Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice

Enrico Zio

Prognostics and Health Management (PHM): where are we and where do we (need to) go in theory and practice

Enrico Zio

MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France enrico.zio@minesparitech.fr

Energy Department, Politecnico di Milano, Milan, Italy, enrico.zio@polimi.it

ABSTRACT

We are performing the digital transition of industry, living the 4th industrial revolution, building a new World in which the digital, physical and human dimensions are interrelated in complex socio-cyber-physical systems. For the sustainability of these transformations, knowledge, information and data must be integrated within model-based and data-driven approaches of Prognostics and Health Management (PHM) for the assessment and prediction of structures, systems and components (SSCs) evolutions and process behaviors, so as to allow anticipating failures and avoiding accidents, thus, aiming at improved safe and reliable design, operation and maintenance.

There is already a plethora of methods available for many potential applications and more are being developed: yet, there are still a number of critical problems which impede full deployment of PHM and its benefits in practice. In this respect, this paper does not aim at providing a survey of existing works for an introduction to PHM nor at providing new tools or methods for its further development; rather, it aims at pointing out main challenges and directions of advancements, for full deployment of condition-based and predictive maintenance in practice.

Keywords: *Prognostics and Health Management (PHM), predictive maintenance, Recurrent Neural Networks (RNNs), Reservoir Computing (RC), Generative Adversarial Networks (GANs), Deep Neural Networks (DNNs), Optimal Transport (OT)*

NOMENCLATURE

AAKR	Auto-Associative Kernel Regression
AANN	Auto-Associative Neural Networks
ADNN	Adjacency Difference Neural Network
AE	Auto-Encoder
AE-GAN	Auto-Encoder aided GAN
ALE	Accumulated Local Effect
ANNs	Artificial Neural Networks
ARM	Association Rule Mining
ARMA	Auto-Regressive Moving Average
BN	Bayesian Network
CatAAE	Categorical Adversarial Autoencoder
CBM	Condition-Based Maintenance
CDT	Cumulative Distribution Transform
CNN	Convolutional Neural Network
CVNN	Complex Valued Neural Network
DAE	Denoising Auto Encoder

DBN	Deep Belief Network
DL	Deep Learning
DNNs	Deep Neural Networks
DT	Decision Trees
ELM	Extreme Learning Machine
EM	Expectation Maximization
EMD	Earth Mover's distance
ESNs	Echo-State Networks
FCM	Fuzzy C-Means
FFT	Fast Fourier Transform
GANs	Generative Adversarial Networks
GLRT	Generalized Likelihood Radio Test
GRU	Gated Recurrent Unit
HI	Health Indicator
HMM	Hidden Markov Model
ICE	Individual Conditional Expectation
ICT	Information and Communication Technology
IoTs	Internet of Things
KD	Kantorovich distance
KF	Kalman Filtering
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LIME	Local Interpretable Model Explanation
LS	Least Square
LSTM	Long Short Term Memory
MAR	Missing At Random
ML	Machine Learning
MODE	Multi-Objective Differential Evolution
NPPs	Nuclear Power Plants
OC-SVM	One Class-Support Vector Machine
OT	Optimal Transport
OTT	Optimal Transport Theory
PCA	Principle Component Analysis
PDP	Partial Dependence Plot
PF	Particle Filtering
PFSA	Probabilistic Finite State Automation
PHM	Prognostics and Health Management
PPIs	Prognostic Performance Indicators
RC	Reservoir Computing
RF	Random Forest
RNNs	Recurrent Neural Networks
RUL	Remaining Useful Life
RVM	Relevance Vector Machine
SA	Sensitivity Analysis

SaNSDE	Self-adaptive Differential Evolution with Neighborhood Search
SC	Spectral Clustering
SOM	Self-Organizing Map
SPRT	Sequential Probability Ratio Test
SSCs	Structures, Systems and Components
STPN	Spatio-Temporal Pattern Network
SVM	Support Vector Machine
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution

1. INTRODUCTION

Prognostics and Health Management (PHM) is a computation-based paradigm that elaborates on physical knowledge, information and data [1] of structures, systems and components (SSCs) operation and maintenance, to enable detecting equipment and process anomalies, diagnosing degradation states and faults, predicting the evolution of degradation to failure so as to estimate the remaining useful life (Figure 1). The outcomes of the PHM elaboration are used to support condition-based and predictive maintenance decisions for the efficient, reliable and safe operations of SSCs [2]–[5]. In fact, the capability of deploying these maintenance strategies provides the opportunity of setting efficient, just-in-time and just-right maintenance strategies: in other words, providing the right part to the right place at the right time. This opportunity is big because doing this would maximize the production profits and minimize all costs and losses, including asset ones [6]. As a result, in the past decade PHM research and development has intensified, both in academia and industry, involving various disciplines of mathematics, computer science, operations research, physics, chemistry, materials science, engineering, etc. [7], [8].

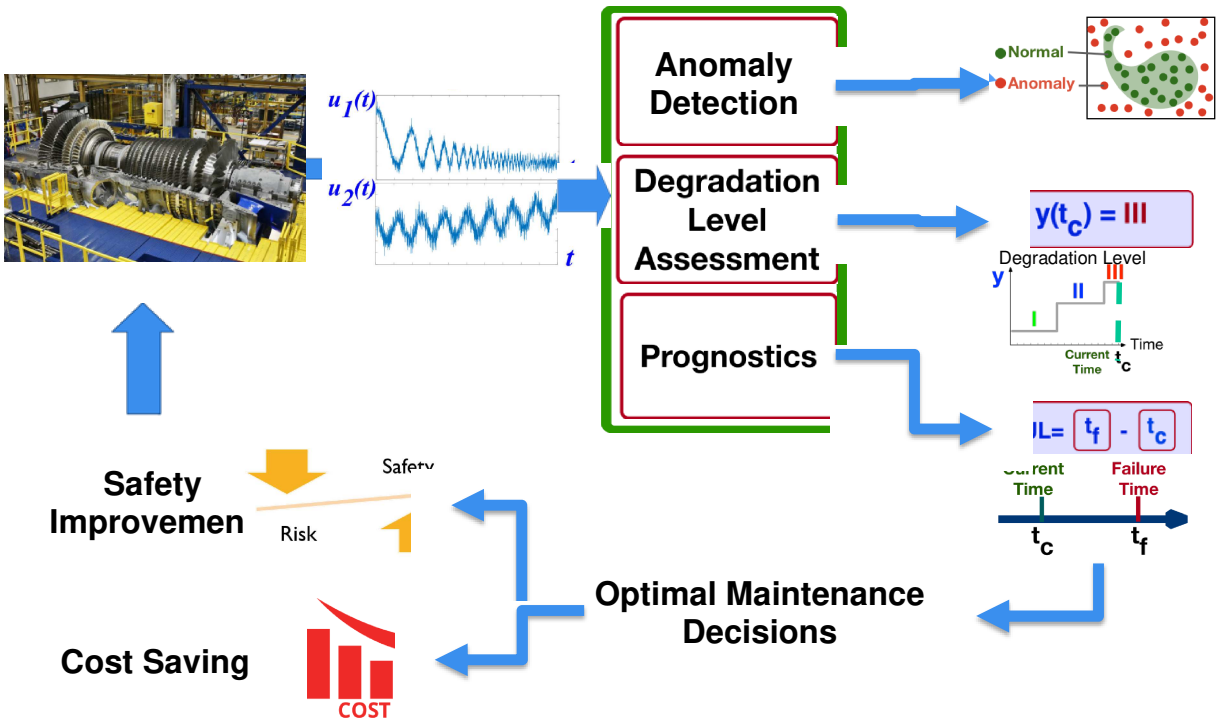

Figure 1. PHM tasks. The data collected from industrial component sensors feeds three major PHM tasks: fault detection (anomaly detection), fault diagnostics (degradation

level assessment) and fault prognostics (remaining useful life prediction). The successful deployment of PHM provides solid foundations for the optimal maintenance decisions, and thus improve the safety of industrial SSCs while reducing cost.

For making reliability and safety decision using PHM outcomes in practice, identifying, understanding and quantifying the impacts and benefits that the development of a PHM system can have on the health management of a SSC is necessary (e.g. avoid unexpected catastrophic failures, reduce maintenance frequency, optimize spare parts and storage, optimize resources, etc.). Then, the practical implementation of PHM includes data acquisition to enable detection, diagnostics and prognostics tasks, and maintenance decision making [9] (Figure 1). The supporting PHM development framework (Figure 2) and its requirements must, then, be properly defined to perform well in real industrial scenarios [9]–[11]. Given the increasing complexity, integration and informatization of modern engineering SSCs, PHM can no longer be an isolated addition for supporting maintenance but must be closely linked to the other structure, power, electromechanical, information and communication technology (ICT), control parts of the systems. Then, PHM must be included at the beginning of the system conceptualization, and carried through its design and development in an integrated framework capable of satisfying the overall operation and performance requirements [12], [13].

Finally, for the use of PHM in practice, the question of which methods to use is fundamental. For example, referring, in particular, to the prognostic task of PHM, the prediction capability of a prognostic method refers to its ability to provide trustable predictions of the Remaining Useful Life (RUL), with the quality characteristics and confidence level required for making decisions based on such predictions. Indeed, this heavily influences the decision makers' attitude toward taking the risk of using the predicted RUL outcomes to inform their decisions [14]. The choice of which method to use is typically driven by the data available and/or the physics-based models available, and the cost-benefit considerations related to the implementation of the PHM system. A set of Prognostic Performance Indicators (PPIs) must be used to guide the choice of the approach to be implemented, within a structured framework of evaluation. These PPIs measure different characteristics of a prognostic approach and need to be aggregated to enable a final choice of prognostic method, based on its overall performance [15]. For this reason, various performance metrics have been defined to enable the evaluation of the performance of PHM methods [16]. These metrics are needed to guide the PHM system development (Figure 2).

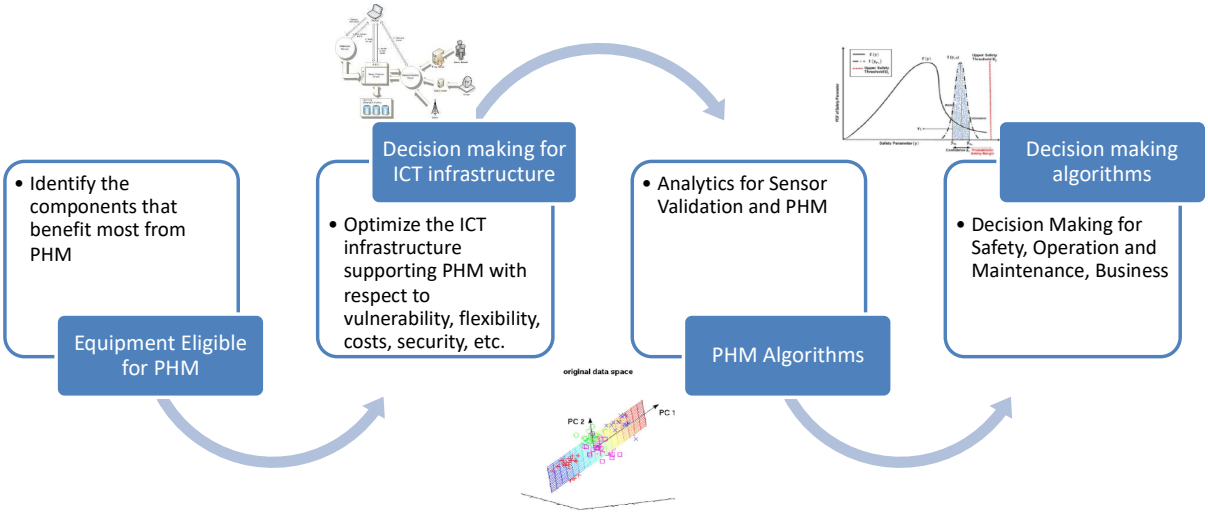

Figure 2 PHM development framework for informed decision-making

Up to now, for the maturation of PHM, the main efforts have been mainly devoted to the development of hardware (i.e., Internet of Things (IoTs), smart meters, etc. [17]–[20] and software for tracking the health state of monitored equipment (e.g., data analytics, platforms for IoT interconnection and clouding for computing, etc. [21]–[23]). On the other hand, the full deployment of PHM in practice involves other aspects, including design (e.g. the use of smart components may lead to different reliability allocation solutions), and impacts various work units involved in maintenance decisions and actuations (e.g., workers can use smart systems, maintenance engineers can analyze big data), including the supporting logistics (spare parts availability and warehouse management can be driven by the PHM results) [17].

In this paper, we present some main challenges for the development of PHM in practice, corroborated by practical examples, and associate to some of them the developments of Recurrent Neural Networks (RNNs), Reservoir Computing (RC), Generative Adversarial Networks (GANs), Deep Neural Networks (DNNs), Optimal Transport Theory (OTT), as potential directions to successfully address them.

2. CHALLENGES TO PHM IN PRACTICE

Main challenges to the deployment of PHM in practice still remain, coming from different sides:

- the physics of the problem
- the data available
- the requirements of the solutions.

The challenges related to the physics of the problem derive from the complexity of the SSCs degradation processes, which are not completely known, dynamic and highly non-linear, and hence their understanding, characterization and modelling are difficult.

The challenges related to the data relate to multiple aspects (Figure 3):

- the many anomalies in the real data collected in the field (including missing data and erroneous data from malfunctioning sensors)
- the scarcity and incompleteness of data recognizably related to the state of degradation of the SSC of interest (labelled patterns)
- the difficulty of managing and treating big data, with a large variety of signals collected by sensors of different types
- the changing operational and environmental conditions which affect the data used to train the PHM models and calibrate their parameters, and on which the models are applied.

