**Politecnico di Torino**

Master's degree in Aerospace Engineering


# AI for Prognostics: a Survey on emerging approaches and implementation


**Supervisors:**
Prof. Matteo Dalla Vedova
Prof. Paolo Maggiore
Eng. Leonardo Baldo

**Author:**
Cosimo Cristian Errico


December 2025

# Abstract

The increasing demand for reliability, safety, and efficiency across complex engineering systems has placed Prognostics and Health Management (PHM) at the centre of industrial innovation. In particular, accurately estimating Remaining Useful Life (RUL) is crucial for predicting failures and extending system lifecycles. The progressive integration of Artificial Intelligence (AI) into PHM has reshaped the way reliability and maintenance are addressed in engineering systems emerging as a key enabler of advanced prognostic methods and complementing or surpassing traditional physics-based models.

This thesis presents a systematic literature review (SLR) on emerging approaches for AI-based prognostics, analysing fifty-four peer-reviewed studies across multiple engineering domains including aerospace, mechanical equipment, manufacturing, wind energy, and automotive. The review identifies clear trajectories ranging from physics-based and model-driven approaches, based on deterministic equations, to data-driven approaches based on machine learning and deep learning architectures, and finally toward hybrid solutions in which physical knowledge and AI models are integrated. Deep Learning techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, autoencoders, Echo State Networks (ESN), and more recently Transformers, are compared in terms of their methodological principles, applications, strengths, and limitations. Particular attention is devoted to macrotrends such as the widespread adoption of deep architectures and the growing emphasis on uncertainty quantification, explainability, and trustability.

By mapping the state of the art and reflecting on their limitations, this paper contributes to a deeper understanding of how AI can be effectively integrated into PHM and sketches the direction in which future research and industrial adoption are expected to move.

# Table of Contents

# List of Tables

# List of Figures

# Acronyms

**SLR**
: Systematic Literature Review

**PHM**
: Prognostics and Health Management

**RUL**
: Remaining Useful Life

**AI**
: Artificial Intelligence

**ML**
: Machine Learning

**DL**
: Deep Learning

**CNN**
: Convolutional Neural Network

**RNN**
: Recurrent Neural Network

**RL**
: Reinforcement Learning

**ANN**
: Artificial Neural Network

**LSTM**
: Long Short-Term Memory

**FNN**
: Feedforward Neural Networks

**DT**
: Digital Twin

**IoT**
: Internet of Things

**STFT**
: Short Time Fourier Transform

**MCSA**
: Motor Current Signature Analysis

**ACMS**
: Aircraft Condition Monitoring System

**MDP**
: Markov Decision Process

**GRU**
: Gated Recurrent Unit

**ESN**
: Echo State Network

**VAE**
: Variational AutoEncoder

**AE**
: AutoEncoder

**GAN**
: Generative Adversarial Network

**SHAP**
: SHapley Additive exPlanations

**SVM**
: Support Vector Machine

**SoH**
: State of Health

**SCADA**
: Supervisory Control and Data Acquisition

**GPU**
: Graphics Processing Unit

**BMS**
: Battery Management System

**RMSE**
: Root Mean Square Error

**MAE**
: Mean Absolute Error

**MAPE**
: Mean Absolute Percentage Error

**NMSE**
: Normalized Mean Square Error

**CRPS**
: Continuous Ranked Probability Score

**PICP**
: Prediction Interval Coverage Probability

**PH**
: Prognostic Horizon

**RA**

Relative Accuracy

**CRA**

Cumulative Relative Accuracy

**XAI**

Explainable Artificial Intelligence

# Chapter 1

# Introduction

In the contemporary industrial context, PHM has been established as a pivotal element in ensuring the reliability and efficiency of complex systems. The implementation of prognostic techniques facilitates the adoption of predictive maintenance methodologies, a strategy that has been demonstrated to engender a reduction in unplanned downtime, whilst concomitantly extending the operational life of components and enhancing overall system performance. In particular, predictive maintenance supported by accurate prognostic methods is now widely recognised as a decisive strategy for reducing maintenance costs and preventing catastrophic failures, especially in highly critical or high-value sectors [1, 2].

Recent advances in AI, and particularly ML and DL, have given rise to a set of data-driven methodologies that are rapidly gaining ground compared to traditional models based exclusively on physical knowledge. The use of data-driven techniques has demonstrated great effectiveness in predicting failures, monitoring system health, and estimating RUL. These solutions have found widespread application in the manufacturing, aerospace, and automotive sectors, where automation and continuous analysis of field data are now integral to production processes [3, 4].

Historically, prognostic methods were primarily based on physical-mathematical or statistical models. Despite the interpretability and theoretical coherence offered by these approaches, they were frequently constrained in terms of scalability and adaptability to dynamic and nonlinear scenarios. The advent of AI has precipitated a paradigm shift within this domain. Data-driven methodologies exhibit enhanced flexibility, adeptness in the management of intricate signals, the capacity to withstand noise, and the ability to accommodate variable operating conditions and phenomena that are intractable to model with deterministic equations [5, 6].

The increasing amount of data provided by sensors, combined with the development of increasingly sophisticated learning algorithms, has accelerated the progress of new prognostic solutions. However, the current literature presents a highly heterogeneous landscape: profoundly different models, applications in widely divergent domains, and a lack of common standards for performance evaluation. This diversity makes it difficult to understand how effective, transferable, and applicable the developed solutions are to real-world industrial systems.

In light of this scenario, the primary objective of this thesis is to conduct a SLR aimed at critically synthesizing the most recent AI-based prognostic approaches. The work aims to compare the performance of data-driven techniques with that of traditional methods, highlighting their strengths, limitations, and potential critical issues. Furthermore, the analysis aims to identify the industrial sectors in which AI has already been most successfully integrated and highlight the barriers that still hinder the widespread adoption of these solutions in industrial practice [7].

