden nodes is sparse autoencoder, where a sparse constraint is added to limit the activation of its nodes. With sparse constraint, some nodes in the hidden layer are active and the other nodes are inactive. This constraint is achieved by adding a penalty term into the loss function. The third variant is called variational autoencoder [88]. The basic idea is that instead of mapping an input to a fixed vector, input is mapped to a distribution. When decoding from the layer, samples from each distribution are randomly selected to generate a vector. Variational autoencoder provides a probabilistic manner for describing the input. In the predictive maintenance field, autoencoders currently are mainly used to reduce the dimension and eliminate the redundancy of the data. Autoencoders are commonly employed as a feature extractor or a tool for health index construction. Some examples can be found in [95–99]. To learn sequential information from input signals, LSTM autoencoder is a preferable choice. As an autoencoder learns to capture as much information as possible rather than as much relevant information as possible, it may misunderstand important variables of the input data.

2.3.4. Bayesian deep learning

Bayesian Deep learning method is an extension of deep learning in a probabilistic manner. The fundamental idea is to adopt Bayesian inference as the learning tool for quantifying uncertainty of the model by treating the deep learning architectures as probabilistic models. Typically, a Bayesian Deep learning places prior distributions over the network's weights and then learn the corresponding posterior distributions over the weights. Then each forward pass will have different weights and therefore providing potentially different outputs. As exact Bayesian inference is computationally intractable for most of the NN structures, some sampling strategies are used to learn the parameters of Bayesian deep learning models, which is often computationally expensive. From the literature review, we observed that some popular approximations have been proven effective to alleviate the computational burden, such as variational inference, expectation propagation, Laplace approximation, Hamiltonian methods, bootstrapping, Monte Carlo dropout, etc. Among these, Monte Carlo dropout, which combines approximate Bayesian inference with dropout, has drawn considerable attention in many research fields due to its simplicity, scalability, and computational efficiency. It is well known that data-driven models inevitably face two

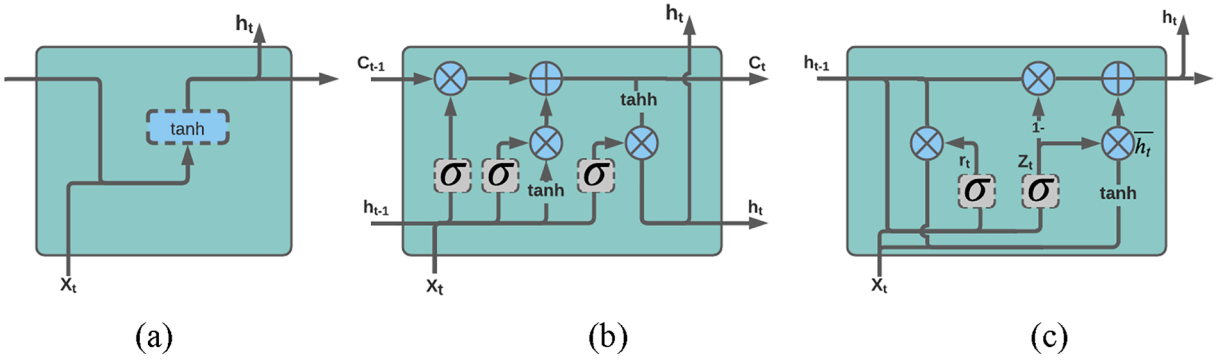

Fig. 4. Memory cell structure, (a) standard RNN; (b) LSTM; (c) GRU LSTM.

types of uncertainties: aleatoric uncertainty, reflecting the noise pollution in data collection and transmission, and epistemic uncertainty, reflecting the ignorance of model property. To address these issues, a variety of researchers have been exploring Bayesian Deep learning to account for the uncertainties to improve model accuracy. References [100,101] use Bayesian deep learning based methods to quantify the uncertainty of point prediction for bearings and Gas turbine engine. Li et al. [102] proposed a Bayesian deep learning based methods, in which a sequential Bayesian boosting algorithm was executed to improve the prediction accuracy. A Bayesian deep learning model can be treated as an ensemble of multiple models, which may naturally reduce the risks of over-fitting issues. Another benefit of Bayesian deep learning models is that they allow to quantify the uncertainty, which is very important for RUL prediction considering the limited data availability and the stochastic nature of degradation processes in the manufacturing field. But the computational cost is heavier for online inference comparing with other deep learning models as the model needs to run multiple times to get the distribution of outputs.

2.3.5. *Transfer learning*

Deep learning models excel at learning from a large number of labeled examples, learn a very accurate mapping from the inputs to outputs. But it lacks capability to generalize to different application scenarios. The reason is that many machine learning algorithms assume that the training and test data are in the same feature space and have the same distribution. This assumption may not hold in many real-world applications. Transfer learning is developed to tackle this issue through storing knowledge learned from one domain (called the source domain) and transferring it to a different but related problem (called the target domain). Mathematically speaking, a domain can be defined as $D = \{\chi, P(X)\}$, where χ represents a feature space, $P(X)$ is a marginal distribution where $X = \{x_1, x_2 \cdots, x_n\} \in \chi$. A task can be defined as $T = \{y, f(x)\}$, where y is a label space, $f(\cdot)$ is a predictive function. Given a source domain D_s and learning task T_s, a target domain D_t and learning task T_t, where $D_s \neq D_t$ or $T_s \neq T_t$. Transfer learning aims to help improve the learning of the target predictive distribution based on D_s and T_s. Based on what to transfer, transfer learning can be conducted at several levels: instance-transfer, feature-representation-transfer, parameter-transfer, and rational-knowledge-transfer [103]. Instance-transfer tries to reweight some the samples from the source domain in an attempt to correct for the distribution difference, then apply them in the target domain for training. Feature-representation-transfer tries to get good feature representations that can reduce the difference between the source domain and the target domain. Parameter-transfer discovers shared parameters or prior knowledge between the source domain and the target domain. Parameter-transfer models believe that a well-trained model on the source domain has learned a well-defined structure, and if two tasks are related, this structure can be transferred to the target model. Relational-knowledge-transfer works by mapping some similar patterns from the inputs to the outputs between both domains. Examples for RUL prediction based on transfer learning can be found from [89,98,104–106]. Transfer learning is especially useful in the situation where source data and target data are in different feature spaces or have different distributions in which training data for the target problem are limited but data for a related problem are abundant. Transfer learning provides an effective way for RUL prediction with limited historical failure data. However, transfer learning only performs better under the condition that the domain and target problems of both models are similar enough. Otherwise, it will end up with a negative transfer. Currently, it is still challenging to find solutions to negative transfer.

3. Applications in predictive maintenance

In this section, we first provide an overview of the workflow of the predictive maintenance and how the prognostic approaches are applied for predictive maintenance. Then we share some statistics and give a comprehensive review of applying RUL predictions to various fields.

Fig. 6 summarizes the workflow of the predictive maintenance. First, data are collected intermittently or continuously from an interactive physical system of interest. Various sensors are installed to collect the degradation signals in semi-observable or fully online systems. Some commonly used sensors are pressure sensors, force sensors, speed sensors, temperature sensors, torque sensors, proximity probes, accelerometers, etc. Apart from the sensor data, quantitative data are sometimes collected based on the purpose and application domain. For example, the RUL of many power storage systems depends on the charging and discharging cycles. In such cases, a number of cycles are recorded to collect the data. Following the data collection and processing, feature extraction plays a critical role in model development and RUL prediction. The straightforward use of raw data is inconvenient due to the high complexity and nonlinearity of sensor signals. Hence, the underlying motivation of feature extraction is to utilize the patterns and trends in the sensor signals to predict RUL. Literature in RUL prediction has focused on many time-domain and frequency-domain features such as root mean square, kurtosis, short-time Fourier transform [107], wavelet transform [108], empirical mode decomposition [109]. Recently, a number of machine learning based approaches have been utilized to learn the learn the features and mapping the raw signals to the associated RUL. Several such machine learning algorithms are described in Section 2.2. These machine learning algorithms are that they are capable to extract the useful features and information within the data with a very limited human intervention. Section 2.3 demonstrates several deep learning architectures, which are emerging and highly effective techniques for patterns and trends recognition. Their deep networks are capable of obtaining high-level abstractions of data to improve the performance in intelligent prognostics. The best part of these deep learning architectures is that they are capable of feature extraction without human intervention. The extracted features are then used as inputs of developed RUL prediction model. Once the predicted RUL is obtained, the last step is to decide the optimal time to send out an

Table 3
Variants of RNNs and their use for RUL prediction.

Variants	Distinctions	References
Recurrent Neural Network (RNN)	• Can utilize the sequential information • Able to retain the short-term information • Able to capture the temporal correlations in sequence data	• Heimes [83] utilized the RNN incorporating with Kalman filtering training and evolutionary algorithm for prognostics problem. • Liu et al. [84] proposed an adaptive RNN for dynamic state forecasting to leverage the RUL prediction for Lithium-ion-batteries. • Liang et al. [85] proposed a RNN based health indicator for RUL prediction of bearings.
Long Short-Term Memory (LSTM)	• Solve the vanishing gradient or exploding problem • Able to retain both Long and short term information • It is easy to detect and capture important features over a long distance	• Zhang et al. [86] employed LSTM RNN to learn the long-term dependencies among the degraded capacities and construct a RUL predictor. • Zhao et al. [87] developed an RUL predictor using the LSTM, which can evaluate the trend features. • Xiang et. al. [88] introduced a new type of LSTM with weight amplification for accurate prediction of gear remaining life.
Gated Recurrent Unit (GRU) RNN/LSTM	• Can deal with long term relationship in RNN • Able to capture dependencies at different time scales • Able to capture the inherent relation for long-term prediction	• Song et al. [80] proposed a battery RUL prediction approach based on the RNN with gated recurrent unit. • Chen et al. [81] incorporated kernel principal component analysis and GRU RNN to predict RUL. • Wang et al. [82] presented a hybrid RUL prediction model by adapting a deep heterogeneous GRU model.
Bi-directional Long Short-Term Memory (BLSTM)	• Utilize the information in both forward and backward direction • Suitable for intermediate prediction	• Zhang et al. [89] proposed a transfer learning based BLSTM network for turbofan engine RUL prediction. • Wang et al. [90] proposed a data-driven approach with BLSTM network for RUL prediction, which can make full use of sensor data sequence in bidirectional.
Bi-directional Handshaking Long Short-Term Memory (BHLSTM)	• Can utilize two sets of LSTM cells in reverse order • Allow forward and backward unit collaboration in the learning process	• Elsheikh et al. [91] proposed the BHLSTM to predict the RUL of a system, which is capable to process maximum information for any given subset of sequence.

alert of failure to help maintenance decision making. If the predicted time is shorter than the actual time, maintenance will be implemented earlier, the benefits of more extended usage are lost. If the prediction is too late, the equipment may fail and result in a more significant loss. To determine the optimal maintenance policy, the inventory management associated with the spare parts cost also needs to be considered in practice. Timely acquisition of the inventory number and status of spare parts is challenging. Some researchers have investigated this issue by joint optimization of maintenance and inventory management [110–112]. Another exigent issue in the decision making process is that spare parts inevitably deteriorate over time due to the inner mechanism and imperfect storage conditions, which will shorten the storage lifetime and eventually affect inventory management. How to estimate the storage lifetime with arbitrary number of spare parts based on the operating and storage degradation processes is attracting researchers' attention [113,114]. Interested readers are referred to the recent articles for more details [113,114]. In this paper, we mainly focus on the RUL prediction in predictive maintenance.