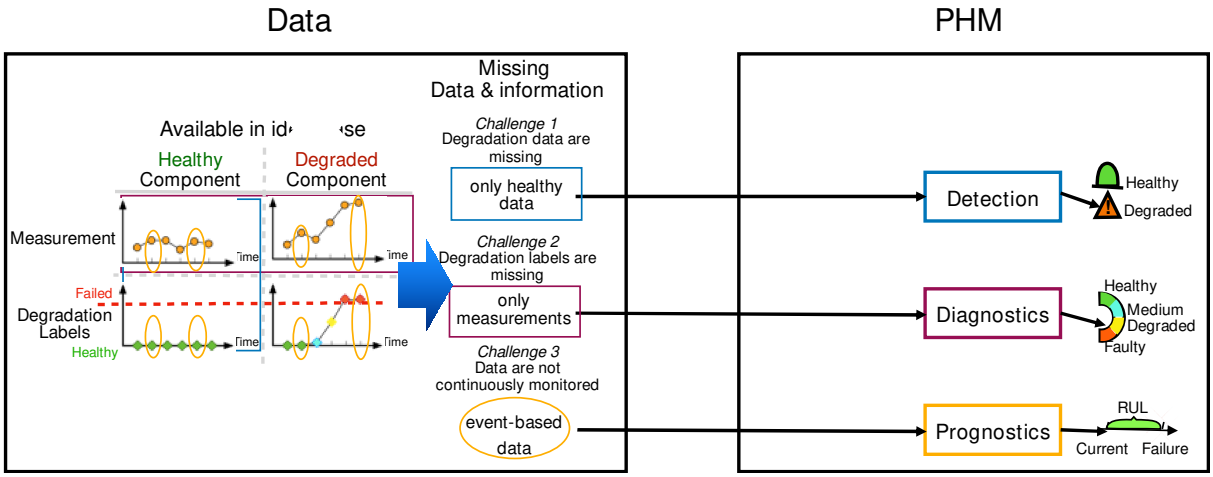

Figure 3 PHM challenges from data. Fault detection is affected by the challenge of missing data and erroneous data from malfunctioning sensors; fault diagnostics is affected by the challenge of missing labels relating to the state of degradation of the SSCs; fault prognostics is affected by the challenge that data are collected in event-based scenarios because of the difficulty of managing and treating big data.

The challenges related to the requirements of the PHM solutions come from the multiple objectives that they must achieve, depending on the applications. The obvious ones are accuracy and precision, quantified with defined performance indicators and measured against the decisions that they support: in some cases, very high accuracy and precision is required to be able to take confident decisions (e.g. of stopping a system upon an alert of fault detection, of replacing a component upon a fault diagnosis, of anticipating or postponing a scheduled maintenance based on accurate remaining useful life predictions); in other cases, accuracy and precision need not be so high, and may be compromised for other objectives. For example, transparency, explainability and interpretability of PHM models are attributes of particular interest, if not demanded, for decision making in safety-critical applications, for which they may also be a regulatory prerequisite. Also, PHM as a data-dependent enabling technology for smart condition-based and predictive maintenance has issues regarding security. Indeed, the technological network supporting PHM is made of devices, communication technologies and various protocols, so that security issues regarding availability, data integrity, data confidentiality and authentication exist. As these issues hamper operational efficiency, robustness and throughput, they must be adequately addressed.

Finally, an enveloping challenge to the deployment of PHM in practice comes from the fact that the PHM tasks of fault detection, diagnostic and prognostic are inevitably affected by various sources of uncertainty, such as incomplete knowledge on the present state of the equipment, randomness in the future operational usage profile and future evolution of the degradation of the equipment, inaccuracy of the PHM model and uncertainty in the values of the signal measurements used by the PHM model to elaborate its outcomes, etc. Therefore, any outcome of a PHM model should be accompanied by an estimate of its uncertainty, in order to confidently take robust decisions based on such outcome, considering the degree of mismatch between the PHM model outcomes and the real values.

As these issues hamper operational inefficiency, robustness and throughput, they must be adequately addressed.

With specific reference to data-driven methods and models for the tasks of fault detection, fault diagnostics and failure prognostics in PHM, the next section addresses some

of the above challenges with the focus on advanced methods that are proving as promising for their solution.

3. ADVANCING METHODS OF FAULT DETECTION, FAULT DIAGNOSTICS AND FAILURE PROGNOSTICS FOR MEETING THE CHALLENGES OF PHM IN PRACTICE

Methods of fault detection, fault diagnostics and failure prognostics within the PHM framework are continuously being developed and advanced, and applications to various SSCs are being deployed, supported by the technology of sensors and monitoring systems, the techniques of data analytics, image processing and text mining, mostly based on the Artificial Intelligence (AI) and Machine Learning (ML) paradigms, and the computational power [24]. The objective of fault detection is to recognize abnormaities/anomalies in SSCs behavior. The objective of fault diagnostics is to identify the SSCs degradation states and the causes of degradation. Prognostics aims at predicting the SSCs Remaining Useful Life (RUL), i.e. the time left before it will no longer be able to perform its intended function. Fault detection and diagnostics, and failure prognostics are the enablers of condition-based and predictive maintenance, which offers major opportunities for Industry 4.0 and smart SSCs, as they can allow reducing failures, increasing SSCs usage, and reducing operation and maintenance costs, with tangible benefits of reduction of production downtime, risk and asset losses, and consequent increase of production profit [24].

A number of challenges still remain, arising from the complexity of the physics which PHM is addressed to in practice, from the data available and from the requirements to the PHM solutions for practical applications. In this Section, we go through some of these challenges, to see where we stand, and where we are going and need to go.

3.1 Data Challenges

3.1.1 Fault detection

As mentioned above, fault detection is the PHM task which aims at identifying the presence of abnormalities/anomalies during the operation of a SSC. While such abnormalities/anomalies are commonly referred to as faults in certain disciplines, such as energy and mechanical engineering, the term damage is commonly used in some other disciplines such as structural engineering. In practical applications, fault/damage detection is challenging because it is necessary to assess the presence of the fault/damage based on signals of physical variables measured during the SSC operation and such process is complicated by the various sources of uncertainty that can render the signal processing extremely difficult.

Fault detection methods are classified as model-based and data-driven [25]. Model-based methods use first principles and physical laws to describe the physical phenomena and processes of interest [26][27][28]. For example, [26] builds a model of the behavior of a rotor using the finite element method and successfully applies it to fault detection. [27] introduces a model-based fault detection and isolation technique for manufacturing machinery based on a defined relationship between a fault signal and observer theory. [28] presents a two-level Bayesian approach based on the use of Hidden Markov Model (HMM) and Expectation Maximization (EM) to detect early faults in a milling machine. However, the practical application of model-based methods is limited by the difficulty of developing accurate mathematical models of the processes and behaviors of complex modern SSCs [29].

For this reason, data-driven fault detection methods are more popular than model-based ones, as they rely only on data for the recognition of anomalous patterns attributable to faults [30][31][32][33][34][35]. For example, [30] develops a fault detection method for power generation systems, by combining Principle Component Analysis (PCA) for feature

extraction and Random Forest (RF) for fault behavior pattern learning. Support Vector Machine (SVM) techniques are introduced to detect faults considering concept drift in nuclear power plants [31], and to detect faults in high speed train brake systems in case of highly imbalanced data [35]. Neural Network based approaches attract attention in fault detection, e.g. [32] combines a set of Artificial Neural Networks (ANNs) through Bayesian statistics for heavy-water nuclear reactor fault detection and uncertainty quantification, [34] uses ANN to detect false alarms in wind turbines for reliability centered maintenance, [33] introduces a Recurrent Neural Network (RNN) with optimized hyperparameters for the detection of software faults.

These methods can be divided in those which rely on one-class classification models and those which use residuals, i.e., the differences between the real measurements and the reconstructed values of the signals in normal conditions, to identify the normal/abnormal conditions [36].

The former require training of a one-class classification model on signal measurements collected from both normal (healthy) and abnormal/anomalous (faulty) conditions of SSCs. However, in practical applications, faults are rare and the data have manifold distributions embedded in high-dimensional spaces. Distributions with non-smooth densities and the curse of dimensionality of the data in the long-term multivariate time series collected from sensors on real industrial SSCs, can cause model overfitting and render difficult the empirical reconstruction of the data distribution, which, therefore, leads to unsuccessful detection of abnormal/anomalous conditions in SSCs behavior. These technical issues hamper the successful deployment of one-class classification methods for fault detection in practical applications. The need is, then, to develop methods able to detect anomalous (faulty) conditions given data in normal conditions, and to deal with the manifold distribution and large dimensionality of real data. In this direction, Generative Adversarial Networks (GANs) are an interesting perspective as they can be used to reproduce complex distributions, e.g. manifolds [37], [38]. An example is given in the work by [39], which proposes an Auto-Encoder aided GAN (AE-GAN) model for the detection of abnormal/anomalous conditions in the behavior of a SSC, in which the generator of the GAN and an auxiliary encoder form an AE module, and the reconstruction error generated by the AE is used as score to detect abnormalities/anomalies in the SSC behavior. Adaptive noise is added on the data and AdaBoost ensemble learning is adapted to integrate the AE-GANs applied to detect anomalies in each small time slice of the long-term multivariate time series collected by the sensors [40]. Furthermore, this work derives a lower bound of Jensen-Shannon divergence between generator distribution and normal data distribution as an objective to optimize the AE-GANs hyperparameters; by probing, the optimization works without test data, as commonly needed by other methods. Extensive experiments are conducted on real industrial datasets to demonstrate the usefulness of the developed Adaboost ensembled AE-GAN method for abnormality/anomaly detection in practice.

Residual-based fault detection methods rely on the use of normal-conditions (healthy) data, only [41]. These methods reconstruct the values of the signals expected in normal conditions and use the residuals, i.e., the differences between the real measurements and the reconstructed signals, to identify the normal/abnormal conditions. Examples of residual-based methods include Auto-Associative Kernel Regression (AAKR) [42]–[44], Principal Component Analysis (PCA) [45], One Class-Support Vector Machine (OC-SVM) [46], and Artificial Neural Networks (ANNs) [47]. The empirical model, fitted to the data so as to provide accurate signal reconstructions, plays an essential role in the above procedure. However, its training may require a large amount of healthy data collected under various operating conditions [48]. Besides, different choices of the reconstruction model may yield different detection results [49].

Eventually, the detection of an abnormal condition is confirmed by considering whether the obtained residuals exceed a threshold or by statistical tests. For example, [43] uses the Sequential Probability Ratio Test (SPRT) on the residuals obtained from an AAKR model; [50] applies T2- and Q-statistics of the PCA residuals to detect damages in structures; [51] establishes a statistical hypothesis model in the residual subspace of PCA transform, to detect and isolate sensor faults based on a Bayesian formulation and the generalized likelihood radio test (GLRT). Notice that, although these methods assume a certain distribution of the residuals, most distributions of real-world data may be a priori unknown or may not actually follow the assumed distributions [52].

Another challenge of fault detection lies in the data pre-processing [53] to extract features providing the information useful for enabling the detection. Various pre-processing techniques, such as Fast Fourier Transform (FFT) [54], Continuous Wavelet Transform [55], Mathematical Morphology [56], have been applied to raw signals, and the processed outcomes have been fed to fault detection [57]. The quality of the features selected by pre-processing strongly impacts the detection results, but unfortunately there is no universal rule for choosing the optimal pre-processing method.

Recently, transport-related methods are being considered for applications in PHM. They have already been successfully employed in other domains [58], involving signal and image processing [59], computer vision[60], machine learning and statistics [61], [62]. Commonly used optimal transport distances include Wasserstein distance (or Kantorovich distance) [63] and Earth Mover's distance (EMD) [64]. Wasserstein distance has proved a promising statistic for the nonparametric two-sample test [65].

In the PHM area, [66] has studied the bearing diagnostics problem using EMD combined with dynamical system reconstruction. [67] has used a PCA scheme combined with the Kantorovich distance (KD) for fault detection in the process industry. [68] has developed a method of OT in which the abnormality score is built using the Wasserstein distance and has verified its performance considering the detection of abnormal conditions in bearings. The method differs from other state-of-the-art methods for fault detection, since it directly deals with raw signals and does not require the use of signal reconstruction methods or feature extraction; it is also distribution-free, i.e., it does not require to formulate any a priori hypothesis on the distribution of the data. The basic idea behind the method is to generate an abnormality score, based on Wasserstein distance, to quantify the dissimilarity between the probability distributions of the currently monitored and healthy data. The Cumulative Distribution Transform (CDT) [69] is used to find the univariate Optimal Transport (OT) solution. The method has been applied to a real bearing dataset and successfully compared with two other fault detection methods of literature: a Z-test based method [70] and a PCA-based method for signal reconstruction, combined with the Q-statistic for residual analysis [71]. The Adaboost ensembled AE-GAN method mentioned earlier [39] can also be adapted for application to normal-conditions data only. The generator of the GAN and the auxiliary encoder form the AE module, and the reconstruction error generated by the AE is used as the score to detect abnormalities/anomalies in SSC behavior. For the abnormality/anomaly detection, it is assumed, as usual, that the probability distribution of the abnormal/anomalous-conditions data is significantly different from that of the normal-conditions data: as the generator can only reproduce the distribution of the normal-conditions data, the AE always successfully reconstruct the normal data but fails to reconstruct the abnormal ones. So, any test sample processed through the AE-GAN is declared anomalous if the AE reconstruction error is larger than a certain predefined threshold. For dealing with the high dimensionality of the data, again, an ensemble framework can be used. Non-overlapped sliding time windows are introduced to partition the multivariate time series and a separate data sample for each time window is analyzed by AE-GAN for abnormal/anomaly detection. Finally, the AdaBoost

algorithm is used to aggregate the abnormality/anomaly detection results for each time window. The GAN-based method for addressing the challenge of missing fault data in fault detection is shown in Figure 4.