The thesis therefore offers a structured and critical overview of the state of the art, with the aim of clarifying the maturity of the field, identifying research gaps, and providing useful insights for future developments, both theoretically and in practice. The SLR methodology ensures scientific rigour through predefined inclusion criteria, replicable selection procedures and a systematic comparative analysis of relevant contributions [8].

In an industry that is progressively reliant on real-time data and intelligent automation, the integration of AI into prognostics has become a strategic imperative. The present paper thus seeks to provide an evidence-based synthesis, with a view to highlighting the areas in which AI is effective, the areas in which it encounters difficulties, and the future directions that may facilitate the true operationalisation of these technologies in complex real-world systems.

## 1.1 Systematic Literature Review

In recent decades, the volume of scientific publications has grown at an astonishing rate. In any field of research, and particularly in engineering and computer science, the challenge is no longer just to find information, but also to distinguish, organise and critically evaluate what has already been produced. This is where the SLR comes in: a methodology that has progressively acquired a central role in ensuring rigour, transparency and replicability in bibliographic research.

Unlike traditional narrative reviews, which often reflect the author's subjective perspective and selection, the SLR aims to minimise bias by following a structured, transparent process. As Carrera-Rivera et al.[9] have observed, conducting an SLR involves the establishment of a clear and traceable process that commences with the formulation of research questions, progresses to the establishment of inclusion and exclusion criteria, and culminates in a critical synthesis of the findings. In this sense, SLR does not merely represent a "literature summary"; rather, it constitutes a genuine methodological experiment, the replication of which and subsequent verification by other researchers is essential.

The value of this approach primarily lies in its ability to provide a coherent framework in fields where research is fragmented or excessively broad. Kitchenham and Charters [8], who popularised the use of SLRs in software engineering, emphasise how SLRs allow evidence to be mapped, knowledge gaps to be identified, and a solid foundation to be provided for new lines of inquiry. Similarly, Brereton et al. [10] demonstrate that a well-conducted systematic analysis not only summarises the state of the art, but also constitutes a scientific contribution in its own right, providing guidance for researchers, practitioners and decision-makers. In the context of this thesis, the adoption of SLR is motivated by specific factors. The aim is to collect articles on the use of artificial intelligence in PHM and to create a comprehensive overview of emerging trends and applications across various sectors, as well as the most promising methodologies. Due to the multidisciplinary nature of the topic, which encompasses machine learning, deep learning, mechanical engineering, aerospace, energy and manufacturing, it is impossible to rely on a simple narrative overview. Only a systematic approach can ensure that important contributions are not overlooked and that the analysis remains consistent.

In this particular instance, the methodological dimension does not represent an incidental detail, but rather constitutes an integral component of the research endeavour. An SLR requires the researcher to declare their choices in advance: which databases were consulted, which keywords were used, which time periods were covered, and which criteria were used to select or exclude documents. This level of transparency not only makes the research more robust, but also more useful to the scientific community, as other scholars will be able to replicate, update or critique the process on objective grounds [11].

A further advantage of SLR is the ability to categorise collected data, facilitating systematic comparisons of approaches and results. In the context of PHM, this capability enables differentiation between not only application sectors (e.g. aerospace, automotive, wind energy and manufacturing), but also algorithm types (e.g. Recurrent Neural Networks, autoencoders and Bayesian models) and specific objectives (e.g. diagnostics, prognostics and system health management). This comparative classification facilitates the identification of cross-cutting trends and areas of overlap between apparently disparate fields. It demonstrates the adaptability of approaches developed in one sector for utilisation in others [4, 12].

The choice of SLR responds to two needs: firstly, to ensure the scientific soundness of the thesis and, secondly, to provide the reader with a reliable, critical overview. In an era when artificial intelligence is discussed with great enthusiasm and sometimes without critical analysis, adopting a methodology that minimises the risk of arbitrary selection is crucial for grounding the discussion in concrete and verifiable facts. This is therefore not merely a formal requirement, but an act of scientific responsibility [8].

Finally, it is worth emphasising that an SLR should not be viewed as a static tool.

While it was originally created to "bring order to the chaos" of scientific production, today its value also lies in its ability to adapt to dynamic and interdisciplinary contexts. In this case, the systematic review not only provides a snapshot of the current state of affairs, but also serves as a basis for reflecting on future directions. Which AI approaches are emerging most strongly? In which sectors are there already mature applications? Where are the grey areas or areas of embryonic experimentation? [13]

In conclusion, this research adopted SLR not only because it represents the most established methodological standard, but also because it is the most suitable tool for a broad, rigorous investigation with an innovation focus. Its purpose is not to replace the researcher's creativity, but to guide it through a transparent process that distinguishes between well-founded and uncertain ideas and clarifies the path for future AI-based PHM developments.

## 1.2 AI

### 1.2.1 Definition

Artificial intelligence (AI) is a branch of computer science that involves the study of methods and techniques for designing systems that can perform tasks requiring human-like intelligence. Unlike traditional software programmed to execute fixed commands, AI systems learn, reason, and adapt.
In their seminal textbook [14], Stuart Russell and Peter Norvig define AI through the lens of rational agents, based on two key components:

- **Agent**, entity that perceives its environment (via sensors, data stream, etc.) and acts upon environment (through actuators, decisions, predictions)
- **Rationality**: The agent chooses actions that optimally achieve its goals given available perceptual evidence or computational constraints

The rational-agent approach, favoured by the author, has two advantages over competing approaches. Firstly, the present approach is more general than the "laws of thought" approach, since correct inference is merely one component of this broader framework. The paper posits that there are several possible mechanisms for achieving rationality. Secondly, it is more conducive to scientific development than approaches that are based on human behaviour or human thought.

*"Artificial Intelligence is the study of agents that receive percepts from the environment and perform actions. A rational agent acts to maximize its expected performance measure based on its percept sequence and built-in knowledge"*[14]

The conceptual foundations of this field can be traced back to a provocative question posed by Alan Turing in 1950 [15], which served as a catalyst for further research in this area. The question of whether machines are capable of thought is a compelling one. The following query initiated a period of research and philosophical debate that lasted for decades.