Over the last decade, researchers showed intensive interest in RUL prediction for predictive maintenance and published excellent research papers in this area. However, we only focus on the recent advancement and applications within the time frame of the year 2015–2020. We reviewed 253 published papers under a broad point of view of data-driven approaches such as statistical approaches, deep learning, conventional artificial intelligence and hybrid approaches (combination of statistical and AI-based methods). The papers were collected from Google scholar using a thorough key word search of "remaining useful life (RUL)", "predictive maintenance (PM)", "prognostic and health management (PHM)", "health index (HI)" and "machine health". It is observed that the concept of predictive maintenance and RUL prediction are widely applied for aircraft engines, bearings, gears, motors, machine tools, wind turbines, batteries, computer hard disks, and many more. Based on our observations, we categorized all of these applications into four broad categories as "aircrafts", "rotating machinery", "power systems" and "electronic systems". We also break down the proportion of published papers in these four categories as shown in Fig. 7. Surprisingly, most of the researchers showed their interest in the field of rotating machineries (48%) and aircraft system (23%) applications. One potential reason for this is the availability of standard datasets in these two fields. We found that the NASA C-MAPSS, PRONOSTIA, and Center of Intelligence at the University of Cincinnati dataset are widely used in this area. However, some researchers are also trying to generate the dataset from their own lab and focusing on the field of the power and electronic system components and products.

We also capture the trend (number of published papers) of different data-driven methods for the individual applications as shown in Fig. 8. We can see that the number of publications based on statistical approaches is notably large except for hard drives. Due to persuasive mathematical properties and physical interpretations, also the capabilities to capture the uncertainty of parameters, statistical approaches have been attracted widespread attention. However, some statistical models are complex and can be computationally expensive. Due to the fact that deep learning provides good functional mappings between inputs and outputs, which is powerful to capture dynamic information, the number of deep learning prognostic models is also escalating. The main disadvantage of deep learning is the operating and training processes are "black boxes" and a majority of proposed approaches are only focused on prediction on a population level, it is hard to catch the uncertainty and individual heterogeneity. It would be interesting to combine the advantages of each approach in the future. In the following subsection, we will provide details review of research of RUL prediction for those applications (i.e., aircraft, rotating machinery, electronic systems, power systems) of predictive maintenance. For each field, we first describe the popular and open access dataset, then summarize the contemporary research work in the recent five years in the corresponding field. It is worth noting that the number of papers on data-driven prognostics is enormous. Consequently, some papers would be omitted inevitably.

3.1. Application in rotating machinery

Bearing, gears, motor, and shaft are the most common rotating parts

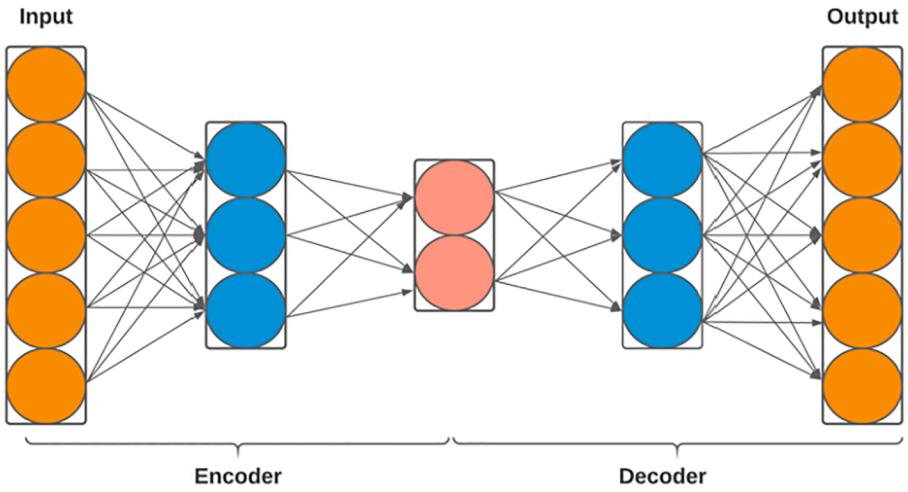

Fig. 5. Architecture of an autoencoder.

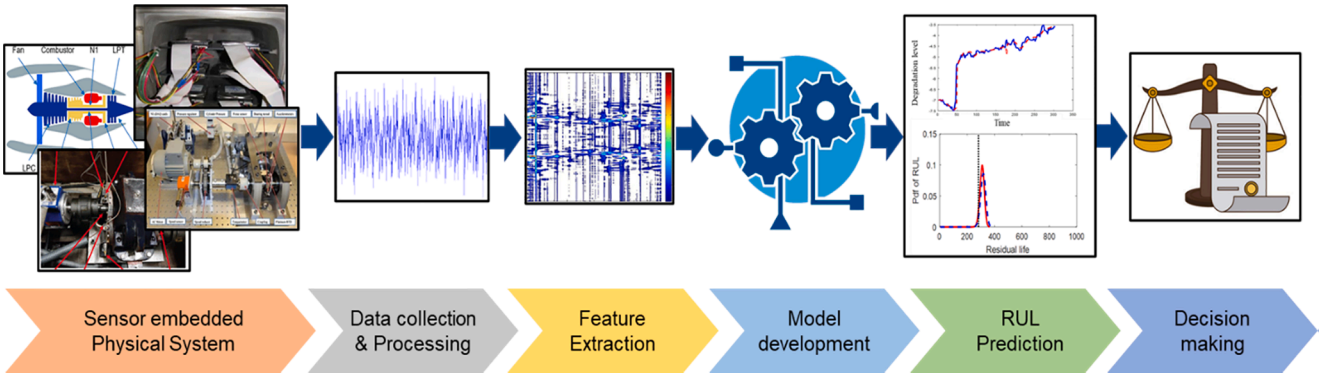

Fig. 6. Workflow of predictive maintenance.

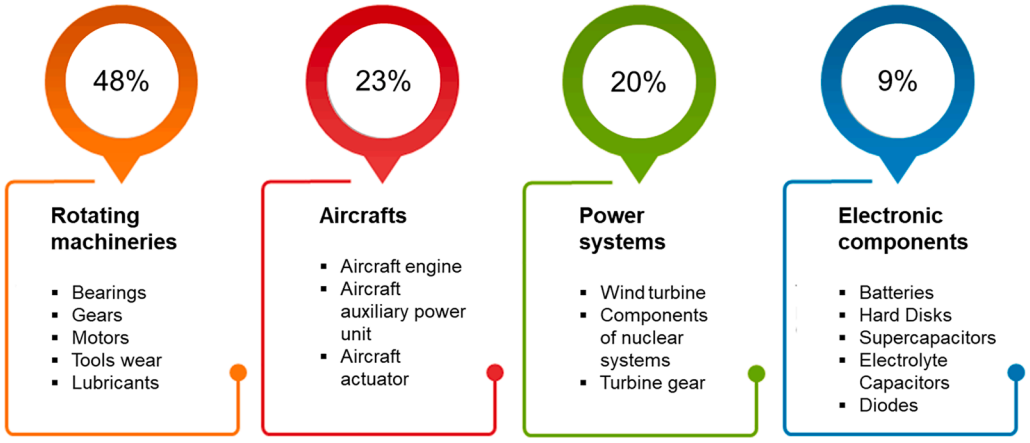

Fig. 7. List of recent application fields of predictive maintenance.

which are widely used in machinery and industry. Rotating machinery is critical for the machine heath and ensures safety as any rotating parts' flaws can lead to significant consequences and failure. The identified significant causes of bearing failures are generally subject to excessive balance, improper installations, poor lubrication practices, alignment tolerances, poor storage, and handling techniques. RUL prediction of bearings and gears is derived based on the vibration or mechanical signals obtained from the bearings. In literature, we found two experimental setups for the run-to-failure and historical data collection of bearings and gears: The Center of Intelligence at the University of Cincinnati [134] and PRONOSTIA [115]. Most of the researches in the field of bearings and gears used the datasets from these two experimental setups. Lubricant is another important substance for rotating machinery. It helps to operate machines in safe and healthy conditions. The deterioration trends can be described by observing the characteristics of lubricants and the operation conditions of the machine. Over time and usage, lubricants degrade and produce acidic substances, moisture, and insoluble deposits, such as carbon deposits, sludges [116], which lead to

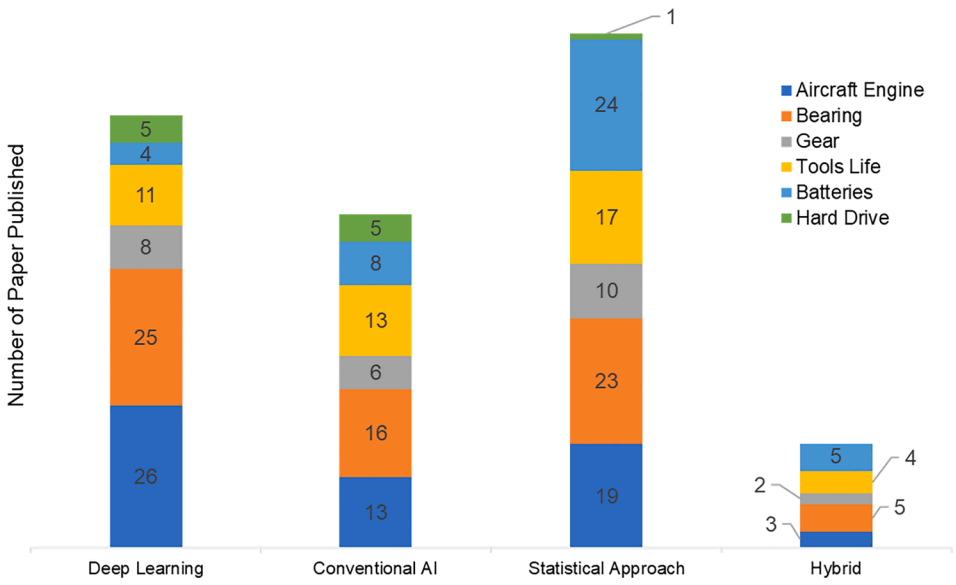

Fig. 8. Recent research trends (2015–2020) on different application fields & methods.

machine failure. Hence, it is necessary to understand the deterioration trend in advance to reduce wear and friction from the mobile components and avoid machine failure. Recently, condition monitoring of lubricating oil has attracted considerable attention in research. The lubricant data could be obtained from either a physical experiment or simulation procedure by mimicking the real environment under a certain assumption. One of the procedures of collecting lubricant data is the four-ball test found in [117]. To obtain real-time data of lubricating oil, a wear debris sensor and an oil property sensor were employed in the oil cycling line. Besides, a temperature control system was set to offer a fixed temperature for the oil property sensor. By doing this, wear conditions and dynamic viscosity and permittivity of lubricating oil can be monitored simultaneously. To accelerate oil degradation, a test composed of various loads and speeds was carried out. The test was stopped and restarted when working conditions were changed. Another lubricant data was found in [118]. The lubricating oil used for large machines was collected from the engine of a loader. In the following paragraphs, we articulate different research methodologies and approaches win this application field along with the description of the commonly used datasets in this area.