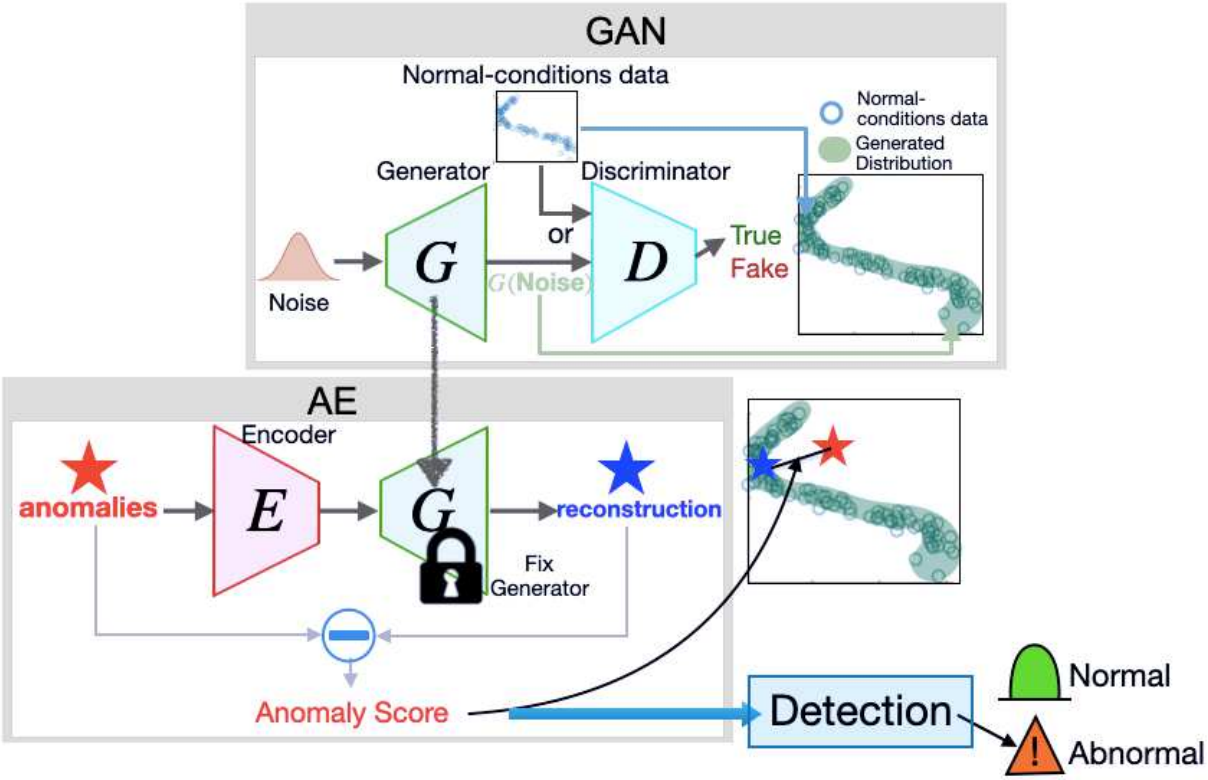

Figure 4 Illustration of GAN-based method in fault detection w.r.t. the challenge of missing fault data [39]. *GAN-based method is a type of distribution reconstruction method which reproduces the normal-conditions data distribution by the Generator and uses an extra Encoder to form an Auto-Encoder, which can obtain anomaly scores (reconstruction errors) to distinguish whether samples are anomalous or not.*

Table 1 summarizes the fault detection techniques, with specific regard to the challenge of missing fault data.

Table 1 Fault detection techniques with regard to the challenge of missing fault data.

Model-based	Data-driven			
	Supervised learning *(Limitation: need both healthy and fault condition data)*	One-class classification *(Advantage: addressing the challenge of missing fault data)*		
		Residual-based *(Limitation: need to apply statistical test)*	Transport-related	Distribution reconstruction-based
Finite Element Method: [26] rotor crack diagnostics	**RF:** [30] power generators fault detection	**AAKR:** [42] non-linear multimode processes fault detection [43] reactor coolant pump fault detetion [44] power plant	**EMD:** [66] bearing fault detection	**GAN-based:** [39] high-speed train automatic door

		fault detection		
Observer theory: [27] rotor fault detection	**SVM:** [31] early fault detection of numerical case [35] high-speed train brake fault detection	**PCA:** [45] air handling unit fault detection	**Kantorovich Distance:** [67] tank heater simulation case fault detection	
HMM: [28] mechanical equipment early fault detection	**ANN:** [32] heavy-water reactor early fault detection [34] wind turbine false alarm detection	**OC-SVM:** [46] building air conditioning system fault detection	**Wasserstein Distance:** [68] bearings fault detection	
	RNN: [33] software fault detection	**ANN:** [47] wind turbine gearbox fault detection	**CDT:** [69] numerical case	

3.1.2 Fault diagnostics

Fault diagnostics requires data analytics capable of identifying the equipment fault state, mode, location and other characteristics of interest, based on monitored signals (temperature, pressure, current, acceleration, etc.). As for the detection task previously discussed, in practical applications it also suffers from the presence of uncertainty coming from the processing of data of the measured signals. Common approaches make use of historical operational data to build empirical classifiers capable of discriminating different classes from the data, for fault diagnostics. Different classification techniques, such as Complex Valued Neural Network (CVNN) [72], Deep Belief Network (DBN) [73], Bayesian Network (BN) [74], Decision Trees (DT) [75], Linear Discriminant Analysis (LDA) [75], K-Nearest Neighbor (KNN) [75], Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) [76][75][77], have been successfully used in applications of different industrial and civil sectors [78]. These methods rely on supervised learning of labelled data, which, however, are rarely available in practice, so that their real application is limited [79]: the real application calls for unsupervised learning of unlabeled data.

Unsupervised learning is an important topic in machine learning for time series segmentation [80], [81] and pattern recognition [21], [82], [83]. In fault diagnostic applications, it is used to provide abstract representations of the raw measurement data and obtain various clusters representing healthy and faulty conditions [22], [84]–[86]. In the work of [22], a Categorical Adversarial Autoencoder (CatAAE) has been proposed for unsupervised learning aimed at fault diagnostics of rolling bearings. In the work of [84], a diagnostic methodology based on unsupervised Spectral Clustering (SC) combined with fuzzy C-means (FCM) has been developed for identifying groups of similar shutdown transients performed by a nuclear turbine. In [85], Self-Organizing Map (SOM) has been used for clustering and identifying degradation states of a railway-signal system. In [86], a methodology combining k-means and Association Rule Mining (ARM) has been developed to mine failure data and diagnose interconnections between failure occurrences in wind turbines. Representation learning can disentangle the different explanatory factors of variation behind the data, making it easier to extract and organize the discriminative information when building fault diagnostic models [87]–[92]. In traditional unsupervised methods for fault diagnostics, the features are

extracted applying ad hoc signal processing techniques to the collected signals, e.g. Fourier spectral analysis and Wavelet transformations [93]. The processing is heavily dependent on a priori knowledge and diagnostic expertise [22], [94], and can be quite time consuming and labor-intensive [88]. Since representation learning is adaptively capable of learning features from raw data, it can constitute an excellent a priori choice for the development of diagnostic techniques. In the work of [95], an unsupervised sparse filtering method based on a two-layer neural network is used to directly learn features from mechanical vibration signals. In the work of [96], a Spatio-Temporal Pattern Network (STPN) based on Probabilistic Finite State Automation (PFSA) and Markov machines is proposed to represent temporal and spatial structures for fault diagnostics in complex systems. However, these conventional representation learning methods cannot capture long-term temporal dependencies in the time series and they typically require high computational complexity.

From the above, it is seen that traditional fault diagnostic approaches typically require the acquisition of signal measurements from SSCs whose true degradation state is known. However, to acquire such labelled data is a difficult, expensive and labor-intensive task. Furthermore, streaming data collected in online-monitored SSCs have long-term temporal dependencies. However, unsupervised learning methods have a hard time dealing with long-term time dependencies, because these dependencies are limited by the size of the sliding time window which can be used for the analysis. Then, there is a need for advancements in the methods to estimate the degradation level at a given time on the basis of a few run-to-failure trajectories with long-term temporal dependencies and for which the true degradation state is unknown. In the work of [97], for example, a two-stage method for unsupervised learning is proposed for fault diagnostic applications, inspired by the idea of representing temporal patterns by a mechanism of neurodynamical pattern learning, called Conceptor. Considering a reservoir, i.e. a randomly generated and sparsely connected RNN [98], Conceptors can be understood as filters characterizing the geometries of the temporal states of the reservoir neurons in the form of square matrices [99], achieving a direction-selective damping of high-dimensional reservoir states [100]. The proposed method develops in two stages. In the first stage, the Conceptors extracted from the training run-to-failure degradation trajectories are clustered into several non-overlapped time series segments representing different degradation levels. In the second stage, the Conceptors and corresponding labels obtained in the first-stage clustering are used to train a Convolutional Neural Network (CNN) for real-time diagnosing the SSC degradation level. The CNN receives in input the Conceptors extracted from the reservoir states at the current time, which contain information about the long-term evolution of the SSC degradation, and the difference between the Conceptors extracted at the present and previous time steps, which contains information about the short-term degradation variation. The proposed method has been applied to two literature case studies concerning bearings fault diagnostics. The results show satisfactory accuracy and efficiency of the method. The Reservoir computing-based method for addressing the challenge of missing labels of degradation state is shown in Figure 5.

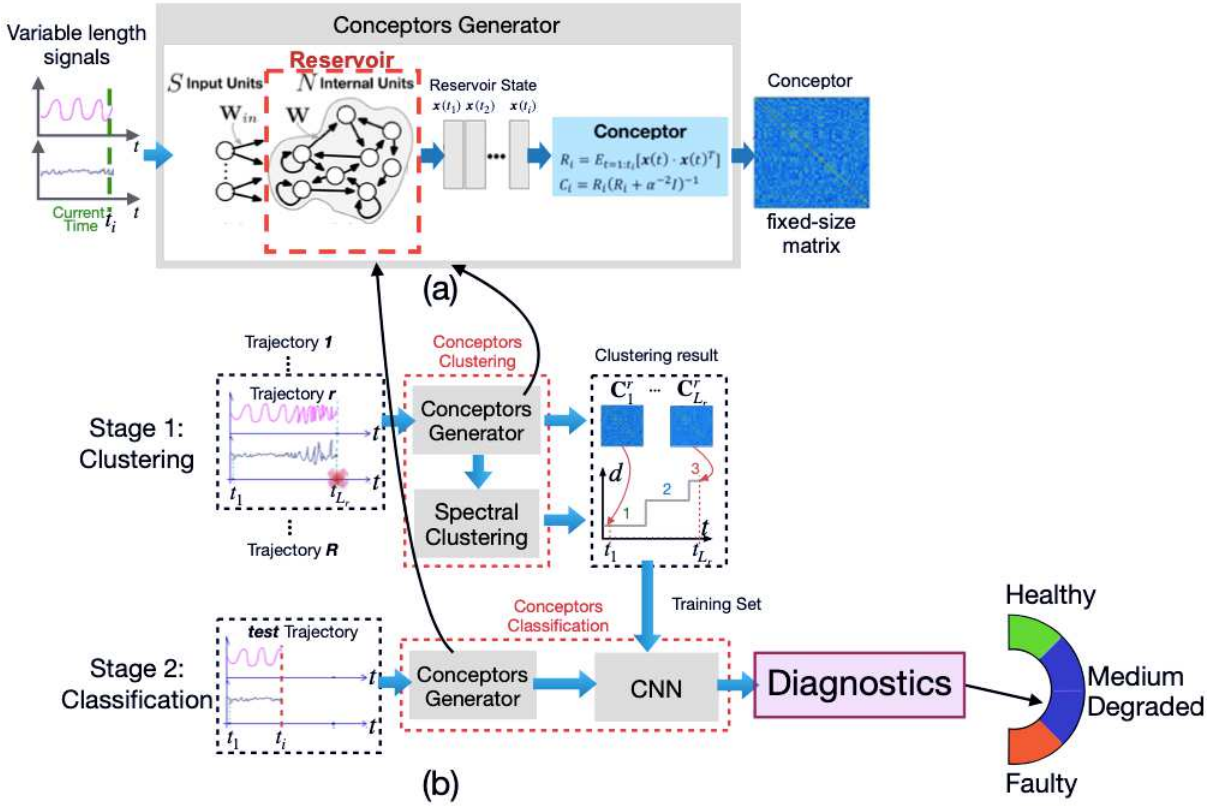

Figure 5 Illustration of Reservoir Computing-based method in fault diagnostics w.r.t. the challenge of missing labels of degradation state [97]. *a) Conceptor Generator converts the variable-length signals into a fixed-size Coceptor matrix by using reservoir computing. The Conceptor matrix is a filtered correlation matrix of reservoir states, which decreases the impact of less important features in the degradation dynamics, i.e. noise and operation conditions. b) stage 1: the combined use of Coceptors matrix and spectral clustering can obtain the pseudo-labels in the run-to-failure trajectories, stage 2:CNN trained by Conceptors with associated pseudo-labels is, then, used for fault diagnostics.*

Table 2 summarizes the fault diagnostics techniques, with specific regard to the challenge of missing labels for degradation states.

Table 2. Fault diagnostics techniques with regard to the challenge of missing labels for degradation state.

