Asset Management decision-making through data-driven Predictive Maintenance – an overview, techniques, benefits and challenges

Madhu Krishna Menon[1], Rabin Tuladhar[2]
[1]College of Science and Engineering, James Cook University, Townsville, Australia
[2]School of Engineering and Technology, Central Queensland University, Rockhampton, Australia
[1]Corresponding author
E-mail: [1]*madhu.krishnamenon@jcu.edu.au*, [2]*rabin.tuladhar@jcu.edu.au*

Check for updates

Abstract. Over the years, industrial asset management has significantly transformed from being an unavoidable resource consumer to a value creator involving multi-criteria decision-making and optimisation. This is particularly important in the scenario of Industry 4.0, which offers more opportunities for improved maintenance effectiveness. This review examines the literature covering the evolving area of data-driven Predictive Maintenance (PdM) within engineering asset management. The work explores current and emerging practices for managing asset degradation, with emphasis on the domain of Prognostics and Health Management (PHM). Next, it examines the opportunities for data-driven methods, associated techniques, and data sources to incorporate data-driven PdM into the maintenance decision-making portfolio. The text concludes by discussing the opportunities and constraints related to data-driven PdM for three identified asset data streams. The paper offers insights for researchers and practitioners interested in utilising data-driven approaches to improve asset reliability, improve maintenance strategies and manage asset complexities.

1. Introduction

Industrial asset management is an interdisciplinary field that balances safety, reliability, maintainability, organisational performance and cost-effectiveness. It aims to optimise the cost, risk and performance across the life cycle of the assets. Asset management provides a systematic approach to new, in-service, and ageing asset decision-making. Such decisions impact the diverse assets throughout their life cycle, including design, manufacture, commissioning, operation, maintenance, upgrades and disposal. The major technical areas identified within the discipline are maintenance strategy, performance measurement, optimization, replacement, outsourcing, risk-based aspects, predictive maintenance, remaining asset life, planning, scheduling and information management [1]. The discipline has gained acceptance as a strategic tool for achieving business goals despite the breadth of subject matter resulting in differing interpretations to suit individual organisations [2]. In addition to physical entities, the scope of asset management encompasses intangible elements such as organisational practices, business objectives, stakeholder interactions, and enterprise management systems, thereby making it a complex socio-technical system of systems. The increasing complexity of systems results in a lack of understanding of asset properties, states, responses and internal interactions, making it more challenging for stakeholders to explain life-cycle decisions. In this context, the transformative technologies from the Industry 4.0 environment can be utilised to improve the understanding of the condition and management of complex assets. Data related to asset conditions, computational capabilities and connectivity facilitate the development of novel methods that learn and improve the knowledge of complex system characteristics over time. In addition to improving the sustenance practices

related to asset health, the newly discovered understanding generates valuable feedback for optimising the design of systems and components. Data-driven Predictive Maintenance (PdM) is a proactive strategy that utilises data streams from asset performance to gain insights and make decisions related to their condition, fault progression, risk assessment and maintenance practices. This review has effectively consolidated existing knowledge on data-driven PdM and lays the groundwork for creating practical workflows utilising the discussed approaches and algorithms.

2. Asset management in the context of Industry 4.0

The key technologies of Industry 4.0 are Cyber Physical Systems (CPS), Internet of Things (IoT), Industrial Internet of Things (IIoT), Internet of Services (IoS), context-aware smart factories, sensors, big data, connectivity and high-performance computing [3,4]. Implementing the elements of Industry 4.0 in manufacturing results in reduced development cycles, customized and flexible production, decentralized operations and improved resource efficiency [5]. These transformative opportunities are referred to as "disruptive" as they fundamentally alter how industrial systems are designed, maintained and operated, thereby introducing new decision-making practices in unprecedented ways. Connected assets, sensors and other devices in an IIoT environment improve the line of sight into operations, maintenance and optimized production processes. Analysis of big data from the assets can uncover hidden patterns, trends and insights, offering a novel alternative to traditional data analysis. An established method in Industry 4.0 is the utilisation of the Digital Twin, which involves creating digital representations of physical processes to enhance the understanding of complex system behaviour and provide early warnings of asset degradation. Digital Twins (DT) are the digital representations of physical processes in the Industry 4.0 ecosystem. When calibrated, they can improve understanding complex system behaviour and provide early warnings of asset degradation. The Digital Twin is widely recognized for its ability to integrate vertically across multiple hierarchical levels of the asset system. In addition, it can also be aligned horizontally with various stakeholders within and outside an organization. [6]. Due to the advancement of Industry 4.0 such as DT, the traditional discrete offerings of products and services are shifting towards an integrated product-service-system offering, also known as servitization in the environment of Industry 4.0.

Asset management systems of the future are expected to bring forth significant changes in business processes, methodologies and the breaking of boundaries across organizations. The "as-a-service" asset management model uses integration, communication and control of industrial assets to improve automated decision-making and reduce human involvement in routine tasks. These changes are facilitated by advancements in big data from assets, DT and engineering approaches such as Prognostics and Health Management (PHM) [7]. This environment is expected to introduce new business models where a consortium of owner-operators, service providers, and remote experts share the benefits and associated risks of the generated value. However, existing processes, technologies, and competencies need to be re-calibrated to ensure successful implementation of the new asset management vision. These gaps include integration and interoperability within the enterprise, effective processing of big data and the implementation of overarching models for service-oriented architecture like the Reference Architecture Model Industrie 4.0 (RAMI 4.0) [8]. Small and Medium Enterprises are found to trail behind bigger organizations in implementing Industry 4.0 practices due to outdated assets, semi-automated control systems, ageing information systems, and asset management practices that lack digital monitoring and reporting [9].

3. Prognostics and health management

Assets are entities that function as integrated systems with different levels of uniqueness to fulfil a functional requirement. Their performance deteriorates over time and may malfunction without proper care. Functional degradation, an inevitable outcome of asset utilisation, can be

reduced by implementing restorative actions. The maintenance function has a strategic role in the profitability, competitiveness, and asset reliability in the short and medium term. It also offers substantial opportunities for creating value in sustainable manufacturing, environmental and social benefits, governance, and the implementation of new technology [10, 11]. Maintenance activities are broadly categorised as Preventive and Corrective. Preventive Maintenance (PM) deals with planned and scheduled tasks carried out to minimise degradation and lower the likelihood of failure. PM tasks encompass physical inspections and tasks based on time, condition, usage, troubleshooting, function tests and statutory inspections. Corrective Maintenance (CM) is performed after recognizing a fault and is intended to restore the asset to an acceptable condition, thereby reducing degradation and function failure. There are two categories within CM. Reactive Maintenance (RM) is the immediate response to a malfunction or breakdown, usually involving restoration or replacement of affected components with little planning process involved. It is associated with unpredictability in cost, downtime, resource utilisation and the subsequent life of the asset. Deferred maintenance is the second type, performed opportunistically at a convenient timeframe. The system-level objective driving a maintenance strategy is to perform just the necessary PM while minimising RM [12-14]. The downside of PM is the possibility of over-maintaining and related costs, while a strategy accommodating RM is not advisable for safety-critical systems. This section discusses the significance of Prognostics and Health Management (PHM) during the operating phase of the asset life cycle in general, an overview of existing practices and the innovative approaches being researched in managing asset health.

Condition Based Maintenance (CBM) is a subset of PM that entails regular monitoring and analysis of parameters like machinery vibration, noise, motor current, thermography, lube oil condition, operational performance. It uses domain-specific diagnostic knowledge and rules to estimate the condition of the machinery [15]. Apart from being non-invasive, CBM can be used for fault diagnosis, optimised PM, potential reduction in CM, and provides a basis for predictive analysis. In addition, this approach allows for focused action without unnecessary maintenance [16]. CBM methods are effective only with asset categories that generate measurable parameters during operation. Despite investment in specialist technologies and trained manpower to interpret the output, traditional CBM faces challenges with accurate state estimation in real-time [3, 12]. The knowledge from CBM can be extended to an integrated, real-time health assessment framework under actual operating conditions and equipped with a layer of failure prediction and advisory in a framework named Prognostics and Health Management (PHM). PHM uses sensor data, failure physics, statistics and machine learning to perform real-time health assessments, predict future state and estimate the Remaining Useful Life (RUL) of an operating asset system [17]. Prominent approaches for RUL estimation and their respective benefits and limitations are summarised below.

3.1. Physics-based approach

The Physics-based approach utilises mathematical models to understand the degradation of a system, its components, and their interactions under different operating conditions. The guiding principle behind this approach is that failure occurs due to fundamental processes such as electrical, chemical, mechanical, thermal or radiation [18]. Applications of this approach include modelling of crack growth propagation, rotating machinery, turbine gas path and wear rate progression. These models require an in-depth understanding of Failure Modes Effect and Criticality Analysis (FMECA), system dynamics and an understanding of specific life-cycle-related properties [19]. The major steps involved in the physics-based approach are determining environmental factors, identifying failure-contributing mechanisms, environmental, operational conditions and functional dependencies of stresses for the best-fit model under a set of specified conditions [20]. This method requires significant computational resources and incurs substantial developmental costs. The complex model developed is tailored to a particular use case. While resource-intensive, once modelled, these artefacts can accurately represent degradation behaviour.