Lei et al. proposed a two-stage method based on a particle filtering algorithm to predict the bearing RUL [119]. They fused multiple features to construct new health indications and then used the maximum-likelihood estimation to initialize the model parameters. Gaussian Process Regression (GPR) is another effective method that was used in [120] with the integration of composite kernels. RMS, Kurtosis and Crest factor are used for feature fusion by self-organizing map. It is experimentally demonstrated that integration of composite kernels improves the prediction accuracy than particle filter method. Liu et al. [121] divided the entire bearing life and built an individual local regression model to get the multiple health states to leverage the RUL prediction. This constituted a semi-supervised approach that can be utilized without having any prior knowledge. Wang et al. [122] proposed a Wiener process model with stage correlation and a Bayesian approach to utilize the prior distribution information into the model parameter. Ahmad et al. [123] proposed a dynamic regression model to capture the trend of bearing health indicator, which is later used to project the future health indicator value and estimate the bearing useful life. The adaptive regression model can determine an appropriate time to start prediction which yields excellent prognostics performance. Kundu et al. [124] used a clustering and change point detection algorithm to identify the failure behaviors and predict the RUL. Sometimes bearing could be operated under multiple conditions. Integrating these multiple conditions would provide better prognosis result.

Literature also states several artificial intelligence based approaches to construct the health indicators for bearing and gear remaining life prediction. Zhao et al. [125] employed the principal component analysis (PCA) and linear discriminant analysis for dimensionality reduction. Then they used a multiple linear regression model to estimate the RUL. In some applications, it is hard to obtain the failure or suspension histories which make the RUL prediction tasks more challenging. To address this issue, Xiao et al. [126] developed an inference method using recorded condition monitoring data. An adaptive time windows was employed to divide the extracted features and to train an ANN for intelligent prognosis. Bastami et al. [127] used the wavelet packet transform to extract signal features and later trained an ANN to estimate the RUL of rolling bearings. The nonlinear nonparametric approach is also considered as a very appealing technique in predicting the RUL of bearings. The ensemble technique is another effective machine learning paradigm that could be used in RUL prediction. Such an ensemble approach, decision tree-based random forest, was proposed in [62] by Kundu et al. for monitoring and detecting the pitting progression in spur gears to predict the RUL of gear.

In recent years, deep learning becomes very effective in machine health monitoring and prognostics due to its capability of learning representation from raw data. With the development of deep learning methods, Guo et al. [85] proposed a deep neural network structure named recurrent neural network based health indicator (RNN-HI), where several classical time–frequency features are combined with the original feature set to get the most sensitive features as the input of RNN-HI model to leverage the RUL prediction of bearings. Deep learning methods can effectively extract the discriminative features for monitoring bearing fault. However, temporal information also plays a critical role in the fault degradation process, which was not considered in many cases. Mao et al. [128] first considered this temporal information in bearing RUL prediction and proposed the LSTM. In another research, Tang et al. [129] proposed an LSTM approach combining the bottleneck features to develop a novel prediction method of bearing performance degradation. In LSTM approaches, first, the feature parameters are extracted from the different domains such as time domain, frequency domain, time–frequency domain. Then the important features are extracted from the original feature set that could better represent the degradation process of bearings. Finally, the selected features are used to train the LSTM network to predict the bearing RUL. Traditionally the

feature extraction is derived from prior knowledge and is separated from the RUL models. Ren et al. [130] proposed a Multi-scale Dense Gated Recurrent Unit Network (MDGRU) to combine the feature extraction into the RUL model by pre-trained Restricted Boltzmann Machine network, multi-scale layers, skip gate recurrent unit layers, dense layers. In [131], Li et al. used a CNN to explore the time–frequency domain information and to extract multi-scale features. Deep learning approaches showed the limitation on predicting less stability for the single sensory information. To address this issue, Wu and Zhang [132] proposed a new cascade fusion convolutional long-short term memory to fuse the information streams in the form of an ensemble model. Lo et al. [133] proposed a one-dimensional CNN for the prognosis of bearing and gear. The network was trained in a hybrid fashion where both the classification loss and clustering loss were combined to estimate the status of prognosis. Xiang et al. [88] proposed an attention based LSTM named LSTM-A. This special type of network utilizes an attention mechanism to amplify the input and hidden layer weights at different degrees to accurately predict the gear remaining life.

The RUL prediction of lubricant is accomplished based on the parameters obtained from oil and degradation trends. Tanwar et al. [134] proposed a degradation model based on continuous time stochastic process, i.e., the Wiener process for lubricating oil degradation tracking and RUL prediction under regular oil top-up effects. In this research, the Oil Replenished Effect was neglected in the prediction of lubricant remaining life. Tanwar and Raghavan addressed this issue in [135] and proposed the use of the GPR model as a non-parametric Bayesian method. Recently, the machine learning approach is also being incorporated for condition monitoring from lubricant oil. In [136], the researcher used the machine learning approach to classify the engine lubricant into three conditions as normal, degraded, and unsuitable. They used a cohort of military land vehicles to collect the data from laboratory test results of lubricants and monitoring system of vehicle health. The proposed machine learning procedure used feature selection methods to identify the best feature set for representing the lubricant oil condition.

3.2. Application in aircrafts

An aircraft engine is the power component of an aircraft propulsion system. Most aircraft engines are either piston engines or gas turbines. An aircraft engine produces thrust to propel an aircraft. Aircraft engine failures may result in significant economic losses and even accidents in extreme cases. Except for the aircraft engine, another two of the most important systems in the aircraft are aircraft auxiliary power unit (APU) and actuators [114]. APU is a small turbine engine installed under the tail of an aircraft. Instead of providing propulsion, its main function is to supply power at a certain flight altitude and provide bleed air for the cabin air condition system on the ground. For some aircrafts, APU can also provide compressed air and backup electric power to compensate for the effect of dead engines. Thus, monitoring of the health state to ensure safety and operation efficiency is essential. Actuators (e.g., Electro-Mechanical Actuators (EMA) and Electro-Hydraulic Actuators (EHA)) play an active role in control systems in aircrafts. They are used to convert electrical signals to mechanical movement or other physical variables, such as pressure or temperature. In this research field, a majority of publications use the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS), which is collected from National Aeronautics and Space Administration (NASA) [137]. The C-MAPSS dataset includes four sub-datasets. All engines work in normal condition first and then degrade continuously until a failure criterion is reached. Every record of the engine state is generated by a set of 24 variables, three of which are operational settings and the other 21 are for engine performance measurements. Currently, there is no public dataset available. Most researchers investigate the performance based on data collected from commercial aircraft fleets. The research in this filed has been summarized in the following paragraphs.

Chehade et al. [138] predicted RUL through individual failure threshold distribution estimation. They developed a convex quadratic formulation that combines the historical population information and the condition monitoring data of an operating unit to online estimate its failure threshold. Some efforts have focused on developing data fusion methodologies for prognostics [139–141]. The main idea is to construct a health index via selecting and fusing multiple degradation signals to track the trajectories of the degradation process. After that, the constructed health index was treated as another sensor signal and then was used for degradation modeling and prognostics. Song and Liu [140] solved the HI construction by the quantile regression technique. Kim et al. [139] proposed a latent linear model for HI construction and a systematic sensor selection procedure for RUL prediction. Chehade et al. [141] extended the data-level fusion techniques to multiple failure mode scenarios. Using constructed HI, Li et al. [142] developed an age- and state-dependent Wiener-process model for RUL prediction with the consideration of the unit-to-unit variability. Son et al. [143] proposed a non-homogeneous gamma process based RUL prediction method. The model considered noisy degradation data and by using the Gibbs sampling technique, the hidden degradation states were approximated by using the Gibbs sampling technique. All of the above mentioned methods were developed based on statistical approaches. Researchers also investigated the machine learning approaches to develop more sophisticated methods. For example, Ordóñez Celestino et al. [144] proposed a hybrid ARIMA (auto-regressive integrated moving average)-SVM model for estimating the RUL. ARIMA model is utilized to estimate the values of the predictor variables in advance. Then, the result of ARIMA is applied as the input of a support vector regression model. Al-Dulaimi et al. [145] proposed a hybrid of LSTM and CNN framework. In their model, LSTM and CNN are constructed in parallel followed by a fully connected multilayer fusion neural network. Zheng et al. [146] proposed an LSTM approach for RUL estimation, which fully utilizes the sensor sequence information and uncovers hidden patterns. Badu et al. [72] developed a deep CNN-based regression method to predict the RUL. The convolution and pooling filters were used along the temporal dimension over the multi-channel sensor data to integrate automated feature learning from raw sensor signals. Wen et al. [147] proposed a residual convolutional neural network (ResCNN), it can help overcome vanishing or exploding gradient problem of deep learning algorithms. Song et al. [96] developed a autoencoder-BLSTM hybrid model. racy of RUL. Autoencoder was used as a feature extractor to reduce the dimension of data. BLSTM was designed to capture the bidirectional long-range dependencies. It showed that the hybrid model had better prediction performance comparing with most existing methods including CNN and LSTM.

To predict the RUL of APU, Chen et al. [148] developed a Gaussian process regression model combined with ensemble empirical mode decomposition. Liu et al.[149] utilized an Extreme learning machine (ELM) to predict the degradation of an APU. They employed a restricted Boltzmann machine (RBM) to optimize the ELM. Wang et al. [150] derived a health index to characterize the APU degradation and then used a Bayesian framework for the RUL prediction. Zhang et al. [151] utilized a Weibull-based generalized renewal process to implement failure rate prediction of APU. Researchers also put their effort for actuators prognostics. Zhang et al. [152] proposed a weighted bagging GPR algorithm. With the idea of ensemble learning, the weighted bagging GPR algorithm uses a series of subsets to train the GPR model. They found that the proposed method can take the randomness of data into consideration. Then Zhang et al. [153] proposed a feature-aided Kalman Filter method for motor voltage estimation, which is an essential parameter for performance degradation assessment of EMA. The dataset they collected through Flyable Electromechanical Actuator, which was made by NASA Ames Research Center. Guo et al. [154] presented an optimized incremental learning and on-line training algorithm based on the relevance vector machine for EHA RUL prediction. In their research, sample entropy was introduced as an effective

signature of the EHA's health.

3.3. Application in power systems

Wind energy generated by wind turbines is a growing and reliable renewable energy source in the world. However, the wind energy industry experiences increasing operation & maintenance costs because of main components failures. The temperature stress caused by the temperature difference along with the machine, e.g., shafts' and gears' temperature, together with lubrication problems accelerate wind turbine faults. To monitor the health of wind turbines, various monitoring techniques have been used, such as acoustic measurement, electrical effects monitoring, equipment vibration monitoring, power temperature monitoring, oil debris monitoring, etc. SCADA (Supervisory Control and Data Acquisition) is a commonly used tool for data collection, which is a system built into turbines to control electricity generation. This system use sensors to collect various functional parameters and data, such as temperature, bearing vibration, wind speed, and phase currents of wind turbines [155]. Carroll et al. [156] ensembled ANN, SVM, and logistic regression to predict wind turbine gearbox failure using SCADA data. This methodology appears to be effective in predicting the failure up to a month before it occurs. Inclusion of high frequency vibration data could extend that prediction capability to 5–6 months before failure occurs with reasonable accuracy. Song et al. [157] introduced a Bayesian framework with three different methods, namely, the bin method, the multivariate normal distribution based method, and the copula method to identify wind turbine health states based on their SCADA data. The results showed that copula method has the best prediction performance. Chen et al. [158] proposed an enhanced particle filtering algorithm for wind turbine drivetrain gearboxes RUL prediction using vibration data, In their method, an adaptive neuro-fuzzy inference system was used to learn the health state transition. Hu et al. [159] explored the Wiener process for the prediction of wind turbine health status using the temperature characteristics of operational SCADA data. Nielsen and Sørensen [160] proposed a Markov deterioration model to predict the deterioration and RUL of wind turbine blades. In their model, a dynamic Bayesian network was used to obtain probabilities of inspection outcomes and the maximum likelihood method was applied to estimate the transition probabilities for a hidden Markov model. Saidi et al. [161] proposed a vibration-based prognostic and health monitoring methodology for wind turbine high-speed shaft bearing using a spectral kurtosis and SVR. Reviews about wind turbine condition monitoring can be found from [155,162,163].