Supervised classifiers *(Limitation: need degradation state labels)*	Unsupervised approaches		
	Unsupervised learning *(Limitation: depend on diagnostic expertise for feature extraction)*	Representation learning	
		Neural Network-based *(Limitation: difficult to capture long-term temporal dependencies)*	Reservoir Computing-based
CVNN: [72] railway track turnouts degradation level assessment	**CatAAE:** [22] rolling bearing fault diagnositcs	**Unsupervised Sparse Filtering Neural Network:** [95] motor bearing fault diagnostics	**Conceptor + SC:** *(Conceptor as a representation capable of encoding long-term temporal*
DBN: [73] aircraft engine	**SC + FCM:** [84] Nuclear Power		

health state classification	Plant (NPP) steam turbine transient identification		*dependencies)* [97] degradation level assessment of bearings
BN: [74] failure type classification of water distribution system	**SOM:** [85] fault mode classification for railway monitoring equipment	**STPN:** [96] fault severities diagnostics of wind turbine	
DT: [75] anomaly (mud) diagnostic on wind turbine blade	**k-means + ARM:** [86] failure cause and weather condition correlation diagnostics of wind turbine system		
LDA: [75] anomaly (mud) diagnostic on wind turbine blade			
KNN: [75] anomaly (mud) diagnostic on wind turbine blade			
ANNs: [101] fault diagnostics of bearings			
SVMs: [75] anomaly (mud) diagnostic on wind turbine blade [76] bearing defects diagnostics [77] pipe failure prediction in water supply networks			

3.1.3 Fault prognostics

Prognostics is concerned with the prediction of the future evolution to failure of the state of a SSC. It involves the processing of data to predict the future degradation of the SSC structural and functional attributes, based on which to estimate the SSC failure probability and RUL. The prognostic outcomes are used for the health management of the SSC, which seeks to use the prognosis to decide on and actuate operational actions and maintenance interventions. To the uncertainties coming from the use of the data available from the sensors, like for the detection and diagnosis tasks, prognostics adds further challenges related to the future evolution of the usage profile and operational environment, whose uncertainties affect the degradation state evolution. This makes it practically impossible to precisely predict the future evolution of the SSC state of health and it is necessary to account for the different sources of uncertainty that affect prognostics, within a systematic framework for uncertainty quantification and management [102].

Prognostics is dependent on the available knowledge, information and data on the process of degradation. There may be situations in which a sufficient quantity of run-to-failure data has been collected during the life of the SSCs, and these can be used to develop empirical (data-driven) models. In other cases, the degradation mechanism is known and a physics-

based model is available. On these bases, prognostics approaches can be grouped into three categories: (i) model-based, (ii) data-driven and (iii) hybrid:

- (i) Model-based approaches use physics-based degradation models to predict the future evolution of the SSCs degradation state and infer the time at which the degradation will reach the failure threshold. These approaches have been applied with success in various practical cases, e.g., to pneumatic valves [103], Li-Ion batteries [104], the residual heat removal subsystem of a nuclear power plant [105], and structures subject to fatigue degradation [106]. In the case of complex SSCs, subject to multiple and competing degradation mechanisms, accurate physics-based models are, however, often not available.
- (ii) Data-driven approaches directly extract from the data the degradation law for SSCs RUL prediction [107]. Such approaches include conventional numerical time series techniques, as well as AI intelligence and data mining algorithms, such as similarity-based [108] and regression-based methods [107]. A variety of AI techniques, such as Convolutional Neural Network (CNN) [109][110], Denoising Auto Encoder (DAE) [111], Long Short Term Memory (LSTM) [112][109][113][114], Gated Recurrent Unit (GRU) [115], SVM [116], Adjacency Difference Neural Network (ADNN) [117], are applied to RUL estimation of different industrial systems and components. The performances of data-driven approaches depend on the quantity and informative quality of the data available to develop the predictive models.
- (iii) Hybrid approaches combine, all the available sources of knowledge, information and data. They bring the advantages of both model-based and data-driven methods. Specifically, they can integrate the robustness and interpretability of model-based methods with the specificity and accuracy of data-driven methods. For instance, [118] combines Kalman Filtering (KF) with data-driven approaches, [119] integrates the Health Indicator (HI) and regression model, [120] combines Relevance Vector Machine (RVM) and Particle Filtering (PF), and [121] integrates a physical model and the Least Square (LS) method to estimate RUL of a variety of industrial equipment.

Traditional fault prognostic methods face the challenge of dealing with incomplete and noisy data collected at irregular time steps, e.g. in correspondence of the occurrence of triggering events in the system. For example, for monitoring the degradation and failure processes of bearings in large turbine units, signal measurements collection (e.g., vibration signals measured by eddy current displacement sensors measuring the radial vibration of the rotor at both ends, the axial vibration of the rotor, and sensors measuring the unit rotating speed) is only triggered by abnormal behaviors of the units, such as large environmental noise and anomalous vibration behavior. These “snapshot” datasets are often encountered in industrial applications, dominated by the necessity of cost saving in storing and managing the databases, and of reducing energy consumption and bandwidth resources. Since failure events are rare, event-based datasets are dominated by missing measurements, where the values of all signals are missing at the same time. With these characteristics, traditional methods for missing data management, e.g. case deletion, imputation [122]–[125] and maximum likelihood estimation [126], are difficult to apply. For instance, since case deletion methods discard patterns whose information is incomplete, they are not useful in case of event-based datasets where a pattern is either present or absent for all signals [126]. Imputation techniques, which are based on the idea that a missing value of a signal can be replaced by a statistical indicator of the probability distribution generating the data, such as the signal mean value [127] or a value predicted by a multivariable regression model, have been shown inaccurate in case of large fractions of missing values in the dataset [128]. Maximum Likelihood methods use the available data to identify the values of the probability distribution parameters with the largest probability of producing the sample data. They typically require

the Missing At Random (MAR) assumption, i.e. the probability of having a missing value is not dependent on the missing values[127],[129], which is not met in event-based datasets.

Few research works have considered fault prognostics in presence of missing data. A model based on Auto-Regressive Moving Average (ARMA) and Auto-Associative Neural Networks (AANN), has been developed for fault diagnostics and prognostics of water process systems with incomplete data [130]. An integrated Extreme Learning Machine (ELM)-based imputation-prediction scheme for prognostics of battery data with missing data [125] and an hybrid architecture of physics-based and data-driven approaches have been proposed to deal with missing data in a rotating machinery prognostic application [131]. In the medical field, a Bayesian simulator has been used to generate missing data for developing prognostic models [132] and a Multiple Imputation approach has been embedded within a prognostic model for assessing overall survival of ovarian cancer in presence of missing covariate data [133]. Notice that all these methods are based on the two successive steps of missing data reconstruction and prediction.

Then, advancements and new methods are still needed to enable predicting the RUL of a SSC on the basis of measurements collected only when triggering events occur, such as SSC faults or extreme operational conditions, and providing an estimate of the uncertainty affecting the RUL prediction. As an example, [134] has developed a method based on Echo-State Networks (ESNs) to directly predict the RUL of a SSC without requiring to reconstruct the missing data. ESNs are considered because of their ability of maintaining information about the input history inside the reservoir states. The main difficulty is that, contrarily to the typical applications of ESNs, the time intervals at which the data become available are irregular. Two different strategies have been considered to cope with the event-based data collection. In one strategy, the ESN receives an input pattern only when an event occurs. The pattern is formed by the measured signals and the time at which the event has occurred. In a second strategy, the reservoir states are excited at each time step. If an event has occurred, the reservoir states are excited both by the previous reservoir states and the measured signals, whereas, if an event has not occurred, they are excited only by the previous reservoir states. By so doing, the connection loops in the reservoir allow reconstructing the SSC dynamic degradation behavior at those time steps in which events do not occur. Multi-Objective Differential Evolution (MODE) algorithm based on a Self-adaptive Differential Evolution with Neighborhood Search (SaNSDE) [135] is used to optimize the ESN hyper-parameters. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [136] is, then, used to select the optimal solution from the obtained Pareto solutions. Furthermore, a bootstrap aggregating (Bagging) ensemble method is applied to improve the RUL prediction accuracy and estimate the RUL prediction uncertainty. Given that ESNs cannot be fed by random sequences of patterns, the traditional Bagging sampling mechanism used to create the bootstrap training sets has been modified. In the proposed solution, the bootstrap training sets are obtained by concatenating entire run-to-failure trajectories, randomly sampled with replacement. The benefits of the proposed methods are shown by application to the prediction of the RUL of a sliding bearing of a turbine unit. The ESN-based one-step RUL prediction method for the challenge of missing data, i.e., event-based measurements, is shown in Figure 6.

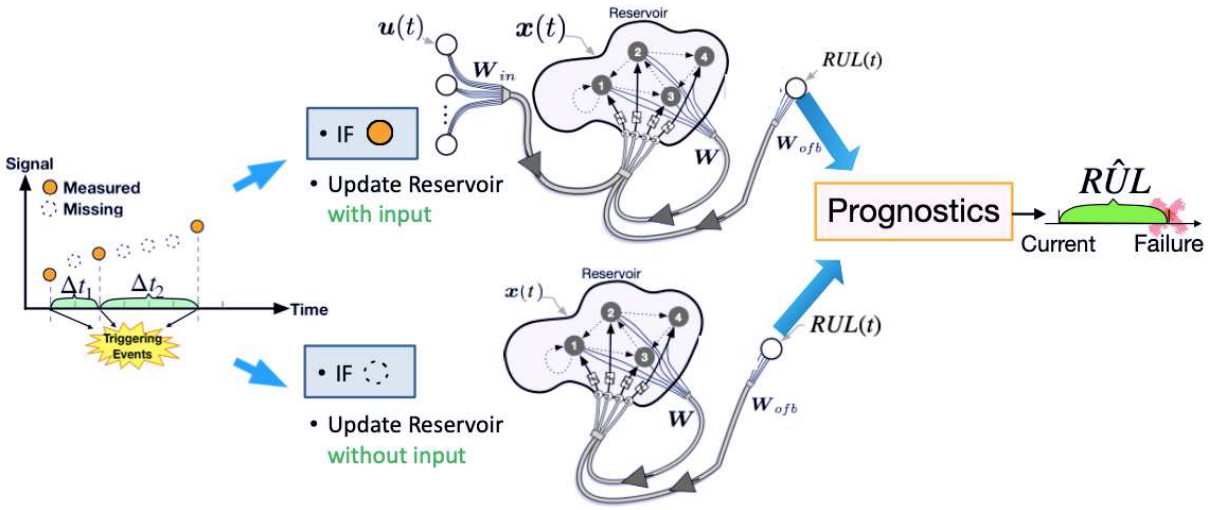

Figure 6 Illustration of ESN-based method in fault prognostics w.r.t. the challenge of missing data, i.e. event-based measurements [134]. *The input neurons of ESN are excited to update the reservoir state when measurements are available (events are triggered), whereas the input neurons are canceled if data are missing (no events occur) and the reservoir is only updated by the reservoir state at the previous time step and the target signal, which force the reservoir to learn from the historical degradation pattern and the target signal evolution pattern.*

Table 3 summarizes fault prognostics techniques, with specific regard to the challenge of missing data, i.e. event-based measurements.

Table 3. Fault prognostics techniques with regard to the challenge of missing data, i.e. event-based measurements.

<table>
<tr><td colspan="3">Traditional fault prognostics
(Limitations: cannot deal with missing data, i.e. event-based measurements)</td><td colspan="2">Fault prognostics in presence of missing data</td></tr>
<tr><td>Model-based</td><td>Data-driven</td><td>hybrid</td><td>Conventional type of missing data
(Limitations: difficult to deal with event-based measurements)</td><td>Event-based measurement</td></tr>
<tr><td rowspan="3">[103] fault prognostics of pneumatic valves
[104] prognostics and health monitoring of Lithium-ion battery
[105] fault</td><td>CNN:
[109] RUL estimation for bearing
[110] RUL estimation for turbofan engine</td><td rowspan="2">KF and data-driven approaches:
[118] RUL estimation for aircraft bleed valve</td><td>Missing at random</td><td rowspan="3">ESN-based one-step RUL prediction without requiring to reconstruct the missing data:
[134] RUL and uncertainty estimation of bearing</td></tr>
<tr><td rowspan="2">Missing data imputation and prognostics by ELM-based method:
[125] RUL estimation of battery</td></tr>
<tr><td>DAE:
[111] RUL estimation for</td><td>HI and regression model:</td><td>Missing not at random</td></tr>
</table>

prognostics in residual heat removal subsystem	centrifugal pumps	[119] RUL estimation for bearings	**Missing data prediction by Quasi-Newton Optimization; Prognostics by AANN and ARMA:** [130] prognostics for wastewater plant	
	LSTM: [109] RUL estimation for bearing [112] RUL estimation for turbofan engine [113] RUL estimation for turbofan engine [114] RUL estimation for turbofan engine	**RVM and PF:** [120] RUL estimation for Lithium-ion battery	Missing at extreme operating condition	
		physical model and LS: [121] RUL estimation for Lithium-ion battery	**Missing data generation by physical model; Prognostics by physical model and data-driven:** [131] prognostics for bearing	
	GRU: [115] RUL estimation for turbofan engine			
	SVM: [116] RUL estimation for aircraft engine			
	ADNN: [117] RUL estimation for aircraft engine			

3.2 Challenges from requirements on practical solutions

3.2.1 Interpretability of models

The ability to correctly interpret a PHM model's output, be it the detection of a fault, its diagnosis or prognosis, is extremely important, and particularly so in safety-critical applications like those concerning the high-risk systems and processes of the chemical, nuclear, aerospace industries, to name a few. It allows understanding of the state of the system or process being modeled and supports analytic reasoning and prescriptive decision making to intervene (or not) and how. It also engenders appropriate trust by the analyst, providing insights on how the model works. The importance of this is such that in some applications, simple models (e.g., even linear models) are preferred for their ease of interpretation, even if they may be less accurate than complex ones. Yet, currently the growing availability of big data for PHM has increased the benefits of using complex models for achieving accuracy, at the expenses of model intelligibility. This brings to the forefront the need of a trade-off between accuracy of the model and interpretability of its output. A wide variety of different

methods have been recently proposed to address this issue, but an understanding of how these methods relate and when one method is preferable to another is still lacking.

Most models and algorithms for PHM are developed and trained to maximize accuracy, neglecting interpretability and causality. Accounting for these aspects may, indeed, lead to a loss in performance but would enhance their safe, reliable and robust use both in terms of undesired biases and uncertainty reduction. Understanding why a PHM model makes a certain prediction can be as crucial as the prediction's accuracy in many applications. However, the highest accuracy for large modern datasets is often achieved by complex models that even experts struggle to interpret, such as ensemble or deep learning models, creating a tension between accuracy and interpretability [137]. Some general attributes sought for in the interpretability of PHM models are:

- fairness: no discrimination in algorithm decisions, which could come from bias in the collected data
- robustness: small changes in input should not cause big changes in output
- causality: causal relations are picked up from the model and rendered explicit
- quantifiable reliability of outcomes and predictions.