In addition, they can be used to simulate synthetic data responses to normal and faulty situations, thereby supporting decision-making for PHM.

3.2. Data-driven approach

The data-driven approach uses historical and real-time data from sensors or simulations to build a model to predict degradation. This approach relies on the knowledge gained from the understanding of the non-linear relationships between inputs and outputs without understanding of the degradation mechanisms. It can be applied to the system, equipment, or component levels [21, 22]. This method is valuable for detecting patterns in data by using specific techniques when dealing with extensive data and failure mode information. Data-driven PHM draws on techniques from statistics, probability, reliability, failure modelling and Artificial Intelligence (AI). This method is considered a "black box" since it depends on patterns and relationships in the data without being able to provide explanations based on engineering loads and material properties. Data-driven models do not require an in-depth understanding of degradation mechanisms and have a faster development cycle than physics-based models. Additionally, the algorithms used in this approach are portable across multiple systems [23].

3.3. Experience-based approach

Experience-based approaches like Expert Systems and Fuzzy Logic are considered when the development of physics-based models is restricted. A collaborative eXpertise-based PHM (X-PHM) approach from the domain of SMEs utilised knowledge from stakeholders integrated into the PHM process was found to improve understanding of the production process, provide explainability to decisions and support the digital transformation [24]. Expert Systems harness the knowledge and engineering experience of subject matter experts. This is codified as rules and if-then statements to address a particular PHM problem, with options available to update the knowledge base. This strategy integrates the basis and the conditions to apply the rules. However, it may lead to subjectivity, noise and an excessive number of rules to handle. The rules are system-specific, which limits their portability to another use case. Additionally, there is a significant dependence on the expert's ability to formulate the rules [25]. Fuzzy Logic-based methods, on the other hand, accommodate partial understanding of system behaviour and are less affected by noise. The integration of neural network and fuzzy logic in Adaptive Neuro-Fuzzy Inference System (ANFIS) is a method that reduces the size of the rule base, handles partial knowledge and offers explainability. They can process vague, imprecise, noisy and missing inputs and performing extensive simulations, thereby offering a compromise solution between analytical models and qualitative estimates [26].

3.4. Hybrid approach

The hybrid approach integrates physics-based, data-driven and experience-based approaches with the case-specific combinations across the three to enhance confidence in predictions. Data-driven approaches establish a baseline for safe operation based on the distribution of non-anomalous data, which can be verified and optimized by physics-based methods in a hybrid model [27]. Alternatively, a physics-based model with identified critical parameters can be validated for anomalies using field measurements from a data-driven model [28]. Hybrid approaches are beneficial when the failure modes are not amenable or too complex to be modelled with a physics-based approach. In the context of gears, rolling bearings, and hydrodynamic bearings, failure modes like misalignment, geometric errors and dirt were difficult to model using a physics-based approach [29]. Chao et al. [30] suggested a hybrid framework combining physics-based and Deep Learning (DL) methods to address the limitations of incomplete physics-based models and the limited representativeness of training sets in data-driven methods.

This framework is reported to improve prediction capabilities with lesser volumes of representative training data compared to purely data-driven models.

4. Data-driven predictive maintenance

PdM is a proactive strategy that utilises digital tools and an on-demand approach to reduce inefficiencies typically found in PM and CM. PdM can be considered an extension of CBM that leverages its monitoring capabilities, includes diagnostic and prognostic elements, and allows for proactive planning of maintenance tasks [31]. In addition to forecasting failures, PdM provides decision guidance on the most suitable time to execute the maintenance intervention, thereby balancing the task frequency and associated cost [32-34]. PdM augments the expertise of subject matter experts by identifying previously unnoticed anomalies and patterns [35]. Furthermore, the use of PdM techniques can decrease the environmental impact and support sustainable practices like zero-waste production [36, 37]. Unlike CBM wherein the trends of abnormal behaviour are monitored retrospectively, PdM tries to recognize non-typical system behaviour in their early stages of occurrence through data analysis, thereby allowing for quicker decision-making [8, 38]. Functioning with traditional PM, CBM, and CM, PdM makes predictions from repeated analysis of known characteristics from the data [39, 40]. PdM has been found to have use cases primarily in the domains of rotating machinery, aircraft systems, power systems, and Electronic Components [41]. However, it is important to note that apart from commercial implementations, many existing solutions do not fully integrate data collection, analysis, prediction and maintenance execution to provide comprehensive end-to-end decision-making support. Integrating PdM with the manufacturing process is a key requirement in developing PdM workflows [42]. Besides, other disruptive factors like the COVID-19 pandemic have also justified retrofitting traditional maintenance practices with collaborative human-machine-technology integration [43]. This section examines data-driven PdM that uses a single or a combination of data sources to generate actionable predictive workflows as part of an enhanced maintenance strategy. This data utilised for PdM can be different types of real data originating from the asset system or synthetic data generated from model simulations.

Life cycle data from assets can be found in different information systems, some of which are proprietary. In addition, not all stakeholders have seamless access to the complete data. To address this, the Machinery Integration, Management and Open Systems Architecture (MIMOSA) has developed specifications, protocols and an open-source data model for a supplier-neutral digital ecosystem that allows interoperability across multiple proprietary CBM systems [44]. The Open System Architecture for Enterprise Application Integration (OSA-EAI) of MIMOSA integrates engineering, maintenance, operation and reliability data for the target audience. Specifications and models within MIMOSA are built on the definitions of ISO 13374 for condition monitoring, ISO 17359 for diagnosis, and ISO 13381 for prognosis. The Open System Architecture for Condition-Based Maintenance (OSA-CBM) developed by MIMOSA consists of six functional block (Fig. 1). It enables the implementation of PHM strategies, integration of various data sources, and enhances interoperability among CBM systems [45]. By integrating sustainability principles, this flow chart can transform into a cyclical process that allows for the monitoring and tracing materials, ensures quality control, and facilitates inventory management [46].

Acquiring and analyzing data from sensors or events during asset usage is crucial for effectively assessing the condition of an asset and detecting potential faults. However, this data can be complex and subject to noise and errors. To ensure the quality of the data used for fault detection, it is important to consider factors such as the selection, calibration and sampling rate of sensors, as well as pre-processing steps like cleaning and de-noising. FMECA, performed at this stage, prioritizes actions and documents the cause-effect relationships. The health assessment stage quantifies the degradation intensity, and the prognostic assessment stage predicts the estimated time or function failure, providing input for proactive decision-making. The final step is advisory generation, which includes recommendations for operational actions and design

improvements. Overall, the chosen workflow for PHM should be able to detect incipient faults, perform failure diagnostics and manage post-prognostic health [47]. A successful prognostic model requires both a mathematical understanding of the model and an appreciation of the business goals it aims to achieve. This can be achieved through an extended three-step process as proposed by Sikorska et al. [48], which considers existing and future failure modes, and post-prognostic actions. Atamuradov et al. [28] proposed a four-step process for implementing PHM in maintenance, which includes (1) critical component analysis, (2) selection of the right sensors for CBM, (3) prognostic feature evaluation, and (4) prognostics methodology and tool evaluation matrices. Critical component analysis involves identifying assets whose unavailability or failures would have significant consequences.
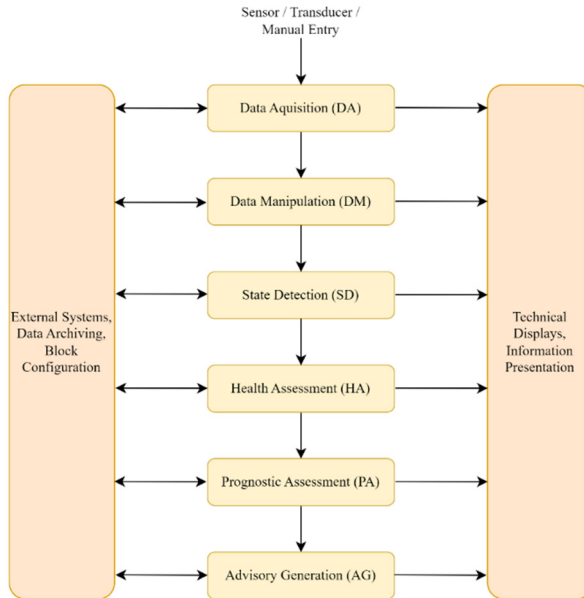

Fig. 1. OSA-CBM functional blocks

Although PHM applications that rely on a single data source can successfully address specific and well-correlated failure modes, their effectiveness is low in the case of concurrent failure modes. Moreover, they are susceptible to noise and environmental factors. Given this scenario, the integration of data from several sensors appears more attractive in offering a comprehensive view of asset condition. However, they require more advanced data fusion and analysis techniques and face challenges in implementation and interpretation due to the increased complexity of data. Lei et al. [49] summarises PHM use cases in both areas. PHM based on multiple data sources generally outperforms models based on single data sources. However, the choice between the approaches depends on the specific application, available resources and the complexity of the problem. An alternative strategy would be to apply single data-source-based PHM for less complex systems with known failure modes and multi-sensor-based PHM for more critical assets.