Electric valves, power transformers, reactor coolant pumps, etc. are also widely used components in many power systems. Though the RUL prediction of these individual components is not trivial, in recent years, researchers showed their interest in prognosis of these components. For example, Wang et al. [164] applied a convolution kernel combined with LSTM for feature extraction. Then, LSTM is utilized for predicting RUL of electric valves. Later, Wang et al. [165] improved the RUL prediction method by combining LSTM and convolutional auto-encoder (CAE). They combined deeper features extracted by CAE and the original features to enrich the dimension of features, and the case study showed an improved predictive capability. Aizpurua et al. [166] focused on lifetime predictions of power transformers in NPPs. They proposed a Bayesian Particle Filtering framework by integrating model-based experimental models, forecasting models and uncertainty modeling concepts together for condition assessment of transformers. Nguyen et al. [167] combined ensemble empirical mode decomposition and LSTM for the prognostics of reactor coolant pumps of NPPs. They observed that multi-step-ahead predictions obtained by an ensemble of separate prediction models are more accurate and less noisy than the predictions obtained by a single model.

3.4. Application in electrical and electronic components

Many electronic systems including consumer electronics, electric vehicles, airplanes, and renewable energy devices use lithium-ion batteries as the main sources of energy storage. The performance of lithium-ion batteries deteriorates over the service time in terms of capacity loss and resistance increase. To ensure the safety and reliability of lithium-ion batteries, accurate estimation of the health state and RUL prediction are essentials to track the actual performance of batteries. RUL of batteries can be determined by the number of charges and discharge cycles to reduce its capacity from the known current value to the threshold value. In the papers [1,168], they provided comprehensive reviews on data-driven methods for battery health diagnostics and prognostics estimation. We found two public datasets available: NASA Ames Prognostic enter of Excellence [169] and Center for Advanced Life Cycle Engineering (CALCE) of University of Maryland [170]. NASA Ames Prognostic enter of Excellence dataset contains 4 batteries' aging processes which were tested under certain conditions. The batteries were run through different charge, discharge, and impedance operational profiles at room temperature. The CALCE provides multiple batteries dataset. For the RUL prediction, the battery capacity was tested as an indicator of battery status. To measure the capacity, all batteries were fully charged under the constant-current/constant-voltage mode. In the discharge period, the cells were applied to a specific load to maintain at a constant current until the voltage was reduced to 2.7 V. Then the discharge capacity was recorded after each full charge–discharge process. The details of the experiments to generate data can be found in [171].

A number of researches has been reported for RUL prediction based on these two public datasets. Zhai and Ye [172] studied a Wiener process model with an adaptive drift for RUL prediction of batteries. They concluded that the proposed model fixed the deficiency of conventional Wiener process models that ignoring the variability of drifts. Shen et al. [173] proposed a Wiener-based model with measurement errors, which were assumed to be a logistic distribution with zero means. They adopted the Monte Carlo expectation–maximization method together with the Gibbs sampling for parameter estimation. Wang et al. [174] proposed a mixed-effects model based on the Wiener process to capture the two-phase degradation pattern. This model accommodated two significant aspects: phase correlation and unit heterogeneity. Zhang et al. [175] presented a stochastic modeling method and took the recovery phenomenon into consideration, which is a common phenomenon for batteries that the system performance degrades with usage and recovers in storage. Si [176] proposed a generic nonlinear stochastic modeling framework, they utilized a time-dependent drift coefficient to characterize the nonlinearity and dynamics of the degradation signals. Chen et al. [177] employed a hybrid method based on SVR and error compensation methods for RUL prediction. They used genetic algorithm to optimize the hyper-parameters of SVR to achieve better accuracy. Xue et al. [178] proposed an integrated algorithm that combines unscented Kalman filter and SVR. Similar to Chen et al. [177], they used a genetic algorithm to optimize parameters of SVR. Khumprom and Yodo [179] proposed a Deep Neural Network (DNN) and compared with other machine learning algorithms, including SVM, k-NN, ANN, and Linear Regression. The results showed that the DNN algorithm could be comparable and outweigh conventional machine learning algorithms. Ren et al. [180] integrated autoencoder with DNN, in which autoencoder was used for multi-dimensional feature extraction and DNN is trained for RUL prediction. Jiao et al. [181] tried to combine both statistical method and deep learning model, they proposed a Particle Filtering framework based on conditional variational autoencoder and a reweighting strategy to predict the RUL. Since battery powered electric vehicles are starting to play a significant role in today's automotive industry. The reliability and safety of batteries are critical. Robust RUL prediction methods for batteries are desirable and the number of publications are expected to increase rapidly.

Through the literature search, it is also observed that the prognosis methodologies are widely being applied for another electronic component namely hard disk drive (HDD). The Hard disk drive (HDD) is a complex system integrating mechanical, electricity and magnetism, which is the most important and robust data storage device for major data storage services. Self-monitoring, analysis and reporting technology (SMART) is a commercial health monitoring system, which can detect and report various indicators of HDD reliability to facilitate the HDD prognosis. However, the conventional SMART can only provide a basic evaluation, and the failure detection rate (FDR) is 3–10% [182]. To improve the accuracy of proactive failure prediction, in recent years, statistical and machine learning methods have been adopted to build prediction models based on the SMART attributes. There are two public datasets used in literature: Baidu Inc. [183] and Backblaze Inc (available at: https://www.backblaze.com/b2/hard-drive-test-data.html). The dataset of Baidu Inc. was collected from a total of 23,395 drives, which had the same initial mode. The attributes of those drives were sampled at every hour using the SMART and labeled as good or failed, with only 433 drives in the failed class and the rest of 22,962 drives in the good class. The Backblaze company also collect the dataset in the similar fashion. They gathered the SMART attributes on daily basis. In 2013, the company made the dataset available for the research community and provides update quarterly [184]. Using Baidu dataset, Xu et al. [185] introduced a RNN based approach to assessing the health status of hard drives based on the gradually changing sequential SMART attributes. Li et al. [186] proposed two hard drive failure prediction models based on Decision Trees (DTs) and Gradient Boosted Regression Trees. Both prediction models showed steady prediction performance, with high failure detection rates (80–96%) and low false alarm rates (0.006–0.31%). Using Backblaze dataset, Lima et al. [187] evaluated the performance of both LSTM and CNN architecture to predict the hard drive failure. The results of this study showed that deep learning models could be the effective alternative for failure prediction.

3.5. Performance analysis

This section first describes some performance evaluation metrics for RUL prediction. Prediction of RUL is a vast research field where researchers develop and proposed a variety of methodologies and algorithms. Hence, different researchers used different performance evaluation metrics. Following our pursuit, we found the four most used evaluation metrics, i.e., Mean Absolutes Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Scoring Function (SF). These metrics are defined as,

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{y}_i - y_i\right| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i - y_i\right)^2} \tag{13}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\widehat{y}_i - y_i\right|}{y_i} \tag{14}$$

$$SF = \begin{cases} \sum_{i=1}^{N}\left(e^{-\frac{\widehat{y}_i - y_i}{13}} - 1\right), if\, \widehat{y}_i - y_i < 0 \\ \sum_{i=1}^{N}\left(e^{\frac{\widehat{y}_i - y_i}{10}} - 1\right), if\, \widehat{y}_i - y_i \geq 0 \end{cases} \tag{15}$$

In the above equations, y_i and $\widehat{y}_i$ represent true RUL and predicted RUL, respectively. N is the total number of units or systems. All of these metrics are negatively oriented scores, which means a lower value indicates a better model. Both the MAE and RMSE report the average RUL prediction error to the model. The value of these two metrics could range between 0 and ∞. MAPE is a variant of MAE which is the absolute error normalized over the data. This metric is useful when the errors need to be compared across data with different scales. The SF, as shown in Equation (15), is asymmetric around the true time of failure. It is defined in a way that late predictions are more heavily penalized than early predictions [137]. We only list a subset of metrics in evaluating the performance of methodologies for different applications or datasets. For example, C-MAPSS dataset, researchers who developed deep learning approaches for the dataset commonly used the scoring and RMSE metrics to quantitively evaluate the proposed methodologies in this field. This provides us an opportunity to compare different deep learning approaches directly. It is worth mentioning that, the literature states a wide variety of performance criteria to evaluate different methods. This is usual and expected, as the datasets used in this research field are characterized by different parameters, scales, variations and operating conditions. Sometimes, a dataset has multiples of sub-datasets and different experimental setups. Considering this limitation, in the comparison table, we only report the common evaluation metrics for any particular dataset having the same parameters, scales, and operating conditions. IEEE PHM2012 challenge dataset, provided by the FEMTO-ST institution in France, is a popular dataset used for experimenting and predicting the bearing remaining useful life. For this dataset, MAE and RMSE are often used to evaluate the proposed methodologies. A comparison table for C-MAPSS and IEEE PHM 2012 challenge dataset is reported in Table 4 and Table 5, respectively. From Table 4, we can observe that deep learning approaches applied to the C-MAPSS dataset are only started in 2016. Since then, the number of publications has grown exponentially, which has been demonstrated in Fig. 8. The best three performances in each column of Table 4 are labeled in bold. Due to the calculation biases from different authors, the RMSE and Scoring are not consistent. Generally, we can conclude that capsule NN [188] and MSCNN [77] show superior performance and a good generalization capability for all sub-datasets. Bayesian Deep learning [101] also demonstrates satisfying results. It is worth mentioning that unlike other deep learning approaches only focusing on point estimation of RUL, Bayesian Deep learning enhanced the model interpretability by providing non only point estimation, but also uncertainty quantification, which is highly desirable in practice. For IEEE PHM2012 challenge dataset, the performance metrics, dataset and data prediction points vary, we only provide a rough comparison based on the work from [224]. As we can see, DNN performs the best comparing other traditional machine learning approaches.

4. Challenges and future trends

4.1. Challenges and future trends for the predictive maintenance

With advancements in the industrial Internet of things and artificial intelligence, predictive maintenance has become more and more efficient. The core point to apply predictive maintenance strategy successfully is to model and predict failure patterns accurately. Based on our studies, it can be seen that this area has been well studied using many methodologies ranging from statistical approaches to machine learning based approaches. Recently, many researchers have focused on machine learning approaches to explore the field of RUL prediction. Deep neural networks such as CNNs, auto-encoder, RNN, LSTM are getting popularity for learning features from raw data. Some researchers integrate the statistical and traditional machine learning approach to explore this field [196,197]. However, the task of RUL is still challenging due to the complex, uncertain, nonlinear features and operational conditions. The main challenges can be summarized in several aspects: (1) Data insufficiency and imbalanced classes: most data-driven models, especially machine learning approaches, predict the RUL based on the extracted

Table 4
Performance comparison for C-MAPSS dataset.