The awareness of the relevance of transparency, explainability and interpretability of PHM models is growing as a need and a requirement, particularly for supporting decision making in safety critical systems, for which it may also be a regulatory prerequisite. For example, in Nuclear Power Plants (NPPs), there is still resistance to the deep penetration of digital I&C systems and PHM, because of the difficulty of testing performance under all postulated conditions, on one side, and guaranteeing reliability based on transparent understanding and interpretation, on the other side. The decision making related to tasks of control, operation, maintenance and safety of NPPs, which have traditionally relied on procedures and expert evaluation and judgment, are gradually being assisted by intelligent machines (i.e. software algorithms) for PHM, developed and trained on the basis of big and customized data: how far and how it can be permitted in safety-critical systems that require licensing depends also on the possibility of interpreting the causality of their output.

For the modelling approaches to PHM based on learning from data, one issue lies in possible biases in the training set that are, then, not present in the test set or contain patterns undesired with respect to the test data, and may be unknown to the user of the trained model output. In this sense, achieving robustness in PHM models is fundamental and one way to proceed is to try to design inherently interpretable models, i.e. so as to exclude all undesired features that are not causally related to the outcome. By examining interpretable models:

- features or functions capturing quirks in the data can be noted and excluded, thereby avoiding related harm in the successive use of the model output, and the understanding of the phenomena analyzed
- knowledge can be extracted, in terms of the interactions among the inputs and how they determine the output
- an evaluation of the reliability of the PHM outcomes can be performed
- some limited extrapolation can be possible, with the aim of gaining knowledge on unexplored scenarios.

Methodologies are used to gain interpretability in a model by looking at the importance of the different input features in determining the model outputs. A distinction is made between model-specific and model-agnostic methodologies for evaluating feature importance. An interesting example of the former is the "attention mechanism" for Neural Networks applied in Prognostics, where importance values are assigned to specific input subsets [138], [139].

As the name implies, model-agnostic feature importance evaluation methodologies can in principle be used for any model. Local approaches are used for online applications and global approaches for offline applications. Local measures focus on the contribution of

features to a specific outcome instance, whereas global measures take all outcomes into account.

The Local Interpretable Model Explanation (LIME) method aims at explaining individual outputs and can be applied to any learning model [140]. Instead of training a global surrogate model, LIME focuses on training local surrogate models to explain individual model outputs. The method works by building for each output instance of interest a local-interpretable model that approximates the original, complex model. Each model output instance is, then, explained by an "explainer-model" that highlights the symptoms that are most relevant to it. With this information about the rationale behind the model, the analyst is now empowered to trust the model output—or not – for her/his decisions and consequent actions.

The idea behind LIME is quite intuitive and it is based on the fact that one can probe the model as many times as desired, by feeding the input data points and retrieving the corresponding outputs of the model. The goal of this is to understand why the learning model gave a certain output. The LIME tests are local sensitivity tests performed in a way to explore what happens to the output when the inputs are locally varied by small perturbations. By so doing, a new dataset is generated, consisting of permuted input samples and corresponding model outputs. For example, the new samples can be created by perturbing each feature individually, drawing from a normal distribution with mean and standard deviation taken from the feature values. On this new dataset, LIME builds and trains the interpretable explainer-model, which is weighed by the proximity of the sampled instances to the instance of interest. The interpretable model should give a good approximation of the original model outputs locally, but it does not have to be a good global approximation of the original model itself. Mathematically, the interpretable explainer model for instance x is the (simple) model g (e.g. a linear regression model) that results as solution of the optimization problem that minimizes the loss function L (e.g. the mean squared error) measuring how close the explanation output of g is to the output of the original model f (e.g. a neural network), while the model complexity $\Omega(g)$ is kept low (e.g. as few features as possible):

$$explanation(x)=\underset{g \in G}{argmin}\ L(f,g,\pi_x)+\Omega(g) \qquad (1)$$

where G is the family of possible explainer models, for example all possible linear regression models, and the proximity measure π_x defines how large is the neighborhood around instance x that is considered for the explanation. In practice, LIME only optimizes the loss part and the user controls the model complexity by $\Omega(g)$, e.g. by selecting by forward and backward feature selection methods the maximum number of features that the linear regression model may use.

The procedure for interpreting locally the complex original model is, then:

i) select the instance of interest x for which an explanation of the original complex model outcome $f(x)$ is needed
ii) perturb the input data and get the original model output values for these new data samples
iii) weigh the new samples according to their proximity to the instance of interest
iv) train a weighed, interpretable model on the new dataset generated in *ii)*
v) explain the local output of the interpretable model g.

LIME has been applied for the interpretation of machine learning models in applications of medical diagnostics [141]. In a recent study about early Parkinson detection, LIME has been used to highlight the features determining the healthy/disease decision of a ML classifier of images of the brain: LIME allows highlighting the super-pixels mostly determining the

classification in healthy or disease states; experts can, then, focus on the super-pixels selected with LIME to interpret and explain the basis for the decision by the ML algorithm, and choose to accept or refuse it.

Shapley values also can be used to assess local features importance [142]. Although they can be used to explain which feature(s) contribute most to a specific model output, Shapley values are not designed to answer the "what would happen if" questions that LIME's local explainer models are designed for. They come from game theory and are designed to construct a fair payout scheme for the players in a game. Suppose one could look at all possible combinations of (a subset of) players in a team replaying a game and observe the resulting team score. One could, then, assign each player of the team a portion of the total payout based on its average added value across all possible subteams to which it was added to play the game repeatedly. Such individual payout is the player's Shapley value and gives the only payout scheme that is proven to be:

- efficient: the sum of the Shapley values of all players should sum up to the total payout
- symmetric: two players should get the same payout if they add the same value in all team combinations
- dummy-sensitive: a player should get a Shapley value of zero if it never improves a subteam's performance when it is added
- additive: in case of a combined payout (say we add two game bonuses), the combined Shapley value of a player across the games is the sum of the individual game's Shapley values; this criterion has no relevant analogy in the context of model interpretability.

In the "game" of our interest for PHM model interpretability, the players are models with different features subsets and they get the same payout mechanism introduced above. The team score in this context is the performance measure of a (sub)model built on a given feature subset. The total payout is the difference between a base value — output of the null model — and the actual output. This difference is, then, divided over all features in accordance to their relative contribution.

Obviously looking at all possible subsets of features is computationally prohibitive in most realistic models with many features. Instead, Shapley value approximations can be computed based on sampling of features.

Other model-agnostic methodologies are based on Sensitivity Analysis (SA), which has been widely applied to models used in various areas, such as nuclear risk assessment [143], industrial bioprocessing [144] and climate change [145]. Indeed, a main application of SA is for identifying the input quantities most responsible of a given output variation [146]. Both local and global approaches to SA have been developed. Local approaches identify the critical input features as those whose variation leads to the most variation in the output. One practical approach for such identification consists in perturbing one single input at a time with small variations around its nominal value, while maintaining the others set at their respective nominal values. The analysis is intrinsically local and the resulting indication can be considered valid for the characterization of the model response around the nominal values. The possibility of extending the results of the analysis to draw global considerations on the model response over the whole input variability space depends on the model itself: if the model is linear or mildly non-linear, then the extension may be possible; if the model is strongly non-linear and characterized by sharp variations, the analysis is valid only locally. Typical local approach techniques are those based on Taylor's differential analysis and on the one-at-a-time simulation, in which the input features are varied one at a time while the others remain set at their nominal values [146].

In those situations (often encountered in practice) in which models are non-linear and non-monotone, the results provided by a local analysis may have limited significance. For this reason, global approaches to SA have been developed. In these approaches, the focus is directly on the uncertainty distribution of the output, which contains all the information about the variability of the model response, with no reference to any particular value of the input (like in the local approaches, where reference is made to the nominal values). The two principal characteristics of the global approaches are somewhat opposite to those of the local ones: 1) the account given to the whole variability range of the input features (and not only to small perturbations around the nominal values); 2) the focus on the effects resulting from considering also the variation of the other uncertain features (instead of keeping them fixed to their nominal values). Many global analysis methods have been developed [146]. The high capabilities of these methods are paid by a very high computational cost.

Another direction to build interpretability into PHM models and algorithms is by integrating prior physical knowledge in the learning models, for providing improved performance and achieving interpretability. This is a promising approach for inducing interpretability into the learning models and different approaches have been proposed where the physical knowledge can be introduced at different levels of the learning process, including in the training data and in the training algorithm [147][148][149][150].

To aid the interpretation of the model, there exists also a suite of methods for the visualization of the relations between input and output. The Partial Dependence Plot (PDP) shows the marginal effect that features have on the output provided by the model [151]. Intuitively, we can interpret the partial dependence as the expected target response as a function of the input features of interest. A partial dependence plot can show whether the relationship between the output and a feature is linear, monotonic or more complex. For example, when applied to a linear regression model, partial dependence plots always show a linear relationship. The computation of partial dependence plots is intuitive: the partial dependence function at a particular feature value represents the average output if we force all data to assume that value for the feature. If the feature for which the PDP is computed is not correlated with the other features, then the PDP perfectly represents how the feature influences the output on average. In the uncorrelated case, the interpretation is clear: the PDP shows how the average output changes when a given feature is changed. The interpretation is more complicated when features are correlated. Also, PDPs are easy to implement and the calculations to obtain them have a causal interpretation which aids model understanding: one intervenes on a feature and measures the corresponding change in the output. By doing so, one analyzes the causal relationship between the feature and the output in the model, and the relationship is causal for the model whose outcome is explicited as a function of the features. However, there are several disadvantages in PDPs. Due to the limits of human perception, the number of features in a partial dependence function must be small (usually, one or two) and, thus, the features considered must be chosen among the most important ones. Some PDPs do not show the feature distribution. Omitting the distribution can be misleading, because one might overinterpret regions with almost no data. This problem is easily solved by showing a rug (indicators for data points on the x-axis) or a histogram. Also, heterogeneous effects might be hidden because PDPs only show the average marginal effects. Suppose that for a feature, half of the input data has a positive correlation with the output (the larger the feature value the larger the output value) and the other half has a negative correlation (the smaller the feature value the larger the output value): then, PDP could be a horizontal line, since the effects of both halves of the dataset could cancel each other out and one would, then, conclude that the feature has no effect on the output. In other words, whereas the PDPs are good at showing the average effect of the target features, they can obscure a heterogeneous relationship created by interactions.

When interactions are present, the Individual Conditional Expectation (ICE) plot can be used to extract more insights [152]. An ICE plot shows the dependence between the output and an input feature of interest. However, unlike a PDP, which shows the average effect of the input feature, an ICE plot visualizes the dependence of the output on a feature for each sample separately, with one line per sample. Again, due to the limits of human perception, only one input feature of interest is supported by ICE plots. On the other hand, in ICE plots it might not be easy to see the average effect of the input feature of interest. Hence, it is recommended to use ICE plots alongside PDPs: they can be plotted together.

Finally, the assumption of independence is the biggest issue with PDPs. It is assumed that the features for which the partial dependence is computed are not correlated with other features. One solution to this problem is Accumulated Local Effect (ALE) plots that work with the conditional instead of the marginal distribution (Apley et al., 2020). ALE plots are a faster than and unbiased alternative to PDPs. Based on the conditional distribution of the features, they calculate differences in outputs instead of averages. ALE plots are unbiased, which means they still work when features are correlated, and are faster to compute than PDPs. The interpretation of ALE plots is also clear: conditional on a given value, the relative effect on the output due to changing the feature value can be read from the ALE plot. Even though ALE plots are not biased in case of correlated features, interpretation remains difficult when features are strongly correlated. Because if they have a very strong correlation, it only makes sense to analyze the effect of changing both features together and not in isolation. This disadvantage is not specific to ALE plots, but a general problem of strongly correlated features. Table 4 summarizes the investigated approaches for interpreting the PHM models.

Table 4 Summary of model interpretability approaches

Approaches	Characteristic	Reference
LIME	Focuces on training local surrogate models to explain individual model outputs	[140]
Shapley value	Uses game theory to construct a fair payout scheme for the player (input features) to obtain features importance	[137]
Sensitivity analysis-based	Identifyies the input quantities most responsible of a given output variation	[143][144][145][146]
Prior physical knowledge-based	Builds interpretability into PHM models and algorithms by integrating prior physical knowledge in the learning models	[147][148][149][150]
PDP	Shows the average marginal effect that features have on the output provided by the model	[151]
ICE	ICE plots visualize the dependence of the output on a feature, for each sample separately	[152]

3.2.2 Security of models

Applications of PHM methods for condition-based and predictive maintenance rely on the exchange and elaboration of data. The models and algorithms used are technological elements of larger socio-human-technical systems that must be engineered with safety and security in mind. They are increasingly used in support of high-value decision-making

processes in various industries, where the wrong decision may result in serious consequences. The underlying models and algorithms are largely unable to discern between malicious input and benign anomalous data. On the contrary, they should be capable of discerning maliciously-introduced data from benign "Black Swan" events. In particular, the learning models and algorithms should reject training data with negative impacts on results. Otherwise, learning models will always be susceptible to gaming by attackers. The specific danger is that an attacker will attempt to exploit the adaptive aspect of a learning model to cause it to fail and produce errors: if the model misidentifies an hostile input as benign, the hostile input is permitted through the security barrier; if it misidentifies a benign input as hostile, the good input is rejected [153]. The adversarial opponent has a powerful weapon: the ability to design training data that cause the learning model to produce rules that misidentify inputs. To avoid this, the models and algorithms used for PHM must have built-in forensic capabilities [154]. These should enable a form of intrusion detection, allowing engineers to determine the exact point in time that an output was given by the model, what input data influenced it and whether or not that data was trustworthy. The data visualization capabilities for the interpretation of the relations between model input features and model output discussed in the previous subsection 3.2.1 show promise to help engineers identify and resolve root causes for these complex issues. Also, specific solutions are required in the areas of Authentication, Input Validation and Denial of Service.