5. Data sources for PdM

Data-driven PdM workstreams employ techniques such as Statistics, Reliability Engineering, Signal Processing, Machine Learning and Deep Learning in workflows to improve the maintenance strategy and enhance decision-making [50]. The data used in PdM applications is obtained from multiple sources, asset levels, and sensors. The primary types of prognostic algorithms in PdM are regression models for tracing the trajectory of degradation, classification models for predicting failure within a time frame and survival models for estimating the risk of

failure over time [41].

This section examines three specific data streams that can benefit data-driven PdM applications. The first category consists of univariate data collected at high sampling rates, typically in vibration measurements obtained using an accelerometer. This information pertains to the health of a component, such as a bearing, located at the lowest level of the asset hierarchy. The second source comprises multi-variate data from the Supervisory Control and Data Acquisition (SCADA) system. This data consists of multiple performance variables that pertain to the overall condition of the system and is collected at longer intervals, usually a few times in an hour. The third source comprises unstructured, short technical text documented by stakeholders. These textual records can be associated with a point in time when an event occurred or the condition of an asset. On its own or in combination with other data streams, these sources have the potential for creating digital twin instances in the field of PdM.

A crucial aspect of data analysis is the extraction of features to analyse asset conditions. This is followed by selecting the most representative among the extracted features. This selection process involves feature analysis and feature evaluation, which deal with understanding feature behaviours and determining their importance, respectively [51]. These tasks, grouped under feature engineering, reduce the dimensions of the feature space and improve the performance of the predictive model. Data from multiple sensors can be processed concurrently to form multi-fusion methods at the data level, feature level and decision level to improve the efficiency of the prediction process [52].

A sizeable number of use cases with data-driven PdM tasks involving vibrations are applied in the detection, diagnostics and prognostics of faults with rotating machinery [53]. Among these cases, the critical components in rotating machinery where vibration measurements have been employed are bearings, gears and shafts. In addition, Lee et al. [54] have summarized the common failure causes, characteristics, CBM parameters, features and algorithms used for critical components of rotating machinery like bearings, gear, shaft, pump and alternator. On the other hand, Hegedüs et al., [55] have grouped 34 algorithms and modelling techniques associated with Root Cause Analysis (RCA), RUL Estimation and Alerts & Fault Prediction. A framework for monitoring bearing vibrations using data-driven techniques proposed dimensionality reduction techniques such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Analysis of Variance (ANOVA), Distance Evaluation Technique (DET), and Two Stage Feature Selection and Weighting Technique (TFSWT) for feature extraction [56].

Commonly used analysis methods and feature extractions on vibration data are in the time, frequency and time-frequency domains [54]. Metrics calculated from the signal, like the mean, Root Mean Square (RMS), standard deviation, peak value, peak-to-peak value, skewness, and kurtosis indexed to the time of occurrence are used to generate features in the time domain. Although they relate to a signal's physical behaviour and are easily interpretable, time-domain features are inadequate in capturing the non-linear and unsteady operation of machines. They fall short in detecting the underlying frequency components associated with specific faults. In contrast, frequency-domain features employ methods like the Fourier Transform and its variations, Residual Signal (RS), Difference Signal (DS), Special Kurtosis (SK) and Envelope Analysis (EA) to detect the prominent frequencies present in the signal. Nevertheless, this approach cannot ascertain the exact time of occurrence of those frequencies, limiting its effectiveness in analysing non-stationary and low-energy vibrations like those from wind turbines. On the other hand, time-frequency analysis methods like Short Time Fourier Transform (STFT), Wigner-Vile Distribution (WVD), Wavelet Transform (WT), Hilbert-Huang Transform (HHT), Empirical Mode Decomposition (EMD) that capture both time and frequency-related information concurrently are found to be suitable for non-stationary vibrations and detecting incipient faults [57, 58]. STFT provides both temporal and spectral information is more suitable for analysing non-stationary signals. Wavelet Transforms represent time signals as fast decaying finite length oscillating waveforms, scaled to match the input signal and reported to have better scalable resolution than time-frequency analysis. Hilbert-Huang Transform is an adaptive and

unsupervised technique where a non-stationary and non-linear signal is decomposed into a finite number of Intrinsic Mode Functions (IMF) [59, 60]. Furthermore, change point detection, independent component analysis and energy concentration distribution are some of the techniques considered in the processing of vibration signals [61].

Fig. 2 illustrates a classification model for evaluating asset condition before and after a maintenance activity based on vibration signal features. This model was developed using the top four signal features (Fig. 3) ranked based on monotonicity [51], a metric used to characterize the trend of features as the system degrades while in operation. The robust performance of the model in terms of classification and test accuracy can be ascribed to the selection of highly representative features.
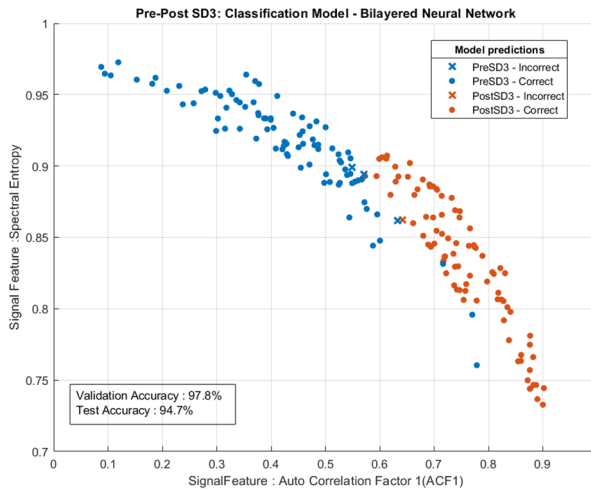

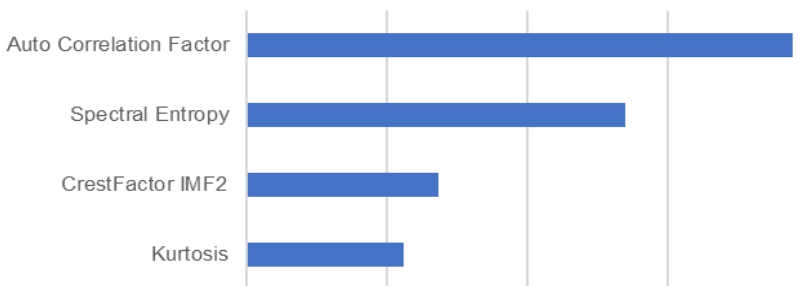**Fig. 2.** Asset condition pre and post maintenance

Fig. 3. Signal feature ranking based on monotonicity

Although the primary objective of SCADA is to control an industrial process, in the context of data-driven PdM, this repository becomes a valuable source of condition information. SCADA offers a multi-variate stream of data, consisting of operating parameters and environmental conditions, is captured typically at intervals of 10 minutes in the case of assets like Wind Turbines (WT). Being a pre-existing data source, PdM workflows based on SCADA data do not incur new capital investment on sensors or data acquisition systems [62]. In addition to component-level prognostics like bearing failure, the predictive models based on SCADA data can be customized for the system level as well [63, 64]. A semi-supervised anomaly detection approach known as "normal behaviour" models is trained on healthy data as a baseline to detect change points from a predominant group of PdM models with SCADA data [65, 66]. Comprehensive degradation modelling using SCADA needs to take care of the complex relationships across the operational parameters of the system and those of the ambient conditions, together with the age of the

asset [67].