It is imperative to define the term "intelligence" in this particular context. Pioneering figures such as John McCarthy conceptualised AI as "the science and engineering of making intelligent machines", systems endowed with the capacity for perception, learning, problem-solving, and decision-making [16]. This development represents a significant advance beyond the mere automation of processes; true AI is characterised by its capacity to understand context, recognise patterns in complex data, and make informed decisions in uncertain circumstances.

In order to comprehend the process through which human beings assimilate knowledge, it is first necessary to consider how this process occurs. A child is capable of recognizing a cat without explicit programming; rather, this ability is developed through repeated exposure to the subject and subsequent positive or negative feedback [17]. In a similar manner, contemporary AI systems, such as deep neural networks, evolve their competencies through the accumulation of experience. These systems are capable of ingesting vast quantities of data, detecting subtle correlations that are imperceptible to human analysts, and refining their internal models through iterative processes. This data-driven approach signifies a paradigm shift from the rigid, rule-based systems that characterised the early era of computing.

The field has evolved through distinct philosophical approaches. In the field of artificial intelligence, a subset of researchers have been motivated by the seminal work of Alan Turing on the imitation game to concentrate on the development of machines that can perform tasks in a manner that is indistinguishable from that of humans [15]. In the field of cognitive science, there exists a school of thought that aligns with the physical symbol system hypothesis proposed by Allen Newell and Herbert Simon [18]. This hypothesis posits that human cognitive processes can be modelled using physical symbols. The aim of this approach is to replicate human thought processes in a manner analogous to the way they utilise physical symbols. In contrast, the rationalist school, championed by Stuart Russell, prioritises practical utility, with agents capable of perceiving their environment and taking optimal actions to achieve goals [14].

At an operational level, AI encompasses a variety of interconnected disciplines.

Machine learning constitutes its dynamic core: algorithms that improve automatically through experience, in a similar way to how a medical diagnostician hones their skills over years of practice. The concept of knowledge representation offers a structural framework for encoding the complexities of the real world into computable formats. Natural language processing acts as a bridge between human communication and digital understanding, while automated reasoning enables logical inference from incomplete information.

The practical ramifications of these definitions are of particular significance for prognostics. In the context of industrial maintenance, conventional methods relied on predetermined thresholds and scheduled inspections. Conversely, AI systems utilise continuous monitoring through vibration sensors, thermal cameras, and acoustic monitors, providing a more comprehensive and real-time approach to machine condition monitoring. These systems are capable of detecting subtle anomalies, such as a faint harmonic resonance or a gradual temperature drift, which indicate impending failures that would not be detected by human operators for several months. This behaviour results from the model's ability to extract latent features and to capture nonlinear relationships that are not easily identifiable through traditional analytical methods. More specifically, it is the outcome of pattern recognition refined through exposure to large numbers of failure scenarios, combined with probabilistic reasoning about the RUL.

Despite its advanced capabilities, contemporary AI remains "narrow" in its expertise, with a focus on specific tasks, while lacking the human capacity for general understanding. While smartphones' voice assistants can flawlessly set reminders, they are incapable of reasoning about abstract concepts. This limitation assumes particular significance in safety-critical prognostic applications, where there is a need to balance predictive power with interpretability.

It can be posited that, at the present moment, the most efficacious definition of AI would be as follows:

*"A suite of technologies that transform data into contextual understanding, thereby enabling decisions that extend beyond the confines of programmed instructions."*

For those engaged in prognostics, this signifies systems that do not merely respond to failures, but rather anticipate them, thereby transforming raw sensor streams into actionable foresight. In the following chapters, an exploration will be conducted of the manifestation of these theoretical foundations in concrete prognostic techniques.

### 1.2.2 AI techniques

AI is not a single school of thought but is an overall domain that comprises a huge set of techniques for simulating human intelligence. There are two broad paradigms to be recognized across history: symbolic AI, which relies upon logical reasoning and rule-based systems, and statistical AI, which focuses on data-driven learning methods. Although earlier AI work was dominated by symbolic approaches, most recent advances have in fact been achieved with statistical methods, particularly in machine learning.

Of the most relevant of such approaches, the following deserve specific mention:

- **Machine Learning (ML)**: Machine Learning can be defined as algorithms that allow a system to improve at accomplishing a task by learning from data as opposed to relying on explicit programming. Depending on the nature of the data that is available, ML can be classified into three distinct categories: supervised learning, unsupervised learning and reinforcement learning. These paradigms have the capacity to facilitate the resolution of a wide range of problems, encompassing classification and prediction, as well as clustering and pattern recognition[19].

- **Deep Learning (DL)**: Deep Learning is a subset of ML that employs multi-layered artificial neural networks with the capacity to learn hierarchical representations of data. The field of deep learning has demonstrated remarkable proficiency in image classification, speech processing, and natural language understanding tasks. Convolutional Neural Networks (CNNs) are widely applied to vision-related tasks, while Recurrent Neural Networks (RNNs) and their extensions (e.g., Transformers) have dominated the sequence modelling and language modelling domains [12].

- **Reinforcement Learning (RL):** RL is defined as the process by which an agent learns to interact with an environment, taking actions and receiving feedback in the form of reward or penalty. The objective is to ascertain an optimal policy that maximises the cumulative reward. The field of reinforcement learning has been applied in various domains, including robotics, autonomous agents, and decision-making problems. A notable achievement in this area is AlphaGo, which has demonstrated significant progress in the field of Go [20].

Beyond these prevailing models, new methods are emerging that aim to combine the advantages of symbolic reasoning and statistical learning. This hybrid field, often referred to as neuro-symbolic AI, is attracting increasing attention due to its potential to overcome the limitations of data-driven systems in isolation, particularly with regard to interpretability and explainability.

The ensuing discourse will focus on the aforementioned three methods with meticulous scrutiny.