3.2.3 Uncertainty

Uncertainty is intrinsically present in the PHM tasks of detection, diagnostics and prognostics, and may adversely affect their outcomes, so to lead to an imprecise assessment of the state and prediction of the behavior of such systems, which could lead to wrongly informed system health management decisions with possibly costly, if not catastrophic, consequences. For practical deployment, it is necessary to be able to estimate the uncertainty and confidence in the outcomes of detection, diagnostics and prognostics activities, for quantifying the risk associated to the PHM decision-making on the operation of engineering systems. Yet, in spite of the recognition of the importance of uncertainty in PHM [155], work is still needed to concretely address the impact of uncertainty on the different PHM tasks and to effectively manage it.

The challenge comes from the fact that there are different sources of uncertainty that affect PHM, whose interactions are not fully understood and, thus, it is difficult to systematically account for them in the PHM tasks. While some sources are internal to the SSC, others are external, and all must be accounted for in the different activities of PHM. There is aleatory uncertainty in the physical behavior of the SSC and epistemic uncertainty in the model of it (developed based on sensors data or physic-based or based on a hybrid combination of both data and physics) and the associated parameters. As mentioned earlier, there is uncertainty in the sensors measurements and in their processing tools. For the prognostic task of PHM, there is also uncertainty on the future SSC operation profile and state evolution.

Given the relevance of uncertainty in the PHM tasks, it becomes necessary to develop systematic frameworks for accounting for such uncertainty in practical applications, in order to enable the robust verification and validation of the solutions developed, with respect to the requirements for their use for decision-making and their contribution to the risk involved in such decisions. Such frameworks must enable the systematic identification, representation, quantification and propagation of the different sources of uncertainty, so that any PHM outcome is provided also with its uncertainty, which needs to be considered for robust decision-making [156].

Focusing specifically on data-driven methods for PHM, the challenge of quantifying the uncertainty in PHM outcomes has rarely been addressed and mostly with ensemble approaches, which can become computationally burdensome, and are highly dependent on how the individual models are developed and how their outcomes are aggregated [157][158][159][160][134][161]. Recently, Bayesian neural networks and variational inference have been used in PHM, for accounting of uncertainty [162][163]. Also, the combination of neural networks and gaussian processes are being considered as a promising direction for providing PHM outcomes equipped with the needed estimates of the associated uncertainty [164].

4. CONCLUSIONS

PHM has become a fashionable area of research and development, due to its promises of enabling condition-based and predictive maintenance, which can be game-changers for the production performance, reliability and safety of industrial businesses. Then, many academic words have been and are developed, and several applications have been attempted, with a more or less significant degree of success. These have been facilitated by the availability of numerous and large data sets, of affordable computational hardware to train the models, of freely available software to implement the models in a reliable and relatively straightforward manner. Yet, quite some work still needs to be done to increase the significance of PHM impacts on industry, due to a number of theoretical and practical issues that still require an effective solution. These come from different perspectives, related to the physics of the problem itself, the nature and type of data, the requirements of the solutions. As for the physics of the problem, it is undoubtful that the SSCs degradation processes in practice are most of the times quite complex and dependent on a large number of parameters and mechanisms, which are dynamic and highly non-linear, and not completely known.

But much of the problem comes from the data and the extraction of informative content for the fault detection, diagnostics and prognostics tasks of PHM. Managing and treating the big condition-monitoring data collected by the sensors and comprised of a large variety of heterogeneous signals is not an easy task and the data are often anomalous, scarce, incomplete and unlabeled. Furthermore, they are collected under changing operational and environmental conditions during the life of the SSC.

Surely, for the effectiveness of extracting informative content from data, undoubtedly Deep Learning (DL) has contributed a great leap by incorporating feature engineering in the process of learning of the models, for automatic processing of big and heterogeneous condition monitoring data and extraction of features relevant for the application. Encouraging results have been obtained already in fault detection and diagnostics, whereas Prognostics remains still a challenge for DL.

Other of the above challenges are being addressed with sophisticated advancements which need to be, then, effectively deployed in practice. These include: Recurrent Neural Networks for PHM applications, and their transformation into images so as to exploit the powerful methods of image processing (including the novel Convolutional Neural Networks (CNNs), particularly for fault detection and diagnostics; signal reconstruction methods

(including Auto-Encoders) of unsupervised and semi-supervised learning for fault detection and diagnostics, and for degradation state prediction, to cope with the frequent practical cases of unlabeled data; Optimal Transport (OT) methods and unsupervised adaptation techniques to cope with the problem that the test data distribution may be a different distribution (or evolve to a different distribution) than that of the training data, with the consequence that the trained data-learned model may perform poorly on the test data.

An issue of particular relevance for the prognostic task of PHM is the proper treatment of the uncertainty in the data and, then, in the models. Several sources of uncertainty exist in practice, as the models are inevitably only representations of the real relationships between input and output, the measured data are inevitably noisy due to measurement errors, and the future operational and environmental profiles of the SSCs are not known. All these uncertainties affect the predictions of the future degradation and failure of the SSCs. With respect to the uncertainty issue in PHM, frameworks are being developed for a probabilistic treatment of the RUL of SSCs: given the potentially costly and catastrophic consequences associated with the decisions that are made based on the PHM outcomes, it is obsolutely necessary to provide also estimates of the uncertainty alongside the predictions. Fore example, frameworks are being developed by Bayesian neural networks and deep gaussian processes.

An issue which is arising with the data-driven models and algorithms used for PHM is that they lack interpretability, which reduces trust in their use particularly for safety-critical applications. This leads to the need to find ways for improving transparency an interpretability for a clearer understanding of what the model predicts and how, and finally for building trust on its use. Methods for injecting physical information in learning models, post-hoc sensitivity approaches and visualization techniques are being studied to provide interpretability from different perspectives, including explaining the learned input-output relation representations, explaining the individual model outputs, explaining the way the output is produced by the model.

Strong concerns are also arising with respect to the security of PHM models for real applications, in particular for safety-critical ones. PHM is increasingly used to support maintenance decision-making processes in various high-value/high-risk industries, where the wrong decision may result in serious consequences. The methods and models used perform exchange and elaboration of data, and must then be secure to reject training data with negative impacts on the results of decision-making.

Acknowledgments
The author is grateful to Dr. Mingjing Xu for helping in the preparation and revision of the paper, and to the four referees whose comments and questions have helped to greatly improve the work.

REFERENCES

[1] E. Zio, "Some challenges and opportunities in reliability engineering," *IEEE Transactions on Reliability*, 2016, doi: 10.1109/TR.2016.2591504.

[2] S. T. Kandukuri, A. Klausen, H. R. Karimi, and K. G. Robbersmyr, "A review of diagnostics and prognostics of low-speed machinery towards wind turbine farm-level health management," *Renewable and Sustainable Energy Reviews*. 2016, doi: 10.1016/j.rser.2015.08.061.

[3] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*. 2018, doi: 10.1016/j.ymssp.2017.11.016.

[4] M. Tahan, E. Tsoutsanis, M. Muhammad, and Z. A. Abdul Karim, "Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review," *Applied Energy*. 2017, doi: 10.1016/j.apenergy.2017.04.048.

[5] D. Wang, K. L. Tsui, and Q. Miao, "Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators," *IEEE Access*, 2017, doi: 10.1109/ACCESS.2017.2774261.

[6] M. Compare, P. Baraldi, and E. Zio, "Challenges to IoT-Enabled Predictive Maintenance for Industry 4.0," *IEEE Internet of Things Journal*, 2020, doi: 10.1109/JIOT.2019.2957029.

[7] C. L. Gan, "Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things," *Life Cycle Reliability and Safety Engineering*, 2020, doi: 10.1007/s41872-020-00119-y.

[8] H. M. Elattar, H. K. Elminir, and A. M. Riad, "Prognostics: a literature review," *Complex & Intelligent Systems*, 2016, doi: 10.1007/s40747-016-0019-3.

[9] I. K. Jennions, O. Niculita, and M. Esperon-Miguez, "Integrating IVHM and asset design," *International Journal of Prognostics and Health Management*, 2016.

[10] J. R. Dumargue, T., Pougeon, J. R., & Massé, "An approach to designing PHM systems with systems engineering," in *European conference of the Prognostics and health management society.*, 2016.

[11] M. Sharp and B. A. Weiss, "Hierarchical modeling of a manufacturing work cell to promote contextualized PHM information across multiple levels," *Manufacturing Letters*, 2018, doi: 10.1016/j.mfglet.2018.02.003.

[12] D. Han, J. Yu, Y. Song, D. Tang, and J. Dai, "A distributed autonomic logistics system with parallel-computing diagnostic algorithm for aircrafts," in *AUTOTESTCON (Proceedings)*, 2019, doi: 10.1109/AUTEST.2019.8878478.

[13] L. Yang, Q. Sun, and Z. S. Ye, "Designing mission abort strategies based on early-warning information: Application to UAV," *IEEE Transactions on Industrial Informatics*, 2020, doi: 10.1109/TII.2019.2912427.

[14] F. Di Maio, P. Turati, E. Z. PHM Society European Conference 2016, "Prediction capability assessment of data-driven prognostic methods for railway applications," *phmsociety.org*.

[15] Z. Zeng, F. Di Maio, E. Zio, and R. Kang, "A hierarchical decision-making framework for the assessment of the prediction capability of prognostic methods," *J Risk and Reliability*, vol. 231, no. 1, pp. 36–52, Feb. 2017, doi: 10.1177/1748006X16683321.

[16] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Metrics for Offline Evaluation of Prognostic Performance," *International Journal of Prognostics and Health Management*, 2010.

[17] D. Kwon, M. R. Hodkiewicz, J. Fan, T. Shibutani, and M. G. Pecht, "IoT-Based Prognostics and Systems Health Management for Industrial Applications," *IEEE Access*, 2016, doi: 10.1109/ACCESS.2016.2587754.

[18] J. Siryani, B. Tanju, and T. J. Eveleigh, “A Machine Learning Decision-Support System Improves the Internet of Things’ Smart Meter Operations,” *IEEE Internet of Things Journal*, 2017, doi: 10.1109/JIOT.2017.2722358.
[19] L. Winnig, “GE’s big bet on data and analytics,” *Proc. MIT Sloan Manag. Rev.*, 2016.
[20] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge Computing: Vision and Challenges,” *IEEE Internet of Things Journal*, 2016, doi: 10.1109/JIOT.2016.2579198.
[21] K. Lin, J. Lu, C. S. Chen, J. Zhou, and M. T. Sun, “Unsupervised Deep Learning of Compact Binary Descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, doi: 10.1109/TPAMI.2018.2833865.
[22] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, “Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks,” *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.07.034.
[23] L. Chen, G. Xu, Y. Wang, and J. Wang, “Detection of weak transient signals based on unsupervised learning for bearing fault diagnosis,” *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.07.004.
[24] K. Goebel, *Prognostics: The Science of Making Predictions*, 1st ed. CreateSpace Independent Publishing Platform, 2017.
[25] E. Zio, “Computational methods for reliability and risk analysis,” in *Computational methods for reliability and risk analysis*, World Scientific Publishing Company, 2009.
[26] A. S. Sekhar, “Model-based identification of two cracks in a rotor system,” *Mechanical Systems and Signal Processing*, 2004, doi: 10.1016/S0888-3270(03)00041-4.
[27] S. L. Jeong Haedong, Bumsoo Park, Seungtae Park, Hyungcheol Min, “Fault detection and identification method using observer-based residuals,” *Reliability Engineering & System Safety*, vol. 184, pp. 27–40, 2019.
[28] Duan, Chaoqun, Viliam Makis, and Chao Deng. "A two-level Bayesian early fault detection for mechanical equipment subject to dependent failure modes." *Reliability Engineering & System Safety* 193 (2020): 106676.
[29] H. P. Wan and Y. Q. Ni, “Bayesian multi-task learning methodology for reconstruction of structural health monitoring data,” *Structural Health Monitoring*, 2019, doi: 10.1177/1475921718794953.
[30] J. A. A. J. et al. Quintanilha, Igor M., Vitor RM Elias, Felipe B. da Silva, Pedro AM Fonini, Eduardo AB da Silva, Sergio L. Netto, “A fault detector/classifier for closed-ring power generators using machine learning,” *Reliability Engineering & System Safety*, no. 107614, 2021.
[31] M. Marseguerra, “Early detection of gradual concept drifts by text categorization and Support Vector Machine techniques: The TRIO algorithm,” *Reliability Engineering & System Safety*, vol. 129, pp. 1–9, 2014.
[32] and E. P. Tolo, Silvia, Xiange Tian, Nils Bausch, Victor Becerra, T. V. Santhosh, Gopika Vinod, “Robust on-line diagnosis tool for the early accident detection in nuclear power plants,” *Reliability Engineering & System Safety*, vol. 186, pp. 110–119, 2019.
[33] Hu, Q. P., Min Xie, Szu Hui Ng, and Gregory Levitin. "Robust recurrent neural network modeling for software fault detection and correction prediction." *Reliability Engineering & System Safety* 92, no. 3 (2007): 332-340.
[34] Marugán, Alberto Pliego, Ana María Peco Chacón, and Fausto Pedro García Márquez. "Reliability analysis of detecting false alarms that employ neural networks: A real case study on wind turbines." *Reliability Engineering & System Safety* 191 (2019): 106574.
[35] J. Liu, Y. F. Li, and E. Zio, “A SVM framework for fault detection of the braking system in a high speed train,” *Mechanical Systems and Signal Processing*, vol. 87, no. October 2016, pp. 401–409, 2017, doi: 10.1016/j.ymssp.2016.10.034.