An example of how dimensionality reduction and feature selection using Principal Component Analysis (PCA) can be applied to SCADA data streams is depicted in Fig 4. Seven sensor measurements related to the operation of a Centrifugal Chiller in HVAC system were reduced to three Principal Components which collectively accounted for approximately 94 % of the variance.
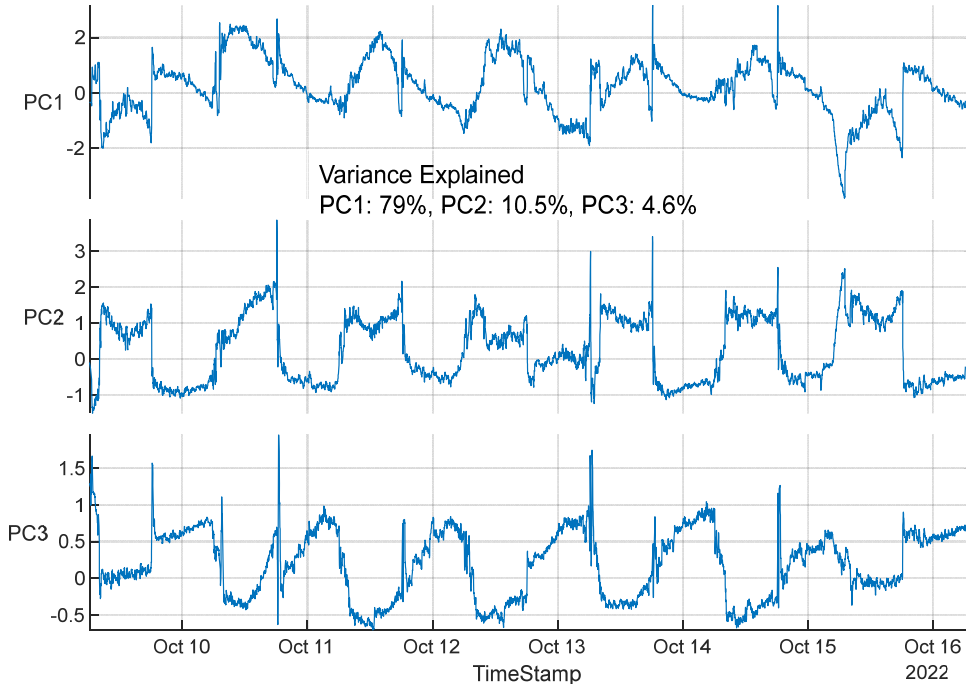

Fig. 4. SCADA features-principal components

Aside from sensor data, asset management repositories contain various sources of concise text, often unstructured, and contain valuable information regarding the condition of assets. These records need to be transformed into numerical representations, in contrast to sensor data originally logged in numerical format for PdM. Although Natural Language Processing (NLP) techniques exist and these repositories have comparable data volumes, such techniques rely on large volumes of structured text for training and hence have limited ability to generalise in the technical context, resulting in shallow actionable outcomes. Technical Language Processing (TLP) is envisaged as NLP customised for engineering data to decipher maintenance knowledge from short text, supported by iterative human action from an SME. TLP incorporates a human-in-the-loop and semi-automated process incorporating tagging of text entities and context-specific dictionary learning from SMEs. Introducing domain knowledge and supervision at key steps in the process involving unstructured text using TLP improves the explainability of the resulting computational data representations [68, 69]. Tagging or annotations on text corpus with TLP embeds domain-specific knowledge on fault characteristics, and their frequency of occurrence indicates the severity of faults.

In addition to tagging, the noise in unstructured text can be reduced by employing pre-processing techniques that extract text features and transform the text dataset into a more compatible format for Machine Learning and Deep Learning workflows [70]. Key pre-processing steps include spell-checking, stop-word removal, lemmatizing, tokenization, Bag of Words (BoW) and n-gram models. These steps assist in detecting the frequently co-occurring named entities, discovering latent topics from the text corpus, training models for text classification and analysing relationships using network representations. The inherent uncertainty associated with language-

based labels presents an opportunity to explore the development of weakly supervised learning methods. The integration of TLP-enabled computational linguistic approaches to PdM workflows has the potential to significantly reduce the effort required in manual annotation by a subject matter expert. Semi-automated workflows utilise clustering methods and association-rule mining from co-occurring terms to infer the underlying cause of failures. Furthermore, they contribute to the establishment of the hierarchy of events, failure ontology of affected entities and facilitate quantitative Failure Mode Mechanism and Event Analysis (FMMEA) [71, 72]. Technical Language Supervision (TLS), an extension of TLP, integrates recent advancements such as word embedding algorithms coupled with manual tagging on industry-specific short text [73]. The tagging exercise on short text related to Corrective Maintenance Work Orders for a Cooling Tower System yielded a notable decrease in vocabulary. It facilitated the categorization of tasks against the appropriate asset hierarchy, as shown in Fig. 5. In addition, the co-occurrence graph representations (Fig. 6) at the level of one of the Cooling Towers identified the interconnected entities associated with Corrective Maintenance work without manual review and annotation.
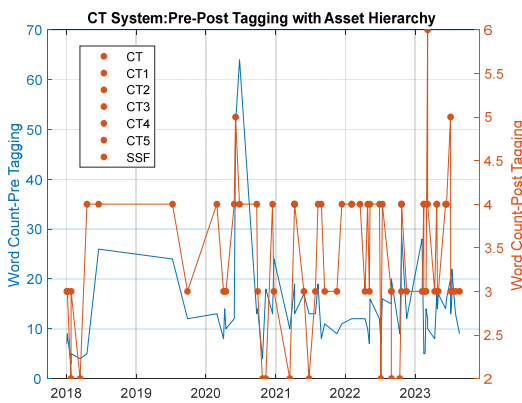

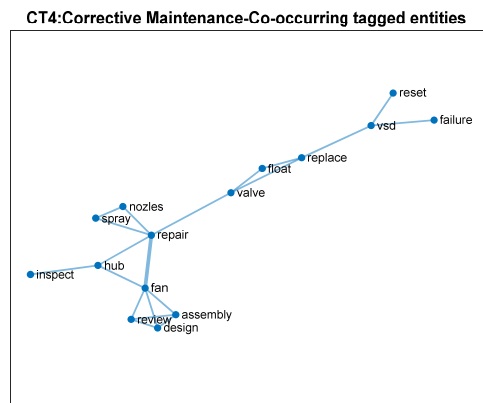**Fig. 5.** Tagging-hierarchy and vocabulary reduction

Fig. 6. Co-occurring entities-corrective maintenance

An asset's performance at any stage in its lifecycle is influenced by several failure modes that contribute to degradation in varying degrees. Furthermore, an available data stream might contain measurement noise, complicating the differentiation between normal and defective states. On the other hand, a best-fit degradation model might not remain the single source of truth across the complete life cycle. Changes in system configuration, affected assets, components or operating states call for redesign and retraining of AI-based models. Uncertainties about managing such modifications to the PHM deliverable can lead to questions on its robustness against model errors and unknown failure phenomena, especially for safety-critical applications. Present-day knowledge on managing and deploying robust AI-based PHM is still early. Consistency assurance, risk quantification, model updating and integrating with FMECA [74] are necessary to address time-dependent errors in the prediction solution caused by state changes.

6. Data processing techniques for PdM

This section summarises the prominent groups of techniques applied to data streams to create PdM workflows, specifically for fault detection, diagnostics, prognostics and RUL estimation. They are grouped under the categories of traditional Machine Learning and Deep Learning as shown in Table 1.

The application of Deep Learning (DL) algorithms in PdM is a relatively new and promising approach to solve problems in PHM covering vibration, images, time series and structured data due to their representing power, automated feature learning capability and performance

capabilities. This is driven by factors such as the availability of large amounts of data, the ability of DL to extract features automatically, and recent advances in theoretical research on algorithms and computing capabilities. Additionally, DL can learn relationships between and within different streams of CBM data and can transfer knowledge across different operating conditions and assets [88]. Wang et al. [89] suggest that in the field of applying DL in machinery, CBM, CNNs and AE are more popular methods, followed by RNN-LSTMs.

Table 1. Techniques – data driven PdM

No.	Technique	Description	Category
1	Support Vector Machines (SVM)	Supervised technique for anomaly detection, classification and regression Applied for RUL estimates in Lithium-ion batteries [75] and bearings [76]	ML
2	Decision Tree (DT)	Flow-chart like structures, consisting of root, multiple branches & nodes, used for fault detection, RUL [77] High explainability, relatively low accuracy and tendency to overfit [78]	ML
3	Random Forest (RF)	Ensemble of multiple, randomized Decision Trees Effective for predictive models with wind turbine performance [79]	ML
4	K-Nearest Neighbour (KNN)	Used for classification and regression. RUL estimates for bearings [80] and solar photovoltaic systems [81]	ML
5	Artificial Neural Network (ANN)	Considered as shallow neural networks, have limited number of inputs, hidden layers and outputs. Rely on hand-crafted features and are prone to over fitting [82]	ML
6	AutoEncoder (AE)	Unsupervised / semi-supervised feed forward neural networks consisting of two smaller networks named Encoder and Decoder. Used for Anomaly Detection as AE or as Variational AutoEncoder (VAE) During training, the Encoder learns features from input data while the Decoder tries to reconstruct the data from the features. Testing with anomalous data results in larger reconstruction errors Fault detection from accelerometer signals of hydro-generators [83]	DL
7	Convolutional Neural Network (CNN)	Effective with timeseries, signal processing and image data streams [84] Fault classification from accelerometer signals using 1D-CNN [85]	DL
8	Recurrent Neural Network (RNN)	Uses past information in the form of sequential data to generate outputs. Well-suited for Natural Language Processing (NLP), signal and video analysis. Limited capabilities when it comes to training long term dependencies, which can result in very small or very large network weights, thereby making them ineffective. RUL based on non-stationary vibration signal, capturing spatial information in the time-frequency domain using Deep RNN [86]	DL
9	Long Short-Term Memory Network (LSTM)	Overcome the limitation with RNN using additional forget gates to control the flow of information through hidden cells and output, more effective in learning long-term dependencies. Experimental verification of LSTM performance on CNC milling machine vibration [87]	DL

7. Ontology management for predictive maintenance

Despite the potential benefits of data-driven predictive methods in improving the quality of a maintenance program, organizations are yet to see significant improvements in their asset performance. There is a gap in integrating and embedding PHM work processes into maintenance work management systems. Internalising and reuse of frameworks, models and deliverables from data-driven PdM through processes that support interoperability and operational excellence is important for an organization undergoing digital transformation with asset management. This is achieved by developing context-specific ontologies with a common vocabulary for information sharing and machine-interpretable definitions related to entities and their relations. The transactional nature of maintenance work management and the involvement of multiple stakeholders makes this approach more relevant. Moreover, ontologies share a common understanding of information structure among systems, stakeholders and software. They facilitate the analysis and re-use of newly developed knowledge of the domain and its operation with explicit considerations and standardized approaches, thereby avoiding the need to reinvent the wheel [90]. Some advantages of ontology management are reduced workload in formalizing new concepts, lower development overhead and enhanced production sustainability.