### 1.2.3 Machine Learning

Machine learning is one of the cornerstones of modern artificial intelligence and is based on the idea that a system can learn from data, rather than being explicitly programmed through rigid rules. In summary, the model is not only endowed with a set of instructions regarding its behavioural parameters, but is also capable of autonomous learning through the recognition of patterns, regularities and correlations in the observed signals. This approach is particularly effective in contexts where physical phenomena are difficult to model deterministically, or when the volume and variability of data exceed the descriptive capabilities of traditional techniques.

In the prognostic domain, machine learning allows us to identify anomalous signals, estimate the health of a component, and predict its future evolution by leveraging historical information provided by sensors. The ability to update models as operating conditions change makes these tools more adaptable than methods based solely on physics, especially in dynamic and real-world environments. Despite its wide variety of algorithms, the common denominator of machine learning is the centrality of data: it is data that drives learning, models relationships, and provides the knowledge needed to make increasingly reliable predictive decisions.

Among the diverse array of AI techniques, ML stands as a preeminent and pervasive approach. ML can be defined as the study of algorithms that enable computers to learn from data and improve their performance on a given task without being explicitly programmed [17]. Rather than being dependent on handcrafted rules, ML systems are capable of deriving patterns and relationships directly from examples. This renders them particularly effective in domains where explicit modelling is either impractical or excessively complex.

The fundamental components of ML can be categorised as follows:

- **Data:** the source of knowledge from which the algorithm extracts patterns.
- **Model:** the mathematical or statistical representation employed to map inputs to outputs.
- **Learning algorithm:** the optimisation process that adjusts the model's parameters in order to minimise prediction errors.

Depending on the availability and type of data, ML can be categorised into three major paradigms:

- **Supervised Learning:** The algorithm is designed to learn from labelled datasets, where each input is associated with a specific target output. The objective is to approximate the mapping function so that the model can predict outcomes for unseen inputs. This approach is commonly applied in classification problems (e.g., medical diagnosis, spam detection) and regression problems (e.g., stock price prediction, energy consumption forecasting).
- **Unsupervised Learning:** In this case, the training data is unlabelled, and the system's objective is to identify latent structures or patterns. Clustering algorithms (e.g., k-means, hierarchical clustering) are employed to group similar data points together, while dimensionality reduction techniques (e.g., PCA, t-SNE) are utilised to extract meaningful representations from high-dimensional data. Applications include customer segmentation, anomaly detection, and data compression.
- **Reinforcement Learning:** Although often considered its own domain, RL can be regarded as a form of learning where an agent interacts with an environment, receiving feedback in the form of rewards or penalties. The objective is to maximize long-term cumulative reward. RL has achieved remarkable success in robotics, control systems, and game-playing AI.

It is evident that, over time, ML has evolved to encompass a diverse array of models and methodologies. These encompass a range of approaches, including linear regression, decision trees, and ensemble methods such as random forests and gradient boosting. In recent times, the field has been dominated by deep learning, a subarea of machine learning based on neural networks with multiple layers, which has revolutionised fields such as computer vision and natural language processing.

The strength of ML lies in its ability to generalise from past experiences to future scenarios, making it a cornerstone of predictive analytics and a fundamental enabler for prognostics. Nevertheless, Machine Learning systems are also subject to limitations, including the requirement of substantial, high-quality datasets, vulnerability to bias, and challenges in interpretability. Addressing these issues remains an active area of research and development.

### 1.2.4 Deep Learning

Deep Learning (DL) is a subfield of Machine Learning that focuses on the use of artificial neural networks with multiple layers to automatically learn hierarchical representations of data. In contradistinction to conventional machine learning techniques, which frequently depend on handcrafted features, DL systems possess the capacity to extract complex and abstract features directly from raw data [21]. This ability renders DL particularly effective in domains such as computer vision, speech recognition, and natural language processing, where data is high-dimensional and difficult to model explicitly.

At the core of deep learning is the concept of Artificial Neural Networks (ANNs), which draw inspiration from the structure and functionality of the human brain. A neural network consists of interconnected layers of nodes, or “neurons,” with each connection possessing an associated weight that undergoes adjustment during the learning process. By employing iterative optimisation methods, such as gradient descent and backpropagation, the network is able to progressively learn to map inputs to outputs with increasing accuracy [22].

A range of deep learning architectures have been developed to address specific types of problems:

- **Feedforward Neural Networks (FNNs):** The most elementary form of neural network is characterised by unidirectional information flow, with information progressing from the input to the output. These tools are primarily employed for basic classification and regression tasks.

- **Convolutional Neural Networks (CNNs):** Designed for grid-structured data such as images, they are used in PHM to analyse one-dimensional vibration signals or two-dimensional time-frequency representations (e.g., spectrograms), excelling in the identification of specific local fault patterns. [21].

- **Recurrent Neural Networks (RNNs):** Being equipped with connections that capture sequential dependencies over time, variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are fundamental architectures for modelling progressive degradation and predicting time series, such as RUL estimation[22].

- **Transformers:** A novel family of architectures based on the self-attention mechanism. Transformers have brought about a paradigm shift in the field of natural language processing, serving as the foundational element of large language models such as BERT and GPT.

The success of deep learning can be attributed to three primary factors: the availability of large-scale datasets, the rapid growth of computational power (particularly GPUs and TPUs), and advancements in optimisation algorithms [21]. Nevertheless, deep learning models are also confronted with considerable challenges. These models generally require substantial amounts of labelled data, are computationally expensive, and often exhibit behaviour that can be considered a “black box," which gives rise to concerns regarding interpretability and trustworthiness [22].

Despite these limitations, deep learning has become a cornerstone of modern AI research and applications. Its influence is expanding rapidly across scientific, industrial and societal domains, and it is becoming one of the driving forces behind the current wave of artificial intelligence.

### 1.2.5 Reinforcement Learning

Reinforcement Learning (RL) is a distinct paradigm of Machine Learning that focuses on how agents learn to make decisions by interacting with an environment in order to maximise cumulative reward [20]. In contrast to supervised and unsupervised learning methods, which depend on labelled or unlabelled datasets, RL is founded on a trial-and-error process. The agent explores a range of actions, receives feedback in the form of rewards or penalties, and uses this feedback to improve its decision-making strategy over time.