[36] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*. 2016, doi: 10.1016/j.jnca.2015.11.016.
[37] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S. K. Ng, "MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, doi: 10.1007/978-3-030-30490-4_56.
[38] H.-J. Xing and W.-T. Liu, "Robust AdaBoost based ensemble of one-class support vector machines," *Information Fusion*, vol. 55, no. August 2019, pp. 45–58, 2019, doi: 10.1016/j.inffus.2019.08.002.
[39] M. Xu, P. Baraldi, X. Lu, F. Cannarile, and E. Zio, "Anomaly Detection for Industrial Systems using Generative Adversarial Networks," in *4th International Conference on System Reliability and Safety (ICSRS 2019)*, 2019.
[40] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class AdaBoost," *Statistics and Its Interface*, 2009, doi: 10.4310/sii.2009.v2.n3.a8.
[41] S. Joe Qin, "Data-driven fault detection and diagnosis for complex industrial processes," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2009, doi: 10.3182/20090630-4-ES-2003.0408.
[42] J. Yu, J. Yoo, J. Jang, J. H. Park, and S. Kim, "A novel hybrid of auto-associative kernel regression and dynamic independent component analysis for fault detection in nonlinear multimode processes," *Journal of Process Control*, 2018, doi: 10.1016/j.jprocont.2018.05.004.
[43] F. Di Maio, P. Baraldi, E. Zio, and R. Seraoui, "Fault detection in nuclear power plants components by a combination of statistical methods," *IEEE Transactions on Reliability*, 2013, doi: 10.1109/TR.2013.2285033.
[44] P. Baraldi, F. Di Maio, P. Turati, and E. Zio, "Robust signal reconstruction for condition monitoring of industrial components via a modified Auto Associative Kernel Regression method," *Mechanical Systems and Signal Processing*, 2015, doi: 10.1016/j.ymssp.2014.09.013.
[45] S. Li and J. Wen, "A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform," *Energy and Buildings*, 2014, doi: 10.1016/j.enbuild.2013.08.044.
[46] K. Yan, Z. Ji, and W. Shen, "Online fault detection methods for chillers combining extended kalman filter and recursive one-class SVM," *Neurocomputing*, 2017, doi: 10.1016/j.neucom.2016.09.076.
[47] P. Bangalore, S. Letzgus, D. Karlsson, and M. Patriksson, "An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox," *Wind Energy*, 2017, doi: 10.1002/we.2102.
[48] P. F. Odgaard, B. Lin, and S. B. Jorgensen, "Observer and data-driven-model-based fault detection in power plant coal mills," *IEEE Transactions on Energy Conversion*, 2008, doi: 10.1109/TEC.2007.914185.
[49] C. Yang, J. Liu, Y. Zeng, and G. Xie, "Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model," *Renewable Energy*, 2019, doi: 10.1016/j.renene.2018.10.062.
[50] L. E. Mujica, J. Rodellar, A. Fernández, and A. Güemes, "Q-statistic and t2-statistic pca-based measures for damage assessment in structures," *Structural Health Monitoring*, 2011, doi: 10.1177/1475921710388972.

[51] H.-B. Huang, T.-H. Yi, and H.-N. Li, "Sensor Fault Diagnosis for Structural Health Monitoring Based on Statistical Hypothesis Test and Missing Variable Approach," *Journal of Aerospace Engineering*, 2017, doi: 10.1061/(asce)as.1943-5525.0000572.
[52] S. Seo and P. D. Gary M. Marsh, "A review and comparison of methods for detecting outliersin univariate data sets," *Department of Biostatistics, Graduate School of Public Health*, 2006.
[53] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7. pp. 1483–1510, 2006, doi: 10.1016/j.ymssp.2005.09.012.
[54] B. Wang, D. Tang, Q. Yue, J. Zhou, and N. Deonauth, "Study on nonlinear dynamic characteristics inherent in offshore jacket platform using long-term monitored response of ice-structure interaction," *Applied Ocean Research*, 2018, doi: 10.1016/j.apor.2017.12.009.
[55] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using continuous wavelet transform," *Applied Soft Computing*, vol. 11, no. 2, pp. 2300–2312, 2011, doi: 10.1016/j.asoc.2010.08.011.
[56] A. S. Raj and N. Murali, "Early classification of bearing faults using morphological operators and fuzzy inference," *IEEE Transactions on Industrial Electronics*, 2013, doi: 10.1109/TIE.2012.2188259.
[57] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing," *Machines*. 2017, doi: 10.3390/machines5040021.
[58] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal Mass Transport: Signal processing and machine-learning applications," *IEEE Signal Processing Magazine*, 2017, doi: 10.1109/MSP.2017.2695801.
[59] P. Li, Q. Wang, and L. Zhang, "A novel earth mover's distance methodology for image matching with gaussian mixture models," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, doi: 10.1109/ICCV.2013.212.
[60] J. Rabin, S. Ferradans, and N. Papadakis, "Adaptive color transfer with relaxed optimal transport," in *2014 IEEE International Conference on Image Processing, ICIP 2014*, 2014, doi: 10.1109/ICIP.2014.7025983.
[61] G. Montavon, K. R. Müller, and M. Cuturi, "Wasserstein training of restricted boltzmann machines," in *Advances in Neural Information Processing Systems*, 2016.
[62] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*. 2012.
[63] M. Cuturi and A. Doucet, "Fast computation of Wasserstein barycenters," in *31st International Conference on Machine Learning, ICML 2014*, 2014.
[64] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proceedings of the IEEE International Conference on Computer Vision*, 2009, doi: 10.1109/ICCV.2009.5459199.
[65] A. Ramdas, N. G. Trillos, and M. Cuturi, "On wasserstein two-sample testing and related families of nonparametric tests," *Entropy*, 2017, doi: 10.3390/e19020047.
[66] S. S. Y. Ng, J. Cabrera, P. W. T. Tse, A. H. Chen, and K. L. Tsui, "Distance-based analysis of dynamical systems reconstructed from vibrations for bearing diagnostics," *Nonlinear Dynamics*, 2015, doi: 10.1007/s11071-014-1857-4.
[67] S. Kammammettu and Z. Li, "Change point and fault detection using Kantorovich Distance," *Journal of Process Control*, 2019, doi: 10.1016/j.jprocont.2019.05.012.

[68] B. Wang, P. Baraldi, X. Lu, and E. Zio, "Fault detection based on optimal transport theory," in *Proceedings of ESREL 2020 – PSAM 15*, 2020.
[69] S. R. Park, S. Kolouri, S. Kundu, and G. K. Rohde, "The cumulative distribution transform and linear pattern classification," *Applied and Computational Harmonic Analysis*, 2018, doi: 10.1016/j.acha.2017.02.002.
[70] M. Panda and P. M. Khilar, "Distributed soft fault detection algorithm in wireless sensor networks using statistical test," in *Proceedings of 2012 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, PDGC 2012*, 2012, doi: 10.1109/PDGC.2012.6449816.
[71] D. Garcia-Alvarez, "Fault detection using Principal Component Analysis (PCA) in a Wastewater Treatment Plant (WWTP)," in *62th International Student's Scientific Conference*, 2009.
[72] O. Fink, E. Zio, and U. Weidmann, "Predicting component reliability and level of degradation with complex-valued neural networks," *Reliability Engineering and System Safety*, vol. 121, pp. 198–206, Jan. 2014, doi: 10.1016/j.ress.2013.08.004.
[73] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering and System Safety*, vol. 115, pp. 124–135, Jul. 2013, doi: 10.1016/j.ress.2013.02.022.
[74] K. Tang, D. J. Parsons, and S. Jude, "Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system," *Reliability Engineering and System Safety*, vol. 186, pp. 24–36, Jun. 2019, doi: 10.1016/j.ress.2019.02.001.
[75] A. Arcos Jiménez, C. Q. Gómez Muñoz, and F. P. García Márquez, "Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers," *Reliability Engineering and System Safety*, vol. 184, pp. 2–12, Apr. 2019, doi: 10.1016/j.ress.2018.02.013.
[76] M. M. Manjurul Islam and J. M. Kim, "Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector Machines," *Reliability Engineering and System Safety*, vol. 184, pp. 55–66, Apr. 2019, doi: 10.1016/j.ress.2018.02.012.
[77] A. Robles-Velasco, P. Cortés, J. Muñuzuri, and L. Onieva, "Prediction of pipe failures in water supply networks using logistic regression and support vector classification," *Reliability Engineering and System Safety*, vol. 196, p. 106754, Apr. 2020, doi: 10.1016/j.ress.2019.106754.
[78] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, "Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications," *IEEE Access*, 2017, doi: 10.1109/ACCESS.2017.2731945.
[79] P. Baraldi, F. Di Maio, M. Rigamonti, E. Zio, and R. Seraoui, "Unsupervised clustering of vibration signals for identifying anomalous conditions in a nuclear turbine," *Journal of Intelligent and Fuzzy Systems*, 2015, doi: 10.3233/IFS-141459.
[80] R. Arn, P. Narayana, B. Draper, T. Emerson, M. Kirby, and C. Peterson, "Motion Segmentation via Generalized Curvatures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, doi: 10.1109/TPAMI.2018.2869741.
[81] F. Cakir, K. He, S. A. Bargal, and S. Sclaroff, "Hashing with Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, doi: 10.1109/tpami.2019.2914897.
[82] Y. Zheng, S. Li, R. Yan, H. Tang, and K. C. Tan, "Sparse Temporal Encoding of Visual Features for Robust Object Recognition by Spiking Neurons," *IEEE Transactions on Neural Networks and Learning Systems*, 2018, doi: 10.1109/TNNLS.2018.2812811.

[83] C. Cao, Y. Huang, Y. Yang, L. Wang, Z. Wang, and T. Tan, “Feedback Convolutional Neural Network for Visual Localization and Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, doi: 10.1109/TPAMI.2018.2843329.
[84] P. Baraldi, F. Di Maio, M. Rigamonti, E. Zio, and R. Seraoui, “Clustering for unsupervised fault diagnosis in nuclear turbine shut-down transients,” *Mechanical Systems and Signal Processing*, vol. 58, pp. 160–178, 2015, doi: 10.1016/j.ymssp.2014.12.018.
[85] C. Bian, S. Yang, T. Huang, Q. Xu, J. Liu, and E. Zio, “Degradation state mining and identification for railway point machines,” *Reliability Engineering and System Safety*, vol. 188, pp. 432–443, Aug. 2019, doi: 10.1016/j.ress.2019.03.044.
[86] M. Reder, N. Y. Yürüşen, and J. J. Melero, “Data-driven learning framework for associating weather conditions and wind turbine failures,” *Reliability Engineering and System Safety*, vol. 169, pp. 554–569, Jan. 2018, doi: 10.1016/j.ress.2017.10.004.
[87] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, doi: 10.1109/TPAMI.2013.50.
[88] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, “An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data,” *IEEE Transactions on Industrial Electronics*, 2016, doi: 10.1109/TIE.2016.2519325.
[89] W. Yang, Y. Shi, Y. Gao, L. Wang, and M. Yang, “Incomplete-data oriented multiview dimension reduction via sparse low-rank representation,” *IEEE Transactions on Neural Networks and Learning Systems*, 2018, doi: 10.1109/TNNLS.2018.2828699.
[90] M. Zhang, N. Wang, Y. Li, and X. Gao, “Deep Latent Low-Rank Representation for Face Sketch Synthesis,” *IEEE Transactions on Neural Networks and Learning Systems*, 2019, doi: 10.1109/TNNLS.2018.2890017.
[91] C. Li, M. Z. Zia, Q. H. Tran, X. Yu, G. D. Hager, and M. Chandraker, “Deep Supervision with Intermediate Concepts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, doi: 10.1109/TPAMI.2018.2863285.
[92] O. Costilla-Reyes, R. Vera-Rodriguez, P. Scully, and K. B. Ozanyan, “Analysis of Spatio-Temporal Representations for Robust Footstep Recognition with Deep Residual Neural Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, doi: 10.1109/TPAMI.2018.2799847.
[93] S. Yin, S. X. Ding, X. Xie, and H. Luo, “A review on basic data-driven approaches for industrial process monitoring,” *IEEE Transactions on Industrial Electronics*. 2014, doi: 10.1109/TIE.2014.2301773.
[94] H. Zhu, L. Lu, J. Yao, S. Dai, and Y. Hu, “Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model,” *Solar Energy*, 2018, doi: 10.1016/j.solener.2018.10.054.
[95] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, “An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016, doi: 10.1109/TIE.2016.2519325.
[96] T. Han, C. Liu, L. Wu, S. Sarkar, and D. Jiang, “An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems,” *Mechanical Systems and Signal Processing*, 2019, doi: 10.1016/j.ymssp.2018.07.048.
[97] Xu, M., P. Baraldi, and E. Zio. "Fault diagnostics by conceptors-aided clustering." *30th European Safety and Reliability Conference, ESREL* 2020 *and 15th Probabilistic Safety Assessment and Management Conference, PSAM* 2020.