ISO 14224 [91] recommends a standardized format for the collection of Reliability and Maintenance data for processing equipment, facilitating data exchange across multiple stakeholder organizations. The categories of data collected are: (1) equipment details with taxonomy and attributes, (2) failure causes and consequences, (3) maintenance actions, resources, consequences and downtime. ISO 15926 [92] is an interoperability standard for industrial data that facilitates the integration and exchange of plant data, along with archiving capabilities, throughout the life cycle. It has a shared generic data model with layered classes of objects and maps data at the source into a standardised format with quality assurance before data storage. The Cross Industry Standard Process for Data Mining (CRISP-DM) is one of the early models proposed in data-driven approaches in industry. An extended version of CRISP-DM aligning business understanding and technical aspects, developed as Data Mining Methodology for Engineering Applications (DMME) supports data-driven workflows in the engineering domain [93, 94]. A human-in-the-loop process of DMME that brings together relevant stakeholders, engineers with domain experience and data analysts is represented in Fig. 7.
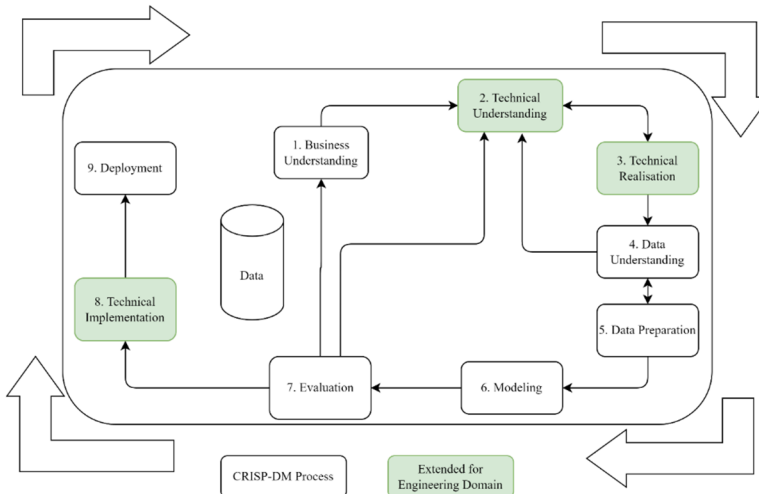

Fig. 7. DMME with extended steps in green for technical workflow management [94]

Nuñez et al. [95] developed a formal ontology application for PHM of machinery, employing the Design Science Research (DSR) methodology. This application featured a five-level

architecture, utilised Ontology Web Language (OWL2) for data modelling and was specifically designed to perform failure analysis on mechanical components. Furthermore, an application named OntoProg, which integrates condition monitoring data, measurement techniques, asset hierarchy, and failure mode information, was demonstrated using a centrifugal pump as a case study [96]. In addition to this, the Reference Ontology for MAINtenance Domain (ROMAIN) offers an open-access framework that is compliant with the Basic Formal Ontology (BFO). Moreover, it aligns with the top-level ontology, re-uses classes from middle levels and empowers researchers to add on new modules [97]. Meanwhile, Jiménez et al. [98] proposed a domain-neutral Ontology Model for Maintenance Strategy Selection and Assessment (OMMSA) serving as a formal terminology framework in maintenance strategies to develop smart computational agents for decision-making. Finally, Polenghi et al. [99], developed the Asset Management Ontology Development methOdology (AMODO) and demonstrated its effectiveness on a laboratory-scale use case involving flexible assembly line manufacturing.

8. Conclusions

Conventional asset management strategies have excessive reliance on PM, fail to cater to the specific needs of assets at different stages in their life cycle and are limited in incorporating the maintenance requirements after an upgrade. They have limitations in addressing the increasing complexity of assets, resulting in an inaccurate assessment of their condition, associated hazards and risks. This necessitates the adoption of a more proactive strategy towards holistic asset management. This provides an opportunity to improve the effectiveness of maintenance and optimise the costs incurred by the maintenance function. Advances in big data, IoT, machine learning, wireless connectivity, sensor technologies, computing power, ontologies, and associated architecture are driving the transformation of the future of maintenance. Data-driven PdM is expected to bring significant changes to the way maintenance is handled in asset management. The goal is to move from ad-hoc problem analysis to customized, economically viable practices based on continuous performance tracking and effective controls. Real-time use of sensor data, machine learning algorithms and PdM workflows generate opportunities for creating self-aware, self-adapting and self-healing systems. Internalising the new deliverables with formal ontologies improves knowledge retention and reuse. Such an environment can facilitate a step change from uncertainties in return on investment of PdM projects to improved confidence in performance-based contracts, where operators and service providers work together to meet key performance indicators. While there are success stories of integrating data from single assets and fleet-wide data for decision-making in critical systems, such as those of NASA and some Original Equipment Manufacturers (OEMs), there is limited evidence of sustained fleet-level PdM solutions in conventional industries. In addition to the use of known failure modes and measurable degradation parameters in state estimation and prognostics, the data-enabled practices need to focus on emerging and undetected failure modes as well.

There are substantial technical challenges while using data streams for data-driven models. Asset performance can be non-ergodic, where the statistical properties of the data streams cannot be averaged over long durations of time. Ageing, wear, environmental factors, quality of maintenance, usage patterns, influence of other assets, human factors including errors and incorrect operation as well as cyber-attacks are some factors contributing to non-ergodicity. While data-driven approaches can be fast and computationally efficient, the main concern is the lack of historical healthy and faulty data. Also, the degradation characteristics of similar assets may not converge to an expected value over time. This takes the scenario of degradation mapping outside the boundaries of a controlled experiment. Using larger datasets is one of the ways to mitigate such limitations affecting generalisation. Including too many input variables from data streams, each representing one dimension into an ML model, results in overfitting and loss of sense of relative distances and neighborhoods among data points, also known as the curse of dimensionality. This can be alleviated through the choice of relevant features by applying Feature

Selection Ranking (FSR) techniques. Developing data-driven techniques without considering domain knowledge poses multiple risks, such as inferior performance and inaccurate predictions due to a lack of understanding of underlying physical processes, sensor data, failure modes, operating conditions, interpretability, validation and safety, which can lead to unplanned downtime and increased costs. Therefore, it's important to include domain knowledge when developing data-driven solutions to ensure models are accurate, robust, reliable, safe and cost-effective. The imbalance caused by limited samples with abnormal asset behaviour and the absence of Run-to-Fail data can also offer challenges in training. Insufficient and excessive sources of training data from different assets belonging to a category lead to concerns with heterogeneity. To mitigate these issues, data fusion algorithms have been proposed to integrate dynamic operational data from assets.

Unlike traditional reliability prediction methods, predictive does not assume a constant failure rate. Additionally, many faults cannot be replicated during verification, which creates uncertainty about the state and risk exposure of the system. Most of the data-driven predictive work is found to be performed on smaller-scale experimental data instead of real-world applications. Hence, it has limitations in generalizing the outcomes. Models using data instances in the past miss out on time-varying characteristics of non-stationary data streams, which limits the life cycle of the trained model. Customised and economically viable practices based on continuous performance tracking and effective controls instead of ad-hoc problem analyses are required to demonstrate confidence in this area. To be effective, the design, structure, and outcomes of the model must be understandable to maintenance practitioners, making explainable prognostic recommendations essential. With lean organizational structures, specialized skills in asset management are crucial for managing risks associated with critical assets. In the future, maintenance practitioners will need not only subject matter expertise but also new skills to effectively manage these risks. This review summarises the limitations of conventional maintenance practices and provides an overview of current trends in maintenance decision-making. Considering an Industry 4.0 environment, it examines the practices associated with data-driven techniques with PdM. Potential areas of application for data-driven workflows are identified in the study. The review concludes by identifying the benefits, challenges and limitations of data-driven PdM. The integration of big data, IoT, machine learning, and sensor technologies in industrial asset management can transform maintenance practices from ad-hoc problem analysis to customized, economically viable, performance-based approaches. Data-driven PdM is a promising approach to navigating the uncertainties of industrial asset management by leveraging substantial amounts of multi-dimensional data generated by assets. However, the implementation of these advanced technologies in the maintenance environment requires addressing challenges such as data quality, scalability, and limited sample sizes. Data fusion algorithms are expected to mitigate these issues. The integration of these technologies also opens the potential for an integrated e-maintenance environment with predictive services outsourced to specialized service providers.