The RL framework is commonly described through a *Markov Decision Process* (MDP), which consists of five key components:

- **States ($S$):** The representation of the environment at a given moment.
- **Actions ($A$):** The set of possible decisions or moves available to the agent.
- **Transition function ($P$):** The probabilistic rules that determine how the environment evolves after the execution of an action.
- **Reward function ($R$):** The feedback signal that evaluates the immediate utility of a given state-action pair.
- **Policy ($\pi$):** The strategy or mapping that the agent follows to choose actions based on states, which can be deterministic or stochastic.

The objective of RL is to identify an optimal policy that maximises the expected cumulative reward, frequently referred to as the return. This is typically accomplished by balancing exploration, defined as undertaking novel actions to ascertain their consequences, and exploitation, which is the selection of the best-known action to maximise reward.

There are several main categories of RL algorithms:

- **Model-free methods:** Algorithms such as Q-learning and SARSA that directly estimate value functions or policies without building a model of the environment.

- **Model-based methods:** Approaches that attempt to learn an explicit model of the environment and use it for planning.
- **Policy gradient methods:** Techniques such as REINFORCE and Actor-Critic methods that optimise the policy directly through gradient-based updates.

In recent years, the integration of RL with DL, known as Deep Reinforcement Learning, has achieved groundbreaking results. For instance, the DeepMind programme AlphaGo defeated world champions in the game of Go [23], a feat previously considered unattainable for machines. RL has also found applications in robotics, autonomous driving, resource management, and recommendation systems.

Despite these achievements, RL faces significant challenges. Training frequently necessitates a considerable number of interactions with the environment, which can be inefficient and costly in real-world applications. Furthermore, the instability of RL algorithms, their sensitivity to hyperparameters, and the difficulty of interpretation represent important limitations that preclude their use in safety-critical domains. Nevertheless, RL continues to be identified as one of the most promising areas of AI research, with the potential to drive significant progress in autonomous decision-making systems.

## 1.3 Prognostics and Health Management

The concept of industrial asset maintenance has undergone a radical evolution over the past decades. The paradigm shift can be described as follows: firstly, from a reactive approach, which might be summarised as "repair when broken"; secondly, to a preventive approach; and finally, to a proactive and predictive paradigm, known as PHM. The PHM does not merely serve to prevent failures; rather, its objective is to predict such failures by quantifying the remaining time within which a component or system will be able to function safely and efficiently. This remaining time is referred to as the RUL of the component or system [24].

The overarching objective is unambiguous: to maximise plant uptime, optimise maintenance costs, reduce waste, and, above all, ensure safety. The Internet of Things (IoT) and smart sensors, in conjunction with the capabilities of contemporary AI techniques, have effectively transformed the realm of PHM from a theoretical concept to a tangible and increasingly pervasive reality.

### 1.3.1 Diagnostics

The PHM framework is based on three fundamental pillars: diagnostics, prognostics, and health management. These components together provide a comprehensive approach to monitoring the condition of engineered systems, enabling predictive maintenance strategies.

Diagnostics answer the question, “What is wrong?” By analysing sensor data (e.g., vibration, temperature, sound, and current), algorithms can identify faults, determine their type, and locate their source. This process recognises and classifies the system’s current health status.

Diagnostics is responsible for interpreting the present. In essence, data analysis provides plant operators with the capability to address fundamental questions in real time. In order to ascertain the machine’s condition and identify any potential malfunctions, it is imperative to inquire about the presence of a fault, the nature of the fault, and its location [24].

Diagnostics is the first step of a PHM strategy: without it, any RUL estimate would rely on incorrect assumptions of an accurate and timely diagnosis, any attempt to predict the remaining useful life of the equipment would be based on incorrect assumptions, risking the generation of false alarms or, worse still, the failure to identify imminent failures. Diagnostics act as an early-warning layer in PHM, isolating anomalies before they evolve into failures.

The diagnostic process can be structured into three consecutive rational steps:

- **Detection**: identifying the occurrence of an event that deviates from standard operating conditions. This deviation can be conceptualised as a warning sign, suggesting the existence of an underlying root cause.

- **Identification or Isolation**: the determination of the nature and location of the fault. Merely detecting the presence of a problem is insufficient; the system must ascertain whether the issue is, for instance, a bearing defect, an imbalance, a misalignment, a lubrication fault, or an electrical malfunction.

- **Severity Assessment**: quantifying the severity of the fault. This phase determines whether a crack has merely initiated or whether a failure is imminent, and is therefore essential for prioritising maintenance actions.

The advent of Machine Learning and Deep Learning has precipitated an epistemological revolution in diagnostic techniques, thereby superseding the inherent limitations of conventional methods predicated on fixed thresholds and human interpretation. Contemporary data-driven models learn to recognise the characteristic signatures of each type of failure directly from raw data, thus obviating the need for preliminary analytical modelling of the system [6].

A review of the extant literature reveals a complex and specialised algorithmic landscape. Convolutional Neural Networks, although originally designed for image processing, have demonstrated extraordinary effectiveness in the analysis of vibration signals. The employment of time–frequency representations, such as spectrograms derived from the Short Time Fourier Transform (STFT), enables these architectures to discern intricate and nuanced patterns indicative of particular failures. This capacity, for instance, encompasses the identification of characteristic frequencies emanating from a damaged bearing, a feat which often surpasses the limits of human perception.

Concurrently, the advent of architectures predicated on attention mechanisms and Transformer-type models is ushering in a new era of diagnostic sophistication. The utilisation of these models, derived from natural language processing, empowers the system to selectively prioritise the most informative segments of a signal, even when they are temporally distant. Consequently, the capability to detect subtle anomalies amidst substantial background noise is significantly enhanced.