Method	year	F001		F002		F003		F004	
		Scoring	RMSE	Scoring	RMSE	Scoring	RMSE	Scoring	RMSE
CNN [72]	2016	1287	18.45	13,570	30.29	1596	19.82	7886	29.16
Multi-objective Deep belief NN [189]	2017	640	17.96	10,851	28.06	683	19.41	7210	29.45
Deep LSTM [129]	2017	338	16.14	4450	24.49	852	16.18	5550	27.17
BiLSTM [22]	2018	295	13.65	4130	23.18	317	13.74	5430	24.86
LSTM-FNN [190]	2018	481	14.89	7982	26.86	493	15.11	5200	27.11
Deep CNN [191]	2018	274	12.61	10,400	22.36	284	12.64	12,500	23.31
hybrid LSTM [87]	2019	262	14.72	6953	29.00	452	17.72	15,069	33.43
Ensemble ResCNN [147]	2019	**212**	**12.16**	**2087**	20.85	**180**	12.01	3400	24.97
capsule NN [188]	2020	276	12.58	**1229**	**16.30**	283	**11.71**	**2625**	**18.96**
MSCNN [77]	2020	**196**	**11.44**	3747	**19.35**	**241**	**11.67**	4844	22.22
CNN + LSTM [192]	2020	231	12.56	3366	22.73	**251**	12.10	**2840**	22.66
Generative Adversarial Networks [193]	2020	**174**	**10.71**	2982	19.49	273	**11.48**	3874	**19.71**
Bayesian Deep learning [194]	2020	267	12.19	**2007**	**18.49**	409	12.07	**2415**	**19.41**

Table 5
Performance comparison for PHM 2012 challenge dataset.

Dataset	Method	MAE (80% training)	RMSE (80% training)
FEMTO-ST institution at PRONOSTIA platform (PHM 2012 challenge dataset)	DNN	**0.0333**	**0.0467**
	GBDT (Gradient boosting decision tree)	**0.0333**	0.0600
	Gaussian	0.0633	0.0933
	SVM	0.0600	0.0800
	BP Neural Network	0.1133	0.1533
	Bayesian Ridge	0.1400	0.1833

**The metrics values are approximated from Ren et al. [195].

features learned from the data. The predictive performance largely relies on the volume and quality of the datasets. However, with the enhanced quality of engineering systems, the probabilities of failure are extremely low. Data collection in the engineering field is time-consuming and costly. Moreover, the collected data usually face the class imbalance issue. As stated in Section 3, most datasets are generated in a simulated environment for case studies. The developed approaches may show poor performance when adapting to real-world scenarios. (2) Poor generalization ability of developed models: many developed models show poor performance when applying to different types or even the same type of equipment under different operational environments. Although transfer learning has been attempted to solve this problem, in practice remains a challenging issue to achieve satisfactory performance in many scenarios. Moreover, the degradation mechanism may be affected by various environmental factors (e.g., speed, loading) in the real world. How to accurately model these variables are still challenging and critical since they will impact the degradation trend which in turn will impact the prediction performance. (3) Late prediction caused by poor predictive capability: if the predicted failure is later than the actual fault occurrence, it may not allow adequate time for remedies, thus causing damage or losses. However, current research pay little attention to time complexity analysis. (4) Noise associated with real-time/online prognostics for in-situ components: the operating components may be affected by random disturbance from the environment. The extensive research on how to incorporate data heterogeneity and uncertainty for real-time prognostics are expected in the future. Moreover, online monitoring and prognostics require fast and efficient algorithms, only a limited research report the computational time. (5) Manual assignment of hyper-parameters estimation and tuning: many algorithms require to assign and tune hyper-parameters, especially for deep learning models, the accuracy of the model is largely dependent on the choice of hyper-parameters. However, the majority of the research only assigns parameters manually. (6) Discrepancy of cross-domain prognosis: it is generally assumed that the training and testing data are from the same distribution. In real world, distribution discrepancy exists between training and testing data due to variation of machine working condition, interference of environmental noise, etc. This may lead to significant prediction performance deterioration.

Opportunities always come along with challenges. These opportunities include, but are not limited to, (1) Dataset enrichment: currently, most of the data-driven algorithms only utilize numeric data to predict the remaining life. It is well known that deep learning algorithms are efficient at extracting features and learn from images. Thus, integration of images and numerical data could bring an effective way to dig into the task of RUL prediction. In addition, the rise and recent development of transfer learning and data augmentation techniques (e.g., Generative adversarial networks) can also alleviate the demand for large datasets and boost the prognostic accuracy. However, currently only a very limited number of research have been conducted for RUL prediction. (2) Uncertainty quantification: a majority of machine learning based approaches are only focused on predicting the mean values of RUL. Combining statistical approaches and then quantifying uncertainty for more informed decision-making will be an interesting direction to investigate. (3) Development of robust yet effective prediction algorithms for online monitoring: with the emerging of machine learning techniques in predictive maintenance field, the proposed models are becoming more and more complex, researchers may focus on how to control the computation time through designing efficient models. (4) Model generalization: most prognostics approaches are typically designed for a specific system or domain. It would be interesting to develop general prognostic methods, which can be implemented in any systems, and to promote in practical applications. (5) Real-time maintenance strategies development: basically, real-time schemes for decision making based on RUL prediction that can deal with more complex and dynamic scenarios are anticipated in society for maintenance and reliability. By utilizing advanced cloud computing, AI and machine learning algorithms, we believe that more automated, intelligent tools will be developed and deployed to improve performance in the industry soon. (6) Integration of attention mechanism into the predictive model: attention mechanism is able to model the dependencies between the target output and the input sequences and has now become an important element in deep learning area, we'll expect to see more and more publications that attention mechanism to RUL prediction in the near future.

4.2. Guidelines for model selection of predictive maintenance implementation

After a thorough review of data-driven algorithms with different applications, we believe the selection of a concrete data-driven model is application-dependent. Also, different models have been their own pros and cons. We would like to provide our perception on how to select and construct an appropriate data-driven model for a specific application

scenario. If the data size is small and degradation data show similar degradation forms, GPMs perhaps are the most suitable and simplest models to use. However, their inability to capture the temporal variability and the uncertainty inherent in the progression of deterioration over time, which is common in practice, limits their engineering applications. In other words, GPMs are applicable only when the unexplained randomness is sufficiently small. To deal with the randomness caused by inherent variability and environmental factors, SPMs are a natural choice. If degradation processes are monotonic and evolving only in one direction, Gamma process and inverse Gaussian process are appropriate to model this type of degradation data. Wiener process is suitable in modeling deterioration which is not monotone. To relax the assumption of parametric forms, Gaussian process models can be well adapted to model the complex data, where do not involve parameters. Both GPMs and SPMs have well-established statistical properties, where a closed form of PDF of RUL is usually available. If not in some cases, filtered based methods or some sampling methods have to be used for finding an approximated RUL. One limitation of both GPMs and SPMs is that they require a pre-defined failure threshold. However, this may not be available or accurate since it requires knowledge from domain experts. Moreover, a fixed failure threshold may not be sufficient to characterize the health status of all products due to their heterogeneous features. In this case, covariate based models are efficient without needing failure threshold assumption. The rapid development of sensing and computing technologies has enriched degradation data significantly. This data-rich environment for degradation modeling and prognostics that could potentially lead to an accurate inference about RUL of products. However, RUL prediction with multi-sensor signals is a more challenging issue than the cases of a single degradation signal. One way to deal with this issue is to combine multi-sensor signals into a composite health index or mapping the correlation between signals and RUL values, then widely-used GPMs and stochastic process models are still applicable. Another option is to use machine learning, which has gradually become a mainstream for RUL prediction. The goal of conventional machine learning and deep learning for RUL prediction is to learn the non-linear mapping between the sensor data and RUL using different network architecture. Among those machine learning models, LSTMs have attracted great attention and presented an outstanding ability in the application of RUL prediction, as they have the capability to learn dependencies of sequential data. While machine models can provide better performance for RUL prediction, they do not have a probabilistic orientation, namely, uncertainty quantification, and therefore, no PDF of the RUL is available. Bayesian neural networks have been used to cover the shortage. If there is a need for cross-domain prognosis, transfer learning is a preferable way to provide a better performance. Attention mechanism can also assist the learning model in yielding potential improvements in the learning tasks.

5. Conclusion

In the context of Industry 4.0, predictive maintenance is transforming the way of thinking maintenance: from cost to business opportunity in the industry. Based on this rationale, predictive maintenance is attracting considerable investment from industries and increasing attention from research societies. Many predictive maintenance techniques have been developed up to now to respond to the demand of high reliability of facilities and equipment but more studies are still required to improve their predictive accuracies and efficiencies. This review provides a comprehensive overview of the most recent data-driven prognostic techniques, which is the indispensable process for predictive maintenance. Specifically, this paper reviews the methodologies, best practices, current challenges, and future trends of machine prognostics. To make accurate prognostics, choosing a proper modeling technique is essential. We provide a detailed summary of statistical based models and machine learning based models. Then, their applications based on these models are demonstrated in detail, which provide a good reference for selecting an appropriate model for a specific application scenario. Moreover, we investigate and pinpoint some challenges and promising directions and opportunities of prognostics for future studies. Lastly, we provide some constructive resolutions to mitigate the predictive maintenance challenges (e.g., data insufficiency and imbalanced classes; poor generalization ability of developed models; late prediction caused by poor predictive capability; noise associated with real-time/online prognostics for in-situ components; manual assignment of hyper-parameters estimation and tuning; and discrepancy of cross-domain prognosis) and guidance for the user to choose appropriate models to support predictive maintenance implementation. This research effort would lead to develop machine learning based predictive maintenance system that enables to sustain effective and accurate preventive maintenance. In summary, this review provides an indication of how to study predictive maintenance problems from data-driven machine prognostics perspective and pave a path for effective further investigation. It can be foreseen that more and more advanced predictive models will be developed in the near future, which will boost predictive maintenance, improve reliability, enhance productivity and achieve intelligent decision-making in the industry.

CRediT authorship contribution statement

Yuxin Wen: Data curation, Writing– original draft, Writing – review & editing. **Md. Fashiar Rahman:** Visualization, Investigation, Validation. **Honglun Xu:** Conceptualization, Methodology. **Tzu-Liang Bill Tseng:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the National Science Foundation (ECR-PEER-1935454), (ERC-ASPIRE-1941524) and Department of Education (Award # P120A180101). The authors wish to express sincere gratitude for their financial support.

References

[1] H. Meng, Y.-F. Li, A review on prognostics and health management (PHM) methods of lithium-ion batteries, Renew. Sustain. Energy Rev. 116 (2019), 109405.
[2] P.G. Ramesh, S.J. Dutta, S.S. Neog, P. Baishya, I. Bezbaruah, Implementation of Predictive Maintenance Systems in Remotely Located Process Plants under Industry 4.0 Scenario, Advances in RAMS Engineering, Springer, 2020, pp. 293–326.
[3] N. Sakib, T. Wuest, Challenges and opportunities of condition-based predictive maintenance: a review, Procedia CIRP 78 (2018) 267–272.
[4] R. Ahmad, S. Kamaruddin, An overview of time-based and condition-based maintenance in industrial application, Comput. Ind. Eng. 63 (2012) 135–149.
[5] A. Jezzini, M. Ayache, L. Elkhansa, B. Makki, M. Zein, Effects of predictive maintenance(PdM), Proactive maintenace(PoM) & Preventive maintenance(PM) on minimizing the faults in medical instruments, in: 2013 2nd International Conference on Advances in Biomedical Engineering, 2013, pp. 53–56.
[6] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Sig. Process. 20 (2006) 1483–1510.
[7] K.L. Tsui, N. Chen, Q. Zhou, Y. Hai, W. Wang, Prognostics and health management: a review on data driven approaches, Math. Probl. Eng. 2015 (2015).
[8] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: a systematic review from data acquisition to RUL prediction, Mech. Syst. Sig. Process. 104 (2018) 799–834.
[9] M.S. Kan, A.C.C. Tan, J. Mathew, A review on prognostic techniques for non-stationary and non-linear rotating systems, Mech. Syst. Sig. Process. 62 (2015) 1–20.
[10] M.G. Pecht, A prognostics and health management roadmap for information and electronics-rich systems, IEICE ESS Fundam. Rev. 3 (2010), 4_25-24_32.