This review offers a synthesis on the developing field of data-driven PdM as a decision-making tool, positioning it in the context of engineering asset management in an Industry 4.0 setting. It establishes a basis for understanding the potential of integrating data-driven PdM with the ideas of digital twin, complexity and an as-a-service model of asset management. The distinctive contribution of this review is the examination of three separate data streams, viz. univariate vibration data, multivariate operational data and unstructured short text and the opportunities in developing data-driven PdM workflows utilizing these sources.

Acknowledgements

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Madhu Krishna Menon: conceptualization, writing-original draft preparation, Rabin Tuladhar: writing-review and editing.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] K. El-Akruti and R. Dwight, "Research methodologies for engineering asset management," in *ACSPRI Social Science Methodology Conference*, Jan. 2010.

[2] P. Konstantakos, P. Chountalas, and A. Magoutas, "The contemporary landscape of asset management systems," *Quality – Access to Success*, Vol. 20, No. 169, pp. 10–17, 2019.

[3] D. Galar, K. Goebel, P. Sandborn, and U. Kumar, *Prognostics and Remaining Useful Life (RUL) Estimation*. New York: CRC Press, 2021, https://doi.org/10.1201/9781003097242

[4] K. Schwab, "The Fourth Industrial Revolution," Crown, 2017.

[5] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business and Information Systems Engineering*, Vol. 6, No. 4, pp. 239–242, Jun. 2014, https://doi.org/10.1007/s12599-014-0334-4

[6] J. Dalzochio et al., "Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges," *Computers in Industry*, Vol. 123, p. 103298, Dec. 2020, https://doi.org/10.1016/j.compind.2020.103298

[7] A. Crespo Márquez, "Digital maintenance management," in *Springer Series in Reliability Engineering*, Cham: Springer International Publishing, 2022, https://doi.org/10.1007/978-3-030-97660-6

[8] A. Bousdekis, K. Lepenioti, D. Apostolou, and G. Mentzas, "A review of data-driven decision-making methods for Industry 4.0 maintenance applications," *Electronics*, Vol. 10, No. 7, p. 828, Mar. 2021, https://doi.org/10.3390/electronics10070828

[9] A. Shaalan, D. Baglee, and M. Knowles, "Are we ready for industry 4.0?," in *Smart Innovation, Systems and Technologies*, pp. 99–113, Aug. 2020, https://doi.org/10.1007/978-3-030-57745-2_10

[10] M. Holgado, M. Macchi, and S. Evans, "Exploring the impacts and contributions of maintenance function for sustainable manufacturing," *International Journal of Production Research*, Vol. 58, No. 23, pp. 7292–7310, Dec. 2020, https://doi.org/10.1080/00207543.2020.1808257

[11] B. R. Karki and J. Porras, "Digitalization for sustainable maintenance services: A systematic literature review," *Digital Business*, Vol. 1, No. 2, p. 100011, Oct. 2021, https://doi.org/10.1016/j.digbus.2021.100011

[12] M. Ben-Daya, U. Kumar, and D. N. P. Murthy, *Introduction to Maintenance Engineering*. Wiley, 2016, https://doi.org/10.1002/9781118926581

[13] "Maintenance. Maintenance Terminology," BS 13308, 2017.

[14] "The Maintenance Framework (II)," GFMAM, 2021.

[15] "Condition monitoring and diagnostics of machines-general guidelines," ISO 17359, 2018.

[16] L. Duan, F. Zhao, J. Wang, N. Wang, and J. Zhang, "An integrated cumulative transformation and feature fusion approach for bearing degradation prognostics," *Shock and Vibration*, Vol. 2018, No. 1, Feb. 2018, https://doi.org/10.1155/2018/9067184

[17] S. Cheng, M. H. Azarian, and M. G. Pecht, "Sensor systems for prognostics and health management," *Sensors*, Vol. 10, No. 6, pp. 5774–5797, Jun. 2010, https://doi.org/10.3390/s100605774

[18] "State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels," *Mechanical Systems and Signal Processing*, Vol. 94, pp. 214–236, 2017.

[19] J. Guo, Z. Li, and M. Li, "A review on prognostics methods for engineering systems," *IEEE Transactions on Reliability*, Vol. 69, No. 3, pp. 1110–1129, Sep. 2020, https://doi.org/10.1109/tr.2019.2957965

[20] Z. Matic and V. Sruk, "The Physics-of-Failure approach in reliability engineering," in *30th International Conference on Information Technology Interfaces (ITI)*, pp. 745–750, Jun. 2008, https://doi.org/10.1109/iti.2008.4588504

[21] R. Karim, D. Galar, and U. Kumar, *AI Factory: Theories, Applications and Case Studies*. CRC Press, 2023.

[22] A. Mosallam, K. Medjaher, and N. Zerhouni, "Nonparametric time series modelling for industrial prognostics and health management," *The International Journal of Advanced Manufacturing Technology*, Vol. 69, No. 5-8, pp. 1685–1699, Jul. 2013, https://doi.org/10.1007/s00170-013-5065-z

[23] H. M. Elattar, H. K. Elminir, and A. M. Riad, "Prognostics: a literature review," *Complex and Intelligent Systems*, Vol. 2, No. 2, pp. 125–154, Jun. 2016, https://doi.org/10.1007/s40747-016-0019-3

[24] N. Omri, Z. A. Masry, N. Mairot, S. Giampiccolo, and N. Zerhouni, "X-PHM: Prognostics and health management knowledge-based framework for SME," *Procedia CIRP*, Vol. 104, pp. 1595–1600, Jan. 2021, https://doi.org/10.1016/j.procir.2021.11.269

[25] L. Liao and F. Kottig, "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction," *IEEE Transactions on Reliability*, Vol. 63, No. 1, pp. 191–207, Mar. 2014, https://doi.org/10.1109/tr.2014.2299152

[26] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: a review," *The International Journal of Advanced Manufacturing Technology*, Vol. 50, No. 1-4, pp. 297–313, Jan. 2010, https://doi.org/10.1007/s00170-009-2482-0

[27] H. Zhang, R. Kang, and M. Pecht, "A hybrid prognostics and health management approach for condition-based maintenance," in *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 1165–1169, Dec. 2009, https://doi.org/10.1109/ieem.2009.5372976

[28] V. Atamuradov, K. Medjaher, P. Dersin, B. Lamoureux, and N. Zerhouni, "Prognostics and health management for maintenance practitioners – review, implementation and tools evaluation," *International Journal of Prognostics and Health Management*, Vol. 8, No. 3, Nov. 2020, https://doi.org/10.36001/ijphm.2017.v8i3.2667

[29] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery," *Advances in Mechanical Engineering*, Vol. 8, No. 8, Aug. 2016, https://doi.org/10.1177/1687814016664660

[30] M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *Reliability Engineering and System Safety*, Vol. 217, p. 107961, Jan. 2022, https://doi.org/10.1016/j.ress.2021.107961

[31] Mobley, R. K., *Maintenance Engineering Handbook*. McGraw-Hill Education, 2014.