Transfer Learning and Unsupervised Learning techniques address additional practical issues. Transfer Learning enables a model to be trained on a reference machine, typically in a controlled environment, and subsequently adapted with limited data to similar machines operating in the field. Unsupervised learning, conversely, is indispensable in the absence of labelled data: by learning the "normal behaviour" of a system, deviations can be flagged as anomalies without requiring prior knowledge of the failure phenomenology [25].

The validity of any advanced diagnostic system is inextricably linked to the quality and comprehensiveness of the underlying data. Sensor placement, selection, and integration are therefore critical design steps. Established data acquisition methods include vibration measurements (the gold standard for rotating machinery), Motor Current Signature Analysis (MCSA), thermal and pressure monitoring for hydraulic and combustion systems, and acoustic or ultrasonic sensing for inaccessible components.

A significant current trend is multimodal data fusion. The synergistic integration of heterogeneous signals, such as vibration, current, and temperature, within a single AI model provides a holistic and contextual view of system health, thereby conferring greater robustness and accuracy than analysis performed on individual signals.

Advanced diagnostic methods are applied across a wide range of industrial use cases. In manufacturing, they enable wear detection in CNC tools, health monitoring of collaborative robots, and acoustic quality control of machining. In the energy sector, they support imbalance and misalignment detection in wind turbines, cavitation monitoring in pumps, and condition monitoring of transformers and generators. In the transportation sector, systems such as the Aircraft Condition Monitoring System (ACMS) identify faults in key aircraft and train components. In process plants, diagnostic models help detect anomalies in heat exchangers, compressors, and valves, where failure may have severe safety or environmental repercussions.

Last but not least, fault diagnosis, as revolutionised by artificial intelligence, is the conceptual and functional cornerstone of an effective PHM system. It transforms a raw stream of sensor data into actionable insight, enabling real-time visibility of asset health. Despite challenges linked to data availability and model transparency, the field continues to progress, pushing industry towards a predictive and intelligent maintenance paradigm in which failure is not a random event, but a foreseeable and manageable outcome [26].

### 1.3.2 Prognostics

While diagnostics focuses on identifying current conditions, prognostics addresses the fundamental issue of temporal projection of degradation and answers the question, “How long can it continue to function?” This is the most advanced and challenging aspect. In particular, using historical and current data, prognostic models estimate how a component will degrade until it fails completely, thus predicting its RUL. It is a prediction of the future based on an understanding of the present and past.

This interdisciplinary approach combines engineering principles, physical models and advanced artificial intelligence algorithms to track the progression of degradation over time and transform raw data into quantitative temporal forecasts. This approach is significant because it can convert technical information into operational decisions, facilitating maintenance planning that optimises asset utilisation while minimising the risk of sudden failure [2].

The contemporary prognostic approach is predicated on three fundamental methodological paradigms, each of which possesses distinctive characteristics and specific fields of application.

Data-driven approaches represent the most innovative frontier, relying on the automatic learning of degradation patterns from historical series of sensory data. It has been demonstrated that recurrent neural networks, and in particular LSTM (Long Short-Term Memory) architectures, have proven to be remarkably efficacious in the capture of temporal dependencies within degradation processes. Recently, transformer models have been developed that possess even more advanced temporal sequence processing capabilities, thanks to attention mechanisms that allow selective weighting of the most significant inputs.

Conversely, model-based approaches maintain their pertinence in applications where the physics of degradation is comprehensively understood and modelable. The aforementioned methodologies are contingent upon mathematical representations of deterioration processes, frequently implemented through Bayesian filters that recursively revise estimates in view of novel observations. The efficacy of these approaches is rooted in their interpretive transparency and reduced reliance on extensive historical data sets.

The emerging trend towards hybrid approaches seeks to synthesise the advantages of the two previous methods, combining the generalisability of data-driven models with the theoretical robustness of physical models. These hybrid architectures have the capacity to utilise neural networks in order to learn the parameters of simplified physical models, or vice versa, employ physical models to generate meaningful features for machine learning algorithms.

A fundamental distinction between advanced prognostics and simple forecasting lies in the capacity to quantify and communicate the uncertainty associated with estimates of RUL. The utilisation of point estimate forecasts, which provide a single numerical value, is being superseded by probabilistic distributions that express the confidence interval of the forecast.

The shift towards probabilistic models represents a significant advancement in the practical implementation of prognostic systems, providing decision-makers with a time estimate and a measure of its reliability. Recent developments in the field have revealed techniques such as ensemble methods and Gaussian processes to be powerful tools for modelling uncertainty. These techniques distinguish between epistemic uncertainty, which is derived from limited knowledge of the system, and aleatory uncertainty, which is inherent in the variability of the process itself [27].

The future trajectory of prognostics involves greater integration with digital twin technologies to create cyber-physical spaces in which predictive models can be updated and validated recursively with real-time data. This will enable RUL predictions to be refined, what-if scenarios to be explored, and maintenance strategies to be optimised within the virtual space prior to physical execution.

Digital twin technology is indeed one of the most promising tools for the industrial implementation of PHM. A Digital Twin is a virtual replica of a physical system that is continuously updated through real-time sensor data. By synchronising the behaviour of the asset and its numerical model, it becomes possible to simulate degradation, evaluate what-if scenarios, and refine RUL estimation during operation. In contrast to conventional prognostic approaches, which rely on static models, Digital Twins allow the prediction to evolve as the physical machine evolves. This provides higher accuracy, improved interpretability, and a more reliable basis for decision-making, especially in safety-critical environments.

The confluence of prognostics, industrial IoT and edge computing is also enabling the deployment of predictive models in the field itself, reducing decision latency and enabling autonomous response to signs of imminent degradation.

In conclusion, prognostics is coming of age as a discipline that is critical to the transition to truly predictive maintenance paradigms. The capacity to convert operating data to quantitative temporal information is a revolutionary advancement in industrial asset management with implications that reach beyond the practice of maintenance to the overall economics of life-cycle management of critical assets [4].

### 1.3.3 Health Management

Health management involves integrating diagnostics and prognostics into decision-making processes to ensure systems operate at their best. Its main purpose is to convert condition monitoring and prediction results into practical maintenance strategies.