[11] Y. Wen, J. Wu, Q. Zhou, T.-L. Tseng, Multiple-change-point modeling and exact bayesian inference of degradation signal for prognostic improvement, IEEE Trans. Autom. Sci. Eng. (2018) 1–16.
[12] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, Remaining useful life estimation – A review on the statistical data driven approaches, Eur. J. Oper. Res. 213 (2011) 1–14.
[13] Z.S. Ye, M. Xie, Stochastic modelling and analysis of degradation for highly reliable products, Appl. Stochastic Models Bus. Ind. 31 (2015) 16–32.
[14] Z. Zhang, X. Si, C. Hu, X. Kong, Degradation modeling–based remaining useful life estimation: a review on approaches for systems with heterogeneity, Proc. Inst. Mech. Eng. Part O: J. Risk Reliab. 229 (2015) 343–355.
[15] D. Wang, K.-L. Tsui, Q. Miao, Prognostics and health management: a review of vibration based bearing and gear health indicators, IEEE Access 6 (2017) 665–676.
[16] S. Khan, T. Yairi, A review on the application of deep learning in system health management, Mech. Syst. Sig. Process. 107 (2018) 241–265.
[17] L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, M. Wei, A review on deep learning applications in prognostics and health management, IEEE Access 7 (2019) 162415–162438.
[18] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Sig. Process. 115 (2019) 213–237.
[19] M. Kordestani, M. Saif, M.E. Orchard, R. Razavi-Far, K. Khorasani, Failure prognosis and applications—a survey of recent literature, IEEE Trans. Reliab. (2019).
[20] J. Guo, Z. Li, M. Li, A review on prognostics methods for engineering systems, IEEE Trans. Reliab. (2019) 1–20.
[21] M. Baur, P. Albertelli, M. Monno, A review of prognostics and health management of machine tools, Int. J. Adv. Manuf. Technol. 107 (2020) 2843–2863.
[22] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications, Mech. Syst. Sig. Process. 42 (2014) 314–334.
[23] C.J. Lu, W.O. Meeker, Using degradation measures to estimate a time-to-failure distribution, Technometrics 35 (1993) 161–174.
[24] N. Gebraeel, A. Elwany, J. Pan, Residual life predictions in the absence of prior degradation knowledge, IEEE Trans. Reliab. 58 (2009) 106–117.
[25] H. Kim, J.T. Kim, G. Heo, Prognostics for integrity of steam generator tubes using the general path model, Nucl. Eng. Technol. 50 (2018) 88–96.
[26] N. Gebraeel, Sensory-updated residual life distributions for components with exponential degradation patterns, IEEE Trans. Autom. Sci. Eng. 3 (2006) 382–393.
[27] Y. Wen, J. Wu, Y. Yuan, Multiple-phase modeling of degradation signal for condition monitoring and remaining useful life prediction, IEEE Trans. Reliab. 66 (2017) 924–938.
[28] R. Zhou, N. Serban, N. Gebraeel, Degradation-based residual life prediction under different environments, Ann. Appl. Stat. (2014) 1671–1689.
[29] N. Chen, K.L. Tsui, Condition monitoring and remaining useful life prediction using degradation signals: revisited, IIE Trans. 45 (2013) 939–952.
[30] G.A. Whitmore, Estimating degradation by a Wiener diffusion process subject to measurement error, Lifetime Data Anal. 1 (1995) 307–319.
[31] X. Wang, Wiener processes with random effects for degradation data, J. Multivariate Anal. 101 (2010) 340–351.
[32] Y. Wen, J. Wu, D. Das, T.-L.-B. Tseng, Degradation modeling and RUL prediction using Wiener process subject to multiple change points and unit heterogeneity, Reliab. Eng. Syst. Saf. 176 (2018) 113–124.
[33] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, M.G. Pecht, Remaining useful life estimation based on a nonlinear diffusion degradation process, IEEE Trans. Reliab. 61 (2012) 50–67.
[34] Z.-S. Ye, Y. Wang, K.-L. Tsui, M. Pecht, Degradation data analysis using Wiener processes with measurement errors, IEEE Trans. Reliab. 62 (2013) 772–780.
[35] X.-S. Si, W. Wang, C.-H. Hu, M.-Y. Chen, D.-H. Zhou, A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation, Mech. Syst. Sig. Process. 35 (2013) 219–237.
[36] C.-Y. Peng, S.-T. Tseng, Mis-specification analysis of linear degradation models, IEEE Trans. Reliab. 58 (2009) 444–455.
[37] H. Wang, X. Ma, Y. Zhao, An improved Wiener process model with adaptive drift and diffusion for online remaining useful life prediction, Mech. Syst. Sig. Process. 127 (2019) 370–387.
[38] Z.-S. Ye, Y. Shen, M. Xie, Degradation-based burn-in with preventive maintenance, Eur. J. Oper. Res. 221 (2012) 360–367.
[39] Z. Zhang, X. Si, C. Hu, Y. Lei, Degradation data analysis and remaining useful life estimation: a review on Wiener-process-based methods, Eur. J. Oper. Res. 271 (2018) 775–796.
[40] Q. Dong, L. Cui, A study on stochastic degradation process models under different types of failure thresholds, Reliab. Eng. Syst. Saf. 181 (2019) 202–212.
[41] J.M. van Noortwijk, A survey of the application of gamma processes in maintenance, Reliab. Eng. Syst. Saf. 94 (2009) 2–21.
[42] R.R. Richardson, M.A. Osborne, D.A. Howey, Gaussian process regression for forecasting battery state of health, J. Power Sources 357 (2017) 209–219.
[43] P. Boškoski, M. Gašperin, D. Petelin, Đ. Juričić, Bearing fault prognostics using Rényi entropy based features and Gaussian process models, Mech. Syst. Sig. Process. 52 (2015) 327–337.
[44] S.A. Aye, P.S. Heyns, An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission, Mech. Syst. Sig. Process. 84 (2017) 485–498.
[45] D. Yang, X. Zhang, R. Pan, Y. Wang, Z. Chen, A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve, J. Power Sources 384 (2018) 387–395.
[46] Z.-S. Ye, N. Chen, The inverse Gaussian process as a degradation model, Technometrics 56 (2014) 302–311.
[47] F. Cartella, J. Lemeire, L. Dimiccoli, H. Sahli, Hidden semi-Markov models for predictive maintenance, Math. Probl. Eng. 2015 (2015).
[48] C.R. David, Regression models and life tables (with discussion), J. Roy. Stat. Soc. 34 (1972) 187–220.
[49] Q. Zhou, J. Son, S. Zhou, X. Mao, M. Salman, Remaining useful life prediction of individual units subject to hard failure, IIE Trans. 46 (2014) 1017–1030.
[50] J. Man, Q. Zhou, Prediction of hard failures with stochastic degradation signals using Wiener process and proportional hazards model, Comput. Ind. Eng. 125 (2018) 480–489.
[51] J. Hu, Q. Sun, Z. Ye, Q. Zhou, Joint modeling of degradation and lifetime data for RUL prediction of deteriorating products, IEEE Trans. Ind. Inf. (2020) 1.
[52] X. Yue, R.A. Kontar, Joint models for event prediction from time series and survival data, Technometrics (2020) 1–26.
[53] J.L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, Y. Kluger, DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network, BMC Med. Res. Method. 18 (2018) 24.
[54] H. Kvamme, Ø. Borgan, I. Scheel, Time-to-event prediction with neural networks and Cox regression, arXiv preprint arXiv:1907.00825, 2019.
[55] P.J.G. Nieto, E. García-Gonzalo, F.S. Lasheras, F.J. de Cos Juez, Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability, Reliab. Eng. Syst. Saf. 138 (2015) 219–231.
[56] T. Qin, S. Zeng, J. Guo, Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO–SVR model, Microelectron. Reliab. 55 (2015) 1280–1284.
[57] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Remaining useful life estimation based on nonlinear feature reduction and support vector regression, Eng. Appl. Artif. Intell. 26 (2013) 1751–1760.
[58] Y. Hu, C. Hu, X. Kong, Z. Zhou, Real-time lifetime prediction method based on wavelet support vector regression and fuzzy c-means clustering, Acta Autom. Sin. 38 (2012) 331–340.
[59] C. Shen, D. Wang, F. Kong, W.T. Peter, Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, Measurement 46 (2013) 1551–1564.
[60] J. Liu, R. Seraoui, V. Vitelli, E. Zio, Nuclear power plant components condition monitoring by probabilistic support vector machine, Ann. Nucl. Energy 56 (2013) 23–33.
[61] H.-Z. Huang, H.-K. Wang, Y.-F. Li, L. Zhang, Z. Liu, Support vector machine based estimation of remaining useful life: current research status and future trends, J. Mech. Sci. Technol. 29 (2015) 151–163.
[62] P. Kundu, A.K. Darpe, M.S. Kulkarni, An ensemble decision tree methodology for remaining useful life prediction of spur gears under natural pitting progression, Struct. Health Monit. 19 (2020) 854–872.
[63] L. Wang, D. Zhou, H. Zhang, W. Zhang, J. Chen, Application of relative entropy and gradient boosting decision tree to fault prognosis in electronic circuits, Symmetry 10 (2018) 495.
[64] M. Ferguson, R. Ak, Y.-T.T. Lee, K.H. Law, Automatic localization of casting defects with convolutional neural networks, in: 2017 IEEE international conference on big data (big data), IEEE, 2017, pp. 1726–1735.
[65] M.K. Ferguson, A. Ronay, Y.-T.T. Lee, K.H. Law, Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning, Smart Sustain. Manuf. Syst. 2 (2018).
[66] M.F. Rahman, J. Wu, T.L.B. Tseng, Automatic morphological extraction of fibers from SEM images for quality control of short fiber-reinforced composites manufacturing, CIRP J. Manuf. Sci. Technol. 33 (2021) 176–187.
[67] W. Hou, Y. Wei, J. Guo, Y. Jin, Automatic detection of welding defects using deep neural network, J. Phys.: Conf. Ser., IOP Publishing (2017), 012006.
[68] Z. Huang, Z. Pan, B. Lei, Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data, Remote Sensing 9 (2017) 907.
[69] M.F. Rahman, T.-L.B. Tseng, M. Pokojovy, W. Qian, B. Totada, H. Xu, An automatic approach to lung region segmentation in chest x-ray images using adapted U-Net architecture, Medical Imaging 2021: Physics of Medical Imaging, International Society for Optics and Photonics, 2021, pp. 115953I.
[70] M.F. Rahman, Y. Wen, H. Xu, T.-L.B. Tseng, S. Akundi, Data mining in telemedicine, Adv. Telemed. Health Monit. (2020) 103.
[71] L. Wen, L. Gao, X. Li, B. Zeng, Convolutional neural network with automatic learning rate scheduler for fault classification, IEEE Trans. Instrum. Meas. 70 (2021) 1–12.
[72] G.S. Babu, P. Zhao, X.-L. Li, Deep convolutional neural network based regression approach for estimation of remaining useful life, International conference on database systems for advanced applications, Springer, 2016, pp. 214–228.
[73] L. Ren, Y. Sun, H. Wang, L. Zhang, Prediction of bearing remaining useful life with deep convolution neural network, IEEE Access 6 (2018) 13041–13049.
[74] B. Yang, R. Liu, E. Zio, Remaining useful life prediction based on a double-convolutional neural network architecture, IEEE Trans. Ind. Electron. 66 (2019) 9521–9530.
[75] S. Kiranyaz, A. Gastli, L. Ben-Brahim, N. Al-Emadi, M. Gabbouj, Real-time fault detection and identification for MMC using 1-D convolutional neural networks, IEEE Trans. Ind. Electron. 66 (2018) 8760–8771.