[32] L. Pinciroli, P. Baraldi, and E. Zio, "Maintenance optimization in industry 4.0," *Reliability Engineering and System Safety*, Vol. 234, p. 109204, Jun. 2023, https://doi.org/10.1016/j.ress.2023.109204

[33] T. Zhu, Y. Ran, X. Zhou, and Y. Wen, "A survey of predictive maintenance: systems, purposes and approaches," *arXiv:1912.07383*, Jan. 2019, https://doi.org/10.48550/arxiv.1912.07383

[34] K. Shukla, S. Nefti-Meziani, and S. Davis, "A heuristic approach on predictive maintenance techniques: Limitations and scope," *Advances in Mechanical Engineering*, Vol. 14, No. 6, Jun. 2022, https://doi.org/10.1177/16878132221101009

[35] T. Cerquitelli, N. Nikolakis, N. O. 'Mahony, E. Macii, M. Ippolito, and S. Makris, "Predictive maintenance in smart factories," in *Information Fusion and Data Science*, Singapore: Springer Singapore, 2021, https://doi.org/10.1007/978-981-16-2940-2

[36] E. Jovicic, D. Primorac, M. Cupic, and A. Jovic, "Publicly available datasets for predictive maintenance in the energy sector: a review," *IEEE Access*, Vol. 11, pp. 73505–73520, Jan. 2023, https://doi.org/10.1109/access.2023.3295113

[37] D. Li, A. Landström, Fast-Berglund, and P. Almström, "Human-centred dissemination of data, information and knowledge in Industry 4.0," *Procedia CIRP*, Vol. 84, pp. 380–386, Jan. 2019, https://doi.org/10.1016/j.procir.2019.04.261

[38] E. Lughofer and M. Sayed-Mouchaweh, "Prologue: predictive maintenance in dynamic systems," in *Predictive Maintenance in Dynamic Systems*, Cham: Springer International Publishing, 2019, pp. 1–23, https://doi.org/10.1007/978-3-030-05645-2_1

[39] M. Molęda, B. Małysiak-Mrozek, W. Ding, V. Sunderam, and D. Mrozek, "From corrective to predictive maintenance-a review of maintenance approaches for the power industry," *Sensors*, Vol. 23, No. 13, p. 5970, Jun. 2023, https://doi.org/10.3390/s23135970

[40] J. J. Montero Jimenez, S. Schwartz, R. Vingerhoeds, B. Grabot, and M. Salaün, "Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics," *Journal of Manufacturing Systems*, Vol. 56, pp. 539–557, Jul. 2020, https://doi.org/10.1016/j.jmsy.2020.07.008

[41] Y. Wen, M. Fashiar Rahman, H. Xu, and T.-L. B. Tseng, "Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective," *Measurement*, Vol. 187, p. 110276, Jan. 2022, https://doi.org/10.1016/j.measurement.2021.110276

[42] T. Zonta, C. A. Da Costa, R. Da Rosa Righi, M. J. de Lima, E. S. Da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers and Industrial Engineering*, Vol. 150, p. 106889, Dec. 2020, https://doi.org/10.1016/j.cie.2020.106889

[43] García, A. Bregon, and M. A. Martínez-Prieto, "A non-intrusive Industry 4.0 retrofitting approach for collaborative maintenance in traditional manufacturing," *Computers and Industrial Engineering*, Vol. 164, p. 107896, Feb. 2022, https://doi.org/10.1016/j.cie.2021.107896

[44] R. Salokangas, E. Jantunen, M. Larrañaga, and P. Kaarmila, "Mimosa strong medicine for maintenance," in *Smart Innovation, Systems and Technologies*, pp. 35–47, Aug. 2020, https://doi.org/10.1007/978-3-030-57745-2_4

[45] J. Campos, P. Sharma, M. Albano, L. L. Ferreira, and M. Larrañaga, "An open source framework approach to support condition monitoring and maintenance," *Applied Sciences*, Vol. 10, No. 18, p. 6360, Sep. 2020, https://doi.org/10.3390/app10186360

[46] F. Polese, C. Gallucci, L. Carrubbo, and R. Santulli, "Predictive maintenance as a driver for corporate sustainability: evidence from a public-private co-financed R&D project," *Sustainability*, Vol. 13, No. 11, p. 5884, May 2021, https://doi.org/10.3390/su13115884

[47] N.-H. Kim, D. An, and J.-H. Choi, *Prognostics and Health Management of Engineering Systems*. Cham: Springer International Publishing, 2017, https://doi.org/10.1007/978-3-319-44742-1

[48] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing*, Vol. 25, No. 5, pp. 1803–1836, Jul. 2011, https://doi.org/10.1016/j.ymssp.2010.11.018

[49] 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*, Vol. 104, pp. 799–834, May 2018, https://doi.org/10.1016/j.ymssp.2017.11.016

[50] S. Werbińska-Wojciechowska and K. Winiarska, "Maintenance performance in the age of Industry 4.0: a bibliometric performance analysis and a systematic literature review," *Sensors*, Vol. 23, No. 3, p. 1409, Jan. 2023, https://doi.org/10.3390/s23031409

[51] J. Sim, S. Kim, H. J. Park, and J.-H. Choi, "A tutorial for feature engineering in the prognostics and health management of gears and bearings," *Applied Sciences*, Vol. 10, No. 16, p. 5639, Aug. 2020, https://doi.org/10.3390/app10165639

[52] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and A. Abraham, "Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach: A review of two decades of research," *Engineering Applications of Artificial Intelligence*, Vol. 123, p. 106139, Aug. 2023, https://doi.org/10.1016/j.engappai.2023.106139

[53] D. Jung, Z. Zhang, and M. Winslett, "Vibration analysis for IoT enabled predictive maintenance," in *IEEE 33rd International Conference on Data Engineering (ICDE)*, pp. 1271–1282, Apr. 2017, https://doi.org/10.1109/icde.2017.170

[54] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems-Reviews, methodology and applications," *Mechanical Systems and Signal Processing*, Vol. 42, No. 1-2, pp. 314–334, Jan. 2014, https://doi.org/10.1016/j.ymssp.2013.06.004

[55] C. Hegedűs et al., "The MANTIS reference architecture," in *The MANTIS Book*, New York: River Publishers, 2022, pp. 37–92, https://doi.org/10.1201/9781003339748-3

[56] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and Health management: a review on data driven approaches," *Mathematical Problems in Engineering*, Vol. 2015, pp. 1–17, Jan. 2015, https://doi.org/10.1155/2015/793161

[57] H. Ahmed and A. K. Nandi, *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*. John Wiley & Sons, 2020.

[58] V. Atamuradov, K. Medjaher, F. Camci, N. Zerhouni, P. Dersin, and B. Lamoureux, "Machine health indicator construction framework for failure diagnostics and prognostics," *Journal of Signal Processing Systems*, Vol. 92, No. 6, pp. 591–609, Jan. 2020, https://doi.org/10.1007/s11265-019-01491-4

[59] U. B. Souza, J. P. L. Escola, and L. D. C. Brito, "A survey on Hilbert-Huang transform: evolution, challenges and solutions," *Digital Signal Processing*, Vol. 120, p. 103292, Jan. 2022, https://doi.org/10.1016/j.dsp.2021.103292

[60] N. E. Huang, "Introduction to the Hilbert-Huang transform and its related mathematical problems," in *Interdisciplinary Mathematical Sciences*, Vol. 16, WORLD SCIENTIFIC, 2014, pp. 1–26, https://doi.org/10.1142/9789814508247_0001

[61] T. D. Popescu, D. Aiordachioaie, and A. Culea-Florescu, "Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: an overview," *The International Journal of Advanced Manufacturing Technology*, Vol. 118, No. 9-10, pp. 2883–2899, Oct. 2021, https://doi.org/10.1007/s00170-021-07703-1

[62] R. K. Pandit, D. Astolfi, and I. Durazo Cardenas, "A review of predictive techniques used to support decision making for maintenance operations of wind turbines," *Energies*, Vol. 16, No. 4, p. 1654, Feb. 2023, https://doi.org/10.3390/en16041654

[63] Encalada-Dávila, B. Puruncajas, C. Tutivén, and Y. Vidal, "Wind turbine main bearing fault prognosis based solely on SCADA data," *Sensors*, Vol. 21, No. 6, p. 2228, Mar. 2021, https://doi.org/10.3390/s21062228

[64] A. Yasuda, J. Ogata, M. Murakawa, H. Morikawa, and M. Iida, "System health monitoring of wind turbines using SCADA data and gaussian mixture models," in *PHM Society Asia-Pacific Conference*, Vol. 1, No. 1, p. 2017, Jul. 2017, https://doi.org/10.36001/phmap.2017.v1i1.1806

[65] X. Chesterman, T. Verstraeten, P.-J. Daems, A. Nowé, and J. Helsen, "Overview of normal behavior modeling approaches for SCADA-based wind turbine condition monitoring demonstrated on data from operational wind farms," *Wind Energy Science*, Vol. 8, No. 6, pp. 893–924, Jun. 2023, https://doi.org/10.5194/wes-8-893-2023

[66] S. Letzgus, "Change-point detection in wind turbine SCADA data for robust condition monitoring with normal behaviour models," *Wind Energy Science*, Vol. 5, No. 4, pp. 1375–1397, Oct. 2020, https://doi.org/10.5194/wes-5-1375-2020

[67] D. Astolfi, R. Pandit, L. Celesti, A. Lombardi, and L. Terzi, "SCADA data analysis for long-term wind turbine performance assessment: A case study," *Sustainable Energy Technologies and Assessments*, Vol. 52, p. 102357, Aug. 2022, https://doi.org/10.1016/j.seta.2022.102357

[68] M. P. Brundage, T. Sexton, M. Hodkiewicz, A. Dima, and S. Lukens, "Technical language processing: Unlocking maintenance knowledge," *Manufacturing Letters*, Vol. 27, pp. 42–46, Jan. 2021, https://doi.org/10.1016/j.mfglet.2020.11.001