The Health Management paradigm is the decision-making and operational part of the PHM paradigm. It provides the intrinsic connection between predictive analytics and real action on the ground. Diagnostics and prognostics generate data on asset health, and Health Management converts the data into optimised maintenance approach, operation plan, and economical decision.

This discipline integrates engineering, economic, and organisational capabilities, marrying technical analysis and operating management. The overriding theme of the approach is the capacity to transfer predictive insight through actionable plans to deliver greater plant availability, improved resource utilisation, and reduced lifecycle costs. This is founded on a basis of guaranteed safety and reliability at all points. Health Management's success is inherently tied to the quality and timeliness of its interaction with the remaining two pillars of PHM. Diagnostic models have been shown to provide early warning and correct indication of failures, and prognostic models provide the key temporal component through RUL estimation.

Such a combination makes possible a transition from planned maintenance, normally inefficient due to being based on conservative estimates, to condition-based maintenance that only acts as needed, and ultimately to real predictive maintenance, with interventions programmed with exact timing based on the actual state of degradation.

The crux of Health Management lies in the decision-making mechanisms that integrate multiple dimensions. Risk assessment integrates the probability of failure, derived from prognostic models, with the potential impact in terms of safety, costs, and production disruption. Risk matrix models facilitate the prioritisation of interventions based on actual criticality.

Concurrently, economic analysis entails a cost-benefit analysis of the trade-off between the costs of preventative maintenance and the potential costs of sudden failure, utilising life-cycle cost analysis techniques to inform replacement or repair decisions. This analysis is integrated with resource planning, ensuring the availability of spare parts, technical expertise, and optimised time slots.

The effectiveness of Health Management is assessed through a balanced set of indicators. Operational metrics, such as MTBF (Mean Time Between Failures) and MTTR (Mean Time To Repair), have been utilised to measure the enhancement of plant availability. Economic indicators, on the other hand, have been employed to quantify the Return On Investment (ROI) of PHM investments and the reduction in maintenance costs.

The final piece of the puzzle is the consideration of safety and environmental aspects, which are measured by the reduction in accidents and environmental impacts. This set of indicators helps quantify the value generated in terms of cost, safety, and sustainability.

In conclusion, Health Management signifies the culmination of the PHM journey, with predictive analytics being transformed into tangible operational value. The implementation of such a system necessitates a holistic approach, integrating technology, processes, and people to create ecosystems in which technical information, economic decisions, and operational execution converge towards the common goal of maximising asset value throughout their entire life cycle.

# Chapter 2

# Research Macroareas

The application of AI to PHM is not confined to a single industrial sector, but extends to domains differing in scale, safety requirements, and operational constraints. For the purpose of this systematic literature review, five macro-areas were selected for investigation: aerospace, manufacturing, mechanical equipment, wind turbines, and the automotive sector. This choice was not arbitrary, but rather informed by three principal criteria: historical relevance, industrial importance, and the richness of the available scientific landscape.

The aerospace sector is widely regarded as the starting point and one of the most advanced areas of PHM. Driven by extremely stringent safety requirements and the high costs of failure, it has historically represented a privileged testing ground for the development of new models, datasets, and evaluation metrics. It is noteworthy that numerous AI techniques which have become pervasive in other fields were initially evaluated through their application to aviation-related data sets.

The manufacturing sector is distinguished by its predominant adoption of PHM, primarily for economic reasons, with the objectives of reducing downtime, enhancing productivity, and ensuring consistent product quality. In this context, the concept of predictive maintenance is directly linked to enhanced efficiency and competitiveness. This renders the sector particularly conducive to the implementation of data-driven methodologies.

The category of mechanical equipment represents a cross-cutting field, encompassing components such as bearings, gears, pumps, and compressors. These elements are ubiquitous in almost all industrial sectors and are frequently subject to complex and nonlinear degradation processes. The ubiquity and criticality of these systems renders them a central topic for prognostics research, with results easily transferable to other contexts.

As representatives of a rapidly expanding sector of the global energy transition, wind turbines have become a prominent feature in the realm of renewable energy systems. The upkeep of these facilities, most notably those situated offshore, is both expensive and logistically complex. Consequently, the application of PHM is of particular pertinence in this context. Moreover, this sector exemplifies the potential of AI to enhance sustainability by ensuring the reliability of clean energy production infrastructure.

The automotive sector was selected for investigation due to its present phase of transformation, characterised by the emergence of electric and autonomous vehicles. In this context, prognostics assumes a pivotal role not only for traditional mechanical systems, but also for emerging components such as high-voltage batteries and advanced sensor platforms.

The five macro-areas under consideration, when taken together, offer a representative overview of the state of the art in AI-based PHM. The disciplines in question encompass both sectors in which the discipline has already reached a satisfactory level of maturity and areas where research is rapidly developing and continually evolving. A comparison of these cases illuminates the methodological trends that are shared, while also highlighting the distinct challenges posed by each context. This approach serves to address one of the research questions posed in this study. Consequently, the analysis assists in delineating the potential future trajectories of AI-based PHM, emphasising areas of convergence and identifying areas that necessitate further progress.

## 2.1 Aerospace

The aerospace sector is a notable and pioneering context for the development of PHM methodologies and this is attributable to the inherently safety-critical nature of the systems in question. Engine malfunctions, avionics failures or structural anomalies have the potential to result in catastrophic consequences with regard to safety and costs. Consequently, the capacity to anticipate and evaluate degradation in advance is regarded as a strategic imperative [28, 29].

The complexity of aviation and space systems is remarkable. The operation of these systems occurs within extreme environmental conditions, subject to high load cycles, and characterised by the necessity of component reliability over extended periods.

Moreover, the financial implications of unanticipated aircraft maintenance are substantial, primarily attributable to the direct impact on fleet operations and the intricate logistics involved. The adoption of PHM strategies within this sector has been demonstrated to result in a substantial reduction in costs, concurrently enhancing system reliability and availability [30, 5].