[98] M. Lukoševičius and H. Jaeger, “Reservoir computing approaches to recurrent neural network training,” *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009, doi: 10.1016/j.cosrev.2009.03.005.
[99] H. Jaeger, “Using Conceptors to Manage Neural Long-Term Memories for Temporal Patterns,” *Machine Learning*, vol. 18, pp. 1–43, 2016.
[100] Qian, Guangwu, and Lei Zhang. "A simple feedforward convolutional conceptor neural network for classification." *Applied Soft Computing* 70 (2018): 1034-1041.
[101] B. Samanta and K. R. Al-Balushi, “Artificial neural network based fault diagnostics of rolling element bearings using time-domain features,” *Mechanical Systems and Signal Processing*, vol. 17, no. 2, pp. 317–328, 2003, doi: 10.1006/mssp.2001.1462.
[102] S. Sankararaman and K. Goebel, “Uncertainty in Prognostics and Systems Health Management,” *International Journal of Prognostics and Health Management*, 2015, doi: 10.36001/ijphm.2015.v6i4.2319.
[103] Daigle, Matthew J., and Kai Goebel. "A model-based prognostics approach applied to pneumatic valves." *International journal of prognostics and health management* 2, no. 2 (2011): 84-99.
[104] J. Zhang and J. Lee, “A review on prognostics and health monitoring of Li-ion battery,” *Journal of Power Sources*. 2011, doi: 10.1016/j.jpowsour.2011.03.101.
[105] J. Liu and E. Zio, “System dynamic reliability assessment and failure prognostics,” *Reliability Engineering and System Safety*, vol. 160, pp. 21–36, Apr. 2017, doi: 10.1016/j.ress.2016.12.003.
[106] J. M. W. Brownjohn, A. de Stefano, Y. L. Xu, H. Wenzel, and A. E. Aktan, “Vibration-based monitoring of civil infrastructure: Challenges and successes,” *Journal of Civil Structural Health Monitoring*, 2011, doi: 10.1007/s13349-011-0009-5.
[107] X. S. Si, W. Wang, C. H. Hu, and D. H. Zhou, “Remaining useful life estimation - A review on the statistical data driven approaches,” *European Journal of Operational Research*. 2011, doi: 10.1016/j.ejor.2010.11.018.
[108] Y. Liu, X. Hu, and W. Zhang, “Remaining useful life prediction based on health index similarity,” *Reliability Engineering and System Safety*, vol. 185, pp. 502–510, May 2019, doi: 10.1016/j.ress.2019.02.002.
[109] X. Li, W. Zhang, and Q. Ding, “Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction,” *Reliability Engineering and System Safety*, vol. 182, pp. 208–218, Feb. 2019, doi: 10.1016/j.ress.2018.11.011.
[110] X. Li, Q. Ding, and J. Q. Sun, “Remaining useful life estimation in prognostics using deep convolution neural networks,” *Reliability Engineering and System Safety*, 2018, doi: 10.1016/j.ress.2017.11.021.
[111] R. He, Y. Dai, J. Lu, and C. Mou, “Developing ladder network for intelligent evaluation system: Case of remaining useful life prediction for centrifugal pumps,” *Reliability Engineering and System Safety*, vol. 180, pp. 385–393, Dec. 2018, doi: 10.1016/j.ress.2018.08.010.
[112] A. Listou Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov, and H. Zhang, “Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture,” *Reliability Engineering and System Safety*, vol. 183, pp. 240–251, Mar. 2019, doi: 10.1016/j.ress.2018.11.027.
[113] K. T. P. Nguyen and K. Medjaher, “A new dynamic predictive maintenance framework using deep learning for failure prognostics,” *Reliability Engineering and System Safety*, vol. 188, pp. 251–262, Aug. 2019, doi: 10.1016/j.ress.2019.03.018.
[114] Z. Shi and A. Chehade, “A dual-LSTM framework combining change point detection and remaining useful life prediction,” *Reliability Engineering and System Safety*, vol. 205, p. 107257, Jan. 2021, doi: 10.1016/j.ress.2020.107257.

[115] J. Chen, H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Reliability Engineering and System Safety*, vol. 185, pp. 372–382, May 2019, doi: 10.1016/j.ress.2019.01.006.
[116] P. J. García Nieto, E. García-Gonzalo, F. Sánchez Lasheras, and 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," *Reliability Engineering and System Safety*, vol. 138, pp. 219–231, Jun. 2015, doi: 10.1016/j.ress.2015.02.001.
[117] Z. Zhao, Bin Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Reliability Engineering and System Safety*, vol. 164, pp. 74–83, Aug. 2017, doi: 10.1016/j.ress.2017.02.007.
[118] M. Baptista, E. M. P. Henriques, I. P. P. de Medeiros, J. P. P. Malere, C. L. L. Nascimento, and H. Prendinger, "Remaining useful life estimation in aeronautics: Combining data-driven and Kalman filtering," *Reliability Engineering and System Safety*, vol. 184, pp. 228–239, Apr. 2019, doi: 10.1016/j.ress.2018.01.017.
[119] W. Ahmad, S. A. Khan, M. M. M. Islam, and J. M. Kim, "A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models," *Reliability Engineering and System Safety*, vol. 184, pp. 67–76, Apr. 2019, doi: 10.1016/j.ress.2018.02.003.
[120] Y. Chang and H. Fang, "A hybrid prognostic method for system degradation based on particle filter and relevance vector machine," *Reliability Engineering and System Safety*, vol. 186, pp. 51–63, Jun. 2019, doi: 10.1016/j.ress.2019.02.011.
[121] A. Downey, Y. H. Lui, C. Hu, S. Laflamme, and S. Hu, "Physics-based prognostics of lithium-ion battery using non-linear least squares with dynamic bounds," *Reliability Engineering and System Safety*, vol. 182, pp. 1–12, Feb. 2019, doi: 10.1016/j.ress.2018.09.018.
[122] I. Eekhout, R. M. de Boer, J. W. R. Twisk, H. C. W. de Vet, and M. W. Heymans, "Missing data: a systematic review of how they are reported and handled.," *Epidemiology (Cambridge, Mass.)*, 2012, doi: 10.1097/EDE.0b013e3182576cdb.
[123] M. Ranjbar, P. Moradi, M. Azami, and M. Jalili, "An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems," *Engineering Applications of Artificial Intelligence*, 2015, doi: 10.1016/j.engappai.2015.08.010.
[124] Y. S. CH Cheng, CP Chan, "A novel purity-based k nearest neighbors imputation method and its application in financial distress prediction," *Engineering Applications of Artificial Intelligence*, vol. 81, pp. 283–299, 2019.
[125] R. Razavi-Far, S. Chakrabarti, M. Saif, and E. Zio, "An integrated imputation-prediction scheme for prognostics of battery data with missing observations," *Expert Systems with Applications*, 2019, doi: 10.1016/j.eswa.2018.08.033.
[126] A. N. Baraldi and C. K. Enders, "An introduction to modern missing data analyses," *Journal of School Psychology*, 2010, doi: 10.1016/j.jsp.2009.10.001.
[127] A. R. T. Donders, G. J. M. G. van der Heijden, T. Stijnen, and K. G. M. Moons, "Review: A gentle introduction to imputation of missing values," *Journal of Clinical Epidemiology*, 2006, doi: 10.1016/j.jclinepi.2006.01.014.
[128] D. Vergouw *et al.*, "The search for stable prognostic models in multiple imputed data sets," *BMC Medical Research Methodology*, 2010, doi: 10.1186/1471-2288-10-81.
[129] J. Honaker and G. King, "What to do about missing values in time-series cross-section data," *American Journal of Political Science*, vol. 54, no. 2, pp. 561–581, 2010, doi: 10.1111/j.1540-5907.2010.00447.x.

[130] H. Xiao, D. Huang, Y. Pan, Y. Liu, and K. Song, “Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and ARMA model,” *Chemometrics and Intelligent Laboratory Systems*, 2017, doi: 10.1016/j.chemolab.2016.12.009.
[131] U. Leturiondo, O. Salgado, L. Ciani, D. Galar, and M. Catelani, “Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data,” *Measurement: Journal of the International Measurement Confederation*, 2017, doi: 10.1016/j.measurement.2017.02.003.
[132] A. Marshall, D. G. Altman, P. Royston, and R. L. Holder, “Comparison of techniques for handling missing covariate data within prognostic modelling studies: A simulation study,” *BMC Medical Research Methodology*, 2010, doi: 10.1186/1471-2288-10-7.
[133] T. G. Clark and D. G. Altman, “Developing a prognostic model in the presence of missing data: An ovarian cancer case study,” *Journal of Clinical Epidemiology*, 2003, doi: 10.1016/S0895-4356(02)00539-5.
[134] M. Xu, P. Baraldi, S. Al-Dahidi, and E. Zio, “Fault prognostics by an ensemble of Echo State Networks in presence of event based measurements,” *Engineering Applications of Artificial Intelligence*, 2020, doi: 10.1016/j.engappai.2019.103346.
[135] Z. Yang, K. Tang, and X. Yao, “Self-adaptive differential evolution with neighborhood search,” in *2008 IEEE Congress on Evolutionary Computation, CEC 2008*, 2008, doi: 10.1109/CEC.2008.4630935.
[136] Yoon, K. Paul, and Ching-Lai Hwang. *Multiple attribute decision making: an introduction.* Sage publications, 1995.
[137] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *In Proceedings of the 31st international conference on neural information processing systems*, pp. 4768-4777. 2017.
[138] Y. Chen, G. Peng, Z. Zhu, and S. Li, “A novel deep learning method based on attention mechanism for bearing remaining useful life prediction,” *Applied Soft Computing Journal*, 2020, doi: 10.1016/j.asoc.2019.105919.
[139] C. Liu, L. Zhang, J. Niu, R. Yao, and C. Wu, “Intelligent prognostics of machining tools based on adaptive variational mode decomposition and deep learning method with attention mechanism,” *Neurocomputing*, 2020, doi: 10.1016/j.neucom.2020.06.116.
[140] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why should i trust you?’ Explaining the predictions of any classifier,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, doi: 10.1145/2939672.2939778.
[141] P. R. Magesh, R. D. Myloth, and R. J. Tom, “An Explainable Machine Learning Model for Early Detection of Parkinson’s Disease using LIME on DaTSCAN Imagery,” *Computers in Biology and Medicine*, 2020, doi: 10.1016/j.compbiomed.2020.104041.
[142] E. Štrumbelj and I. Kononenko, “Explaining prediction models and individual predictions with feature contributions,” *Knowledge and Information Systems*, 2014, doi: 10.1007/s10115-013-0679-x.
[143] T. Aven and E. Zio, “Some considerations on the treatment of uncertainties in risk assessment for practical decision making,” in *Reliability Engineering and System Safety*, 2011, doi: 10.1016/j.ress.2010.06.001.
[144] J. M. P. King, N. J. Titchener-Hooker, and Y. Zhou, “Ranking bioprocess variables using global sensitivity analysis: A case study in centrifugation,” *Bioprocess and Biosystems Engineering*, 2007, doi: 10.1007/s00449-006-0109-5.
[145] A. Saltelli and S. Tarantola, “On the Relative Importance of Input Factors in Mathematical Models,” *Journal of the American Statistical Association*, 2002, doi: 10.1198/016214502388618447.

[146] E. Borgonovo and E. Plischke, “Sensitivity analysis: A review of recent advances,” *European Journal of Operational Research*. 2016, doi: 10.1016/j.ejor.2015.06.032.
[147] A. Karpatne, W. Watkins, J. Read, and V. Kumar, “Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling.” arXiv preprint arXiv:1710.11431.
[148] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686-707.
[149] Tipireddy, R., Perdikaris, P., Stinis, P., & Tartakovsky, A. (2019). A comparative study of physics-informed neural network models for learning unknown dynamics and constitutive relations. arXiv preprint arXiv:1904.04058.
[150] L. Von Rueden *et al.*, “Informed Machine Learning-A Taxonomy and Survey of Integrating Knowledge into Learning Systems.” arXiv preprint arXiv:1903.12394.
[151] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, 2001, doi: 10.1214/aos/1013203451.
[152] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, “Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation,” *Journal of Computational and Graphical Statistics*, 2015, doi: 10.1080/10618600.2014.907095.
[153] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, “Practical black-box attacks against Deep learning systems using adversarial examples,” *ASIA CCS 2017 - Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security*, 2016.
[154] M. Barreno, B. Nelson, A. D. Joseph, and J. D. Tygar, “The security of machine learning,” *Machine Learning*, 2010, doi: 10.1007/s10994-010-5188-5.
[155] S. Sankararaman, S. Mahadevan, and M. E. Orchard, “Uncertainty in PHM,” *International Journal of Prognostics and Health Management*, vol. 6, no. 4, Nov. 2020, doi: 10.36001/ijphm.2015.v6i4.2289.
[156] R. Flage, T. Aven, E. Zio, and P. Baraldi, “Concerns, Challenges, and Directions of Development for the Issue of Representing Uncertainty in Risk Assessment,” *Risk Analysis*, vol. 34, no. 7, pp. 1196–1207, 2014, doi: 10.1111/risa.12247.
[157] P. Baraldi, R. Razavi-Far, E. Z.-R. E. & S. Safety, and undefined 2011, “Classifier-ensemble incremental-learning procedure for nuclear transient identification at different operational conditions,” *Reliability Engineering & System Safety* 96.4 (2011): 480-488.
[158] J. Liu, E. Z.-A. S. Computing, and undefined 2016, “A SVR-based ensemble approach for drifting data streams with recurring patterns,” *Applied Soft Computing* 47 (2016): 553-564.
[159] S. Al-Dahidi, F. Di Maio, P. Baraldi, and E. Zio, “A locally adaptive ensemble approach for data-driven prognostics of heterogeneous fleets,” *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 231, no. 4, pp. 350–363, 2017, doi: 10.1177/1748006X17693519.
[160] Zhang, D., Baraldi, P., Cadet, C., Yousfi-Steiner, N., Bérenguer, C., & Zio, E. (2019). An ensemble of models for integrating dependent sources of information for the prognosis of the remaining useful life of proton exchange membrane fuel cells. *Mechanical Systems and Signal Processing,* 124, 479-501.
[161] Deng, Y., Di Bucchianico, A., & Pechenizkiy, M. (2020). Controlling the accuracy and uncertainty trade-off in RUL prediction with a surrogate Wiener propagation model. *Reliability Engineering & System Safety*, 196, 106727.

[162] Peng, W., Ye, Z. S., & Chen, N. (2019). Bayesian deep-learning-based health prognostics toward prognostics uncertainty. *IEEE Transactions on Industrial Electronics*, 67(3), 2283-2293.
[163] Benker, M., Furtner, L., Semm, T., & Zaeh, M. F. (2020). Utilizing uncertainty information in remaining useful life estimation via Bayesian neural networks and Hamiltonian Monte Carlo. *Journal of Manufacturing Systems.*
[164] Biggio, L., Wieland, A., Chao, M. A., Kastanis, I., & Fink, O. (2021). Uncertainty-aware Remaining Useful Life predictor. *arXiv preprint arXiv:2104.03613.*