[76] J. Zhu, N. Chen, W. Peng, Estimation of bearing remaining useful life based on multiscale convolutional neural network, IEEE Trans. Ind. Electron. 66 (2018) 3208–3216.
[77] H. Li, W. Zhao, Y. Zhang, E. Zio, Remaining useful life prediction using multi-scale deep convolutional neural network, Appl. Soft Comput. 89 (2020), 106113.
[78] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[79] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078, 2014.
[80] Y. Song, L. Li, Y. Peng, D. Liu, Lithium-Ion Battery Remaining Useful Life Prediction Based on GRU-RNN, in: 2018 12th International Conference on Reliability, Maintainability, and Safety (ICRMS), IEEE, 2018, pp. 317–322.
[81] J. Chen, H. Jing, Y. Chang, Q. Liu, Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process, Reliab. Eng. Syst. Saf. 185 (2019) 372–382.
[82] J. Wang, J. Yan, C. Li, R.X. Gao, R. Zhao, Deep heterogeneous GRU model for predictive analytics in smart manufacturing: application to tool wear prediction, Comput. Ind. 111 (2019) 1–14.
[83] F.O. Heimes, Recurrent neural networks for remaining useful life estimation, IEEE, pp. 1–6.
[84] J. Liu, A. Saxena, K. Goebel, B. Saha, W. Wang, An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries, National Aeronautics And Space Administration Moffett Field CA Ames Research … (2010).
[85] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, Neurocomputing 240 (2017) 98–109.
[86] Y. Zhang, R. Xiong, H. He, M.G. Pecht, Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries, IEEE Trans. Veh. Technol. 67 (2018) 5695–5705.
[87] S. Zhao, Y. Zhang, S. Wang, B. Zhou, C. Cheng, A recurrent neural network approach for remaining useful life prediction utilizing a novel trend features construction method, Measurement 146 (2019) 279–288.
[88] S. Xiang, Y. Qin, C. Zhu, Y. Wang, H. Chen, Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction, Eng. Appl. Artif. Intell. 91 (2020), 103587.
[89] A. Zhang, H. Wang, S. Li, Y. Cui, Z. Liu, G. Yang, J. Hu, Transfer learning with deep recurrent neural networks for remaining useful life estimation, Appl. Sci. 8 (2018) 2416.
[90] J. Wang, G. Wen, S. Yang, Y. Liu, Remaining useful life estimation in prognostics using deep bidirectional lstm neural network, in: 2018 Prognostics and System Health Management Conference (PHM-Chongqing), IEEE, 2018, pp. 1037–1042.
[91] A. Elsheikh, S. Yacout, M.-S. Ouali, Bidirectional handshaking LSTM for remaining useful life prediction, Neurocomputing 323 (2019) 148–156.
[92] S. Xiang, Y. Qin, J. Luo, H. Pu, B. Tang, Multicellular LSTM-based deep learning model for aero-engine remaining useful life prediction, Reliab. Eng. Syst. Saf. 216 (2021), 107927.
[93] Q. An, Z. Tao, X. Xu, M. El Mansori, M. Chen, A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network, Measurement 154 (2020), 107461.
[94] Y. Wang, H. Yao, S. Zhao, Auto-encoder based dimensionality reduction, Neurocomputing 184 (2016) 232–242.
[95] J. Ma, H. Su, W.-L. Zhao, B. Liu, Predicting the remaining useful life of an aircraft engine using a stacked sparse autoencoder with multilayer self-learning, Complexity 2018 (2018).
[96] Y. Song, G. Shi, L. Chen, X. Huang, T. Xia, Remaining useful life prediction of turbofan engine using hybrid model based on autoencoder and bidirectional long short-term memory, J. Shanghai Jiaotong Univ. (Science) 23 (2018) 85–94.
[97] C. Su, L. Li, Z. Wen, Remaining useful life prediction via a variational autoencoder and a time-window-based sequence neural network, Qual. Reliab. Eng. Int. (2020).
[98] W. Mao, J. He, M.J. Zuo, Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning, IEEE Trans. Instrum. Meas. 69 (2020) 1594–1608.
[99] M. Xia, T. Li, T. Shu, J. Wan, C.W. De Silva, Z. Wang, A two-stage approach for the remaining useful life prediction of bearings using deep neural networks, IEEE Trans. Ind. Inf. 15 (2018) 3703–3711.
[100] W. Peng, Z.-S. Ye, N. Chen, Bayesian deep-learning-based health prognostics toward prognostics uncertainty, IEEE Trans. Ind. Electron. 67 (2019) 2283–2293.
[101] M. Kim, K. Liu, A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics, IISE Trans. (2020) 1–23.
[102] G. Li, L. Yang, C.-G. Lee, X. Wang, M. Rong, A Bayesian deep learning RUL framework integrating epistemic and aleatoric uncertainties, IEEE Trans. Ind. Electron. (2020).
[103] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2009) 1345–1359.
[104] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, X. Chen, Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing, IEEE Trans. Ind. Inf. 15 (2018) 2416–2425.
[105] Y. Fan, S. Nowaczyk, T. Rögnvaldsson, Transfer learning for remaining useful life prediction based on consensus self-organizing models, Reliab. Eng. Syst. Saf. 203 (2020), 107098.
[106] H. Zhang, Q. Zhang, S. Shao, T. Niu, X. Yang, H. Ding, Sequential network with residual neural network for rotatory machine remaining useful life prediction using deep transfer learning, Shock Vib. 2020 (2020).
[107] Z. Zhang, Y. Wang, K. Wang, Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network, J. Intell. Manuf. 24 (2013) 1213–1227.
[108] Y. Wang, G. Xu, L. Liang, K. Jiang, Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis, Mech. Syst. Sig. Process. 54 (2015) 259–276.
[109] G. Bin, J. Gao, X. Li, B. Dhillon, Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network, Mech. Syst. Sig. Process. 27 (2012) 696–711.
[110] J.-X. Zhang, D.-B. Du, X.-S. Si, C.-H. Hu, H.-W. Zhang, Joint optimization of preventive maintenance and inventory management for standby systems with hybrid-deteriorating spare parts, Reliab. Eng. Syst. Saf. 214 (2021), 107686.
[111] J. Cai, Y. Yin, L. Zhang, X. Chen, Joint optimization of preventive maintenance and spare parts inventory with appointment policy, Math. Probl. Eng. 2017 (2017).
[112] Y. Jiang, M. Chen, D. Zhou, Joint optimization of preventive maintenance and inventory policies for multi-unit systems subject to deteriorating spare part inventory, J. Manuf. Syst. 35 (2015) 191–205.
[113] J.-X. Zhang, X.-S. Si, D.-B. Du, C.-H. Hu, C. Hu, A novel iterative approach of lifetime estimation for standby systems with deteriorating spare parts, Reliab. Eng. Syst. Saf. 201 (2020), 106960.
[114] H. Jia, Y. Ding, R. Peng, Y. Song, Reliability evaluation for demand-based warm standby systems considering degradation process, IEEE Trans. Reliab. 66 (2017) 795–805.
[115] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier, PRONOSTIA: An experimental platform for bearings accelerated degradation tests.
[116] R.M. Mortier, S.T. Orszulik, M.F. Fox, Chemistry and Technology of Lubricants, Springer, 2010.
[117] Y. Du, T. Wu, J. Cheng, R. Gong, Lubricating oil deterioration on a four-ball test rig via on-line monitoring, in: Proceedings of Malaysian international tribology conference, 2015, pp. 185–186.
[118] Y. Du, T. Wu, S. Zhou, V. Makis, Remaining useful life prediction of lubricating oil with dynamic principal component analysis and proportional hazards model, Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol. 234 (2020) 964–971.
[119] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, J. Dybala, A model-based method for remaining useful life prediction of machinery, IEEE Trans. Reliab. 65 (2016) 1314–1326.
[120] S. Hong, Z. Zhou, C. Lu, B. Wang, T. Zhao, Bearing remaining life prediction using Gaussian process regression with composite kernel functions, J. Vibroeng. 17 (2015) 695–704.
[121] Z. Liu, M.J. Zuo, Y. Qin, Remaining useful life prediction of rolling element bearings based on health state assessment, Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci. 230 (2016) 314–330.
[122] H. Wang, Y. Zhao, X. Ma, Remaining useful life prediction using a novel two-stage wiener process with stage correlation, IEEE Access 6 (2018) 65227–65238.
[123] W. Ahmad, S.A. Khan, M.M. Islam, J.-M. Kim, A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models, Reliab. Eng. Syst. Saf. 184 (2019) 67–76.
[124] P. Kundu, S. Chopra, B.K. Lad, Multiple failure behaviors identification and remaining useful life prediction of ball bearings, J. Intell. Manuf. 30 (2019) 1795–1807.
[125] M. Zhao, B. Tang, Q. Tan, Bearing remaining useful life estimation based on time–frequency representation and supervised dimensionality reduction, Measurement 86 (2016) 41–55.
[126] L. Xiao, X. Chen, X. Zhang, M. Liu, A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition, J. Intell. Manuf. 28 (2017) 1893–1914.
[127] A.R. Bastami, A. Aasi, H.A. Arghand, Estimation of remaining useful life of rolling element bearings using wavelet packet decomposition and artificial neural network, Iranian J. Sci. Technol., Trans. Electr. Eng. 43 (2019) 233–245.
[128] W. Mao, J. He, J. Tang, Y. Li, Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network, Adv. Mech. Eng. 10 (2018), 1687814018817184.
[129] G. Tang, Y. Zhou, H. Wang, G. Li, Prediction of bearing performance degradation with bottleneck feature based on LSTM network, in: 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE, 2018, pp. 1–6.
[130] L. Ren, X. Cheng, X. Wang, J. Cui, L. Zhang, Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction, Future Gen. Comput. Syst. 94 (2019) 601–609.
[131] X. Li, W. Zhang, Q. Ding, Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction, Reliab. Eng. Syst. Saf. 182 (2019) 208–218.
[132] Q. Wu, C. Zhang, Cascade fusion convolutional long-short time memory network for remaining useful life prediction of rolling bearing, IEEE Access 8 (2020) 32957–32965.
[133] C.-C. Lo, C.-H. Lee, W.-C. Huang, Prognosis of bearing and gear wears using convolutional neural network with hybrid loss function, Sensors 20 (2020) 3539.
[134] M. Tanwar, N. Raghavan, Lubricating oil degradation modeling and prognostics using the Wiener process, in: 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), IEEE, 2019, pp. 601–605.