The introduction of AI has had a profound impact on the aerospace Health, Maintenance, and Management domains. Historically, predictive methods were predominantly reliant upon physical models or classical statistical analyses. These methods, in fact, were often characterised by rigidity and an inability to adequately capture the intricacies inherent in real-world data. The increased deployment of sensors and the advancement of data acquisition capabilities in modern aircraft has resulted in the generation of substantial amounts of information pertaining to engine operation, structural vibrations and avionics systems and, for this purpose, the advent of Artificial Intelligence has led to the development of sophisticated tools capable of effectively processing data, identifying latent patterns and, above all, providing more accurate estimations of the RUL of components.

Deep neural networks have been shown to be highly effective in analysing complex signals, which are prevalent in the aerospace industry. The employment of CNNs, RNNs, and LSTMs has facilitated the identification of nascent degradation signals in turbine engines, well in advance of their detection by conventional methodologies [6, 12]. Concurrently, methodologies that are more interpretable and lightweight, such as probabilistic regression models or Bayesian approaches, maintain a central role. These methodologies ensure transparency and traceability, which are essential elements for integration into certification processes [31, 32].

Accordingly, the aerospace sector can be regarded as an advanced laboratory for PHM. A significant proportion of the techniques currently employed in domains such as energy or industrial mechanics were originally conceptualised and validated in the aeronautical and space sectors. The study of this macro-area necessitates a meticulous examination of the most intricate challenges of prognostics, namely the necessity to strike a balance between accuracy and explainability, the management of uncertainty, the integration of data-driven models with physics-based approaches, the quality of available datasets, and the computational costs of implemented solutions.

The examples of applications are numerous and concrete: in aerospace engineering, the capacity to anticipate the deterioration of turbine blades or the progression of material fatigue in advance enables the implementation of targeted interventions and a subsequent reduction in downtime. For avionics systems, instead, the implementation of predictive analytics facilitates the identification of emerging electronic anomalies that could compromise navigation or communication functions. Furthermore, in composite materials, which are increasingly prevalent in modern aircraft structures, AI-based models facilitate more precise evaluation of internal damage phenomena that are challenging to discern with the unaided eye [2].

Finally, in the context of space missions, where maintenance is not feasible, AI-supported PHM systems become imperative instruments for autonomous vehicle health management [5, 28].

For all these reasons, aerospace was chosen as the first macro-area of analysis in this systematic literature review. This represents one of the most advanced contexts for the application of AI in prognostics, and it also provides a paradigm against which other sectors can be compared. Moreover, aerospace offers a prime illustration of the potential of AI in enhancing predictive performance, addressing the challenge of explainability, and ensuring the confidence necessary for implementation in real-world scenarios. In this sense, the study of PHM in the aerospace sector offers a valuable perspective on future research and industrial implementation directions, thereby providing a privileged lens through which to observe the evolution of the entire discipline.

## 2.2 Mechanical systems

Mechanical systems represent a significant and practically relevant sector for the application of PHM. In contrast to the highly specialised and extreme conditions prevalent in the aerospace sector, mechanical systems are ubiquitous from an engineering perspective. They include rotating machinery in manufacturing plants, pumps and compressors in power plants, gearboxes and bearings in transportation, and turbines in hydroelectric facilities, among numerous other systems. These systems form the backbone of modern industry. Their failures, whether minor or catastrophic, have direct economic consequences and indirect impacts on safety, productivity, and reliability. It is precisely this ubiquity and centrality that makes the study of PHM in mechanical systems a fundamental pillar for research and application [30, 1].

Conventional maintenance strategies within this sector have historically been predicated on two methodologies: scheduled maintenance, irrespective of the component's actual condition, and reactive maintenance, initiated only in the aftermath of a failure. It is evident that both of these solutions have clear limitations: the former results in excessive maintenance and wasted resources, while the latter exposes organisations to unplanned downtime and safety risks. The adoption of PHM methodologies signifies a paradigm shift. The integration of operating condition monitoring with predictive algorithms is a key aspect of PHM, facilitating the estimation of the RUL and the precise scheduling of interventions. The implications of this are significant, as evidenced by the findings of [4, 33], which demonstrate that optimised spare parts management, drastically reduced unplanned downtime, and more sustainable use of resources can be achieved.

AI has revolutionized this landscape. Mechanical systems now naturally generate large amounts of complex data, such as vibration signals from rotating components, acoustic emissions from gearboxes, thermal profiles from engines, and oil analysis in compressors. Each of these signals contains subtle traces of degradation, which are often invisible to the naked eye or to traditional statistical models. Machine Learning, and in particular Deep Learning, have demonstrated a remarkable ability to extract hidden patterns, offering far superior predictive capabilities compared to conventional thresholding systems. Convolutional Neural Networks (CNNs) have been successfully applied to vibrational spectral analysis, with the capability of automatically learning discriminant features without the requirement of manual preprocessing. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs) have been shown to excel at modelling the sequential nature of temporal data, a key element for identifying anomalies at an early stage. These approaches have enabled the anticipation of degradation phenomena, which were previously observable only retroactively, thus facilitating proactive and informed decision-making [6, 12].

However, the role of AI in this field is not limited to the realm of predictive accuracy. The ability to interpret models and the trustworthiness of the system are two of the most important aspects for the industrial adoption of PHM. Lightweight and more explainable models, such as probabilistic regression, Bayesian networks, support vector machines, or ensemble methods, continue to be central, especially in safety-critical contexts where every decision must be justified and validated against consolidated engineering knowledge. The future of PHM in mechanical systems therefore appears to be oriented toward hybrid paradigms, combining physics-based models that incorporate domain knowledge with data-driven algorithms exploiting the wealth of sensory data. This convergence ensures not only accurate predictions, but also reliability, transparency and acceptance by operators and regulators alike [31].

Applications of PHM in mechanical systems are numerous and concrete. In rotating machines, the detection of an incipient bearing defect has avoided the failure of entire production lines. In gearboxes, the prediction of damage progression through acoustic signals and torque monitoring has enabled the