[69] A. Dima, S. Lukens, M. Hodkiewicz, T. Sexton, and M. P. Brundage, "Adapting natural language processing for technical text," *Applied AI Letters*, Vol. 2, No. 3, Jun. 2021, https://doi.org/10.1002/ail2.33

[70] C. Wu et al., "Natural language processing for smart construction: Current status and future directions," *Automation in Construction*, Vol. 134, p. 104059, Feb. 2022, https://doi.org/10.1016/j.autcon.2021.104059

[71] H. B. Gunay, W. Shen, and C. Yang, "Text-mining building maintenance work orders for component fault frequency," *Building Research and Information*, Vol. 47, No. 5, pp. 518–533, Jul. 2019, https://doi.org/10.1080/09613218.2018.1459004

[72] A. Trilla, N. Mijatovic, and X. Vilasis-Cardona, "Towards learning causal representations of technical word embeddings for smart troubleshooting," *International Journal of Prognostics and Health Management*, Vol. 13, No. 2, Jul. 2022, https://doi.org/10.36001/ijphm.2022.v13i2.3127

[73] K. Lowenmark, C. Taal, J. Nivre, M. Liwicki, and F. Sandin, "Processing of condition monitoring annotations with bert and technical language substitution: a case study," in *PHM Society European Conference*, Vol. 7, No. 1, pp. 306–314, Jun. 2022, https://doi.org/10.36001/phme.2022.v7i1.3356

[74] S. Khan, S. Tsutsumi, T. Yairi, and S. Nakasuka, "Robustness of AI-based prognostic and systems health management," *Annual Reviews in Control*, Vol. 51, pp. 130–152, Jan. 2021, https://doi.org/10.1016/j.arcontrol.2021.04.001

[75] J. Wei, G. Dong, and Z. Chen, "Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression," *IEEE Transactions on*

Industrial Electronics, Vol. 65, No. 7, pp. 5634–5643, Jul. 2018, https://doi.org/10.1109/tie.2017.2782224

[76] W. Sui, D. Zhang, X. Qiu, W. Zhang, and L. Yuan, "Prediction of bearing remaining useful life based on mutual information and support vector regression model," in *IOP Conference Series: Materials Science and Engineering*, Vol. 533, No. 1, p. 012032, May 2019, https://doi.org/10.1088/1757-899x/533/1/012032

[77] Z. Zheng et al., "A novel method for lithium-ion battery remaining useful life prediction using time window and gradient boosting decision trees," in *10th International Conference on Power Electronics and ECCE Asia (ICPE 2019 – ECCE Asia)*, pp. 3297–3302, May 2019, https://doi.org/10.23919/icpe2019-ecceasia42246.2019.8797021

[78] M. Jamil, S. K. Sharma, and R. Singh, "Fault detection and classification in electrical power transmission system using artificial neural network," *SpringerPlus*, Vol. 4, No. 1, Jul. 2015, https://doi.org/10.1186/s40064-015-1080-x

[79] A. Kusiak and A. Verma, "Prediction of status patterns of wind turbines: a data-mining approach," *Journal of Solar Energy Engineering*, Vol. 133, No. 1, Feb. 2011, https://doi.org/10.1115/1.4003188

[80] D. K. Appana, M. R. Islam, and J.-M. Kim, "Reliable fault diagnosis of bearings using distance and density similarity on an enhanced k-NN," in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2016, pp. 193–203, https://doi.org/10.1007/978-3-319-51691-2_17

[81] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis," *IEEE Transactions on Industrial Electronics*, Vol. 63, No. 3, pp. 1793–1803, Mar. 2016, https://doi.org/10.1109/tie.2015.2509913

[82] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. D. P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers and Industrial Engineering*, Vol. 137, p. 106024, Nov. 2019, https://doi.org/10.1016/j.cie.2019.106024

[83] R. Zemouri et al., "Recent research and applications in variational autoencoders for industrial prognosis and health management: a survey," in *Prognostics and Health Management Conference (PHM-2022 London)*, pp. 193–203, May 2022, https://doi.org/10.1109/phm2022-london52454.2022.00042

[84] S. Hijazi, R. Kumar, and C. Rowen, "Using convolutional neural networks for image recognition," Cadence Design Systems Inc., San Jose, CA, USA, 2015.

[85] J.-W. Kim et al., "A study on fault classification of machining center using acceleration data based on 1D CNN algorithm," *Journal of the Korean Society of Manufacturing Process Engineers*, Vol. 18, No. 9, pp. 29–35, Sep. 2019, https://doi.org/10.14775/ksmpe.2019.18.9.029

[86] M. Ma and Z. Mao, "Deep recurrent convolutional neural network for remaining useful life prediction," in *IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–4, Jun. 2019, https://doi.org/10.1109/icphm.2019.8819440

[87] R. Zhao, J. Wang, R. Yan, and K. Mao, "Machine health monitoring with LSTM networks," in *10th International Conference on Sensing Technology (ICST)*, pp. 1–6, Nov. 2016, https://doi.org/10.1109/icsenst.2016.7796266

[88] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Engineering Applications of Artificial Intelligence*, Vol. 92, p. 103678, Jun. 2020, https://doi.org/10.1016/j.engappai.2020.103678

[89] W. Wang, J. Taylor, and R. J. Rees, "Recent advancement of deep learning applications to machine condition monitoring part 1: a critical review," *Acoustics Australia*, Vol. 49, No. 2, pp. 207–219, Apr. 2021, https://doi.org/10.1007/s40857-021-00222-9

[90] N. F. Noy and D. L. Mcguinness, "Ontology development 101: A guide to creating your first ontology," Stanford knowledge systems laboratory technical report KSL-01-05, 2001.

[91] "Petroleum, petrochemical and natural gas industries-Collection and exchange of reliability and maintenance data for equipment," ISO 14224, 2022.

[92] "Industrial automation systems and integration-Integration of life-cycle ata for process plants including oil and gas production facilities," ISO 15926, 2019.

[93] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model," *Procedia CIRP*, Vol. 79, pp. 403–408, Jan. 2019, https://doi.org/10.1016/j.procir.2019.02.106

[94] H. Wiemer, L. Drowatzky, and S. Ihlenfeldt, "Data mining methodology for engineering applications (DMME)-a holistic extension to the CRISP-DM model," *Applied Sciences*, Vol. 9, No. 12, p. 2407, Jun. 2019, https://doi.org/10.3390/app9122407

[95] D. L. Nuñez and M. Borsato, "An ontology-based model for prognostics and health management of machines," *Journal of Industrial Information Integration*, Vol. 6, pp. 33–46, Jun. 2017, https://doi.org/10.1016/j.jii.2017.02.006

[96] D. L. Nuñez and M. Borsato, "OntoProg: an ontology-based model for implementing prognostics health management in mechanical machines," *Advanced Engineering Informatics*, Vol. 38, pp. 746–759, Oct. 2018, https://doi.org/10.1016/j.aei.2018.10.006

[97] M. H. Karray, F. Ameri, M. Hodkiewicz, and T. Louge, "ROMAIN: towards a BFO compliant reference ontology for industrial maintenance," *Applied Ontology*, Vol. 14, No. 2, pp. 155–177, Apr. 2019, https://doi.org/10.3233/ao-190208

[98] J. J. Montero Jiménez, R. Vingerhoeds, B. Grabot, and S. Schwartz, "An ontology model for maintenance strategy selection and assessment," *Journal of Intelligent Manufacturing*, Vol. 34, No. 3, pp. 1369–1387, Nov. 2021, https://doi.org/10.1007/s10845-021-01855-3

[99] A. Polenghi, I. Roda, M. Macchi, A. Pozzetti, and H. Panetto, "Knowledge reuse for ontology modelling in maintenance and industrial asset management," *Journal of Industrial Information Integration*, Vol. 27, p. 100298, May 2022, https://doi.org/10.1016/j.jii.2021.100298

Madhu Krishna Menon pursues his Ph.D. degree in data-driven solutions for Predictive Maintenance at the James Cook University, Townsville, Australia. He is a Chartered Professional Engineer with Engineers Australia (CPEng), a Chartered Engineer with the Institution of Mechanical Engineers (CEng-MIMechE) and a Registered Professional Engineer (RPEQ). Madhu holds a Master's degree in Maintenance and Reliability Engineering from the Federation University, Victoria and Bachelor of Mechanical Engineering from the University of Kerala, India.

Prof. **Rabin Tuladhar** has completed his Ph.D. from Saitama University. He is currently the Dean of Engineering and Technology at the Central Queensland University, Australia. Rabin has previously held various leadership roles at the James Cook University, Townsville, including those of the Head of Engineering (2022-2023), Director of Engineering Studies (2018-2020) and Associate Dean Learning and Teaching (2012-2018). He has previously worked as a Lecturer at the University of Canterbury, Christchurch, New Zealand and Post Doctoral Fellow with the Japan Society for Promotion of Sciences (JCPS).