[135] M. Tanwar, N. Raghavan, Lubricating oil remaining useful life prediction using multi-output gaussian process regression, IEEE Access 8 (2020) 128897–128907.
[136] V.T. Le, C.P. Lim, S. Mohamed, S. Nahavandi, L. Yen, G.E. Gallasch, S. Baker, D. Ludovici, N. Draper, V. Wickramanayake, Condition monitoring of engine lubrication oil of military vehicles: a machine learning approach, in: 17th Australian International Aerospace Congress: AIAC 2017, Engineers Australia, Royal Aeronautical Society, 2017, pp. 718.
[137] A. Saxena, K. Goebel, D. Simon, N. Eklund, Damage propagation modeling for aircraft engine run-to-failure simulation, in: 2008 international conference on prognostics and health management, IEEE, 2008, pp. 1–9.
[138] A. Chehade, S. Bonk, K. Liu, Sensory-based failure threshold estimation for remaining useful life prediction, IEEE Trans. Reliab. 66 (2017) 939–949.
[139] M. Kim, C. Song, K. Liu, A generic health index approach for multisensor degradation modeling and sensor selection, IEEE Trans. Autom. Sci. Eng. 16 (2019) 1426–1437.
[140] C. Song, K. Liu, Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: a composite health index approach, IISE Trans. 50 (2018) 853–867.
[141] A. Chehade, C. Song, K. Liu, A. Saxena, X. Zhang, A data-level fusion approach for degradation modeling and prognostic analysis under multiple failure modes, J. Qual. Technol. 50 (2018) 150–165.
[142] N. Li, Y. Lei, T. Yan, N. Li, T. Han, A Wiener-process-model-based method for remaining useful life prediction considering unit-to-unit variability, IEEE Trans. Ind. Electron. 66 (2018) 2092–2101.
[143] K. Le Son, M. Fouladirad, A. Barros, Remaining useful lifetime estimation and noisy gamma deterioration process, Reliab. Eng. Syst. Saf. 149 (2016) 76–87.
[144] C. Ordóñez, F.S. Lasheras, J. Roca-Pardiñas, F.J. de Cos Juez, A hybrid ARIMA–SVM model for the study of the remaining useful life of aircraft engines, J. Comput. Appl. Math. 346 (2019) 184–191.
[145] A. Al-Dulaimi, S. Zabihi, A. Asif, A. Mohammadi, A multimodal and hybrid deep neural network model for remaining useful life estimation, Comput. Ind. 108 (2019) 186–196.
[146] S. Zheng, K. Ristovski, A. Farahat, C. Gupta, Long short-term memory network for remaining useful life estimation, in: 2017 IEEE international conference on prognostics and health management (ICPHM), IEEE, 2017, pp. 88–95.
[147] L. Wen, Y. Dong, L. Gao, A new ensemble residual convolutional neural network for remaining useful life estimation, Math. Biosci. Eng 16 (2019) 862–880.
[148] X. Chen, H. Wang, J. Huang, H. Ren, APU degradation prediction based on EEMD and Gaussian process regression, IEEE, pp. 98–104.
[149] X. Liu, L. Liu, L. Wang, Q. Guo, X. Peng, Performance sensing data prediction for an aircraft auxiliary power unit using the optimized extreme learning machine, Sensors 19 (2019) 3935.
[150] F. Wang, J. Sun, X. Liu, C. Liu, Aircraft auxiliary power unit performance assessment and remaining useful life evaluation for predictive maintenance, Proc. Inst. Mech. Eng., Part A: J. Power Energy 234 (2020) 804–816.
[151] Y. Zhang, Y. Peng, P. Wang, L. Wang, S. Wang, H. Liao, Aircraft APU failure rate prediction based on improved Weibull-based GRP, IEEE, pp. 1–6.
[152] Y. Zhang, D. Liu, J. Yu, Y. Peng, X. Peng, EMA remaining useful life prediction with weighted bagging GPR algorithm, Microelectron. Reliab. 75 (2017) 253–263.
[153] Y. Zhang, L. Liu, Y. Peng, D. Liu, An Electro-Mechanical Actuator motor voltage estimation method with a feature-aided Kalman Filter, Sensors 18 (2018) 4190.
[154] R. Guo, Z. Liu, J. Wang, Remaining useful life prediction for the electro-hydraulic actuator based on improved relevance vector machine, Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng. 234 (2020) 501–511.
[155] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, G. Nenadic, Machine learning methods for wind turbine condition monitoring: a review, Renewable Energy 133 (2019) 620–635.
[156] J. Carroll, S. Koukoura, A. McDonald, A. Charalambous, S. Weiss, S. McArthur, Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques, Wind Energy 22 (2019) 360–375.
[157] Z. Song, Z. Zhang, Y. Jiang, J. Zhu, Wind turbine health state monitoring based on a Bayesian data-driven approach, Renewable Energy 125 (2018) 172–181.
[158] F. Cheng, L. Qu, W. Qiao, L. Hao, Enhanced particle filtering for bearing remaining useful life prediction of wind turbine drivetrain gearboxes, IEEE Trans. Ind. Electron. 66 (2019) 4738–4748.
[159] Y. Hu, H. Li, P. Shi, Z. Chai, K. Wang, X. Xie, Z. Chen, A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process, Renewable Energy 127 (2018) 452–460.
[160] J.S. Nielsen, J.D. Sørensen, Bayesian estimation of remaining useful life for wind turbine blades, Energies 10 (2017) 664.
[161] L. Saidi, J.B. Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR, Appl. Acoust. 120 (2017) 1–8.
[162] H.D.M. de Azevedo, A.M. Araújo, N. Bouchonneau, A review of wind turbine bearing condition monitoring: state of the art and challenges, Renew. Sustain. Energy Rev. 56 (2016) 368–379.
[163] J.P. Salameh, S. Cauet, E. Etien, A. Sakout, L. Rambault, Gearbox condition monitoring in wind turbines: a review, Mech. Syst. Sig. Process. 111 (2018) 251–264.
[164] H. Wang, M.-J. Peng, Y.-K. Liu, S.-W. Liu, R.-Y. Xu, H. Saeed, Remaining useful life prediction techniques of electric valves for nuclear power plants with convolution kernel and LSTM, Sci. Technol. Nucl. Install. 2020 (2020).
[165] H. Wang, M.-J. Peng, Z. Miao, Y.-K. Liu, A. Ayodeji, C. Hao, Remaining useful life prediction techniques for electric valves based on convolution auto encoder and long short term memory, ISA Trans. (2020).
[166] J.I. Aizpurua, S.D.J. McArthur, B.G. Stewart, B. Lambert, J.G. Cross, V. M. Catterson, Adaptive power transformer lifetime predictions through machine learning and uncertainty modeling in nuclear power plants, IEEE Trans. Ind. Electron. 66 (2018) 4726–4737.
[167] H.-P. Nguyen, P. Baraldi, E. Zio, Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants, Appl. Energy 116346 (2020).
[168] Y. Li, K. Liu, A.M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, H.E. Hoster, Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review, Renew. Sustain. Energy Rev. 113 (2019), 109254.
[169] B. Saha, K. Goebel, Battery data set, NASA AMES prognostics data repository, 2007.
[170] P. Michael, Battery Data Set, CALCE Battery Research Group, Maryland, MD, 2017, 2017, pp. https://web.calce.umd.edu/batteries/index.html.
[171] W. He, N. Williard, M. Osterman, M. Pecht, Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method, J. Power Sources 196 (2011) 10314–10321.
[172] Q. Zhai, Z.-S. Ye, RUL prediction of deteriorating products using an adaptive Wiener process model, IEEE Trans. Ind. Inf. 13 (2017) 2911–2921.
[173] Y. Shen, L. Shen, W. Xu, A Wiener-based degradation model with logistic distributed measurement errors and remaining useful life estimation, Qual. Reliab. Eng. Int. 34 (2018) 1289–1303.
[174] H. Wang, X. Ma, Y. Zhao, A mixed-effects model of two-phase degradation process for reliability assessment and RUL prediction, Microelectron. Reliab. 107 (2020), 113622.
[175] Z.-X. Zhang, X.-S. Si, C.-H. Hu, M.G. Pecht, A prognostic model for stochastic degrading systems with state recovery: Application to Li-ion batteries, IEEE Trans. Reliab. 66 (2017) 1293–1308.
[176] X.-S. Si, An adaptive prognostic approach via nonlinear degradation modeling: application to battery data, IEEE Trans. Ind. Electron. 62 (2015) 5082–5096.
[177] L. Chen, Y. Zhang, Y. Zheng, X. Li, X. Zheng, Remaining useful life prediction of lithium-ion battery with optimal input sequence selection and error compensation, Neurocomputing 414 (2020) 245–254.
[178] Z. Xue, Y. Zhang, C. Cheng, G. Ma, Remaining useful life prediction of lithium-ion batteries with adaptive unscented kalman filter and optimized support vector regression, Neurocomputing 376 (2020) 95–102.
[179] P. Khumprom, N. Yodo, A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm, Energies 12 (2019) 660.
[180] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, L. Zhang, Remaining useful life prediction for lithium-ion battery: a deep learning approach, IEEE Access 6 (2018) 50587–50598.
[181] R. Jiao, K. Peng, J. Dong, Remaining useful life prediction of lithium-ion batteries based on conditional variational autoencoders-particle filter, IEEE Trans. Instrum. Meas. (2020).
[182] J.F. Murray, G.F. Hughes, K. Kreutz-Delgado, Machine learning methods for predicting failures in hard drives: a multiple-instance application, J. Mach. Learn. Res. 6 (2005) 783–816.
[183] B. Zhu, G. Wang, X. Liu, D. Hu, S. Lin, J. Ma, Proactive drive failure prediction for large scale storage systems, IEEE, pp. 1–5.
[184] N. Aussel, S. Jaulin, G. Gandon, Y. Petetin, E. Fazli, S. Chabridon, Predictive models of hard drive failures based on operational data, IEEE, pp. 619–625.
[185] C. Xu, G. Wang, X. Liu, D. Guo, T. Liu, Health status assessment and failure prediction for hard drives with recurrent neural networks, IEEE Trans. Comput. 65 (2016) 3502–3508.
[186] J. Li, R.J. Stones, G. Wang, X. Liu, Z. Li, M. Xu, Hard drive failure prediction using decision trees, Reliab. Eng. Syst. Saf. 164 (2017) 55–65.
[187] F.D.S. Lima, F.L.F. Pereira, L.G.M. Leite, J.P.P. Gomes, J.C. Machado, Remaining useful life estimation of hard disk drives based on deep neural networks, IEEE, pp. 1–7.
[188] A. Ruiz-Tagle Palazuelos, E.L. Droguett, R. Pascual, A novel deep capsule neural network for remaining useful life estimation, Proc. Inst. Mech. Eng. Part O: J. Risk Reliab. 234 (2020) 151–167.
[189] C. Zhang, P. Lim, A.K. Qin, K.C. Tan, Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics, IEEE Trans. Neural Networks Learn. Syst. 28 (2017) 2306–2318.
[190] Y. Liao, L. Zhang, C. Liu, Uncertainty prediction of remaining useful life using long short-term memory network based on bootstrap method, in: 2018 IEEE International Conference on Prognostics and Health Management (ICPHM), IEEE, 2018, pp. 1–8.
[191] X. Li, Q. Ding, J.-Q. Sun, Remaining useful life estimation in prognostics using deep convolution neural networks, Reliab. Eng. Syst. Saf. 172 (2018) 1–11.
[192] A.L. Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov, H. Zhang, Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture, Reliab. Eng. Syst. Saf. 183 (2019) 240–251.
[193] G. Hou, S. Xu, N. Zhou, L. Yang, Q. Fu, Remaining useful life estimation using deep convolutional generative adversarial networks based on an autoencoder scheme, Comput. Intell. Neurosci. 2020 (2020).
[194] M. Kim, K. Liu, A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics, IISE Trans. 53 (2020) 326–340.

[195] L. Ren, J. Cui, Y. Sun, X. Cheng, Multi-bearing remaining useful life collaborative prediction: a deep learning approach, J. Manuf. Syst. 43 (2017) 248–256.
[196] L. Liao, F. Köttig, A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction, Appl. Soft Comput. 44 (2016) 191–199.
[197] M. Djeziri, S. Benmoussa, R. Sanchez, Hybrid method for remaining useful life prediction in wind turbine systems, Renewable Energy 116 (2018) 173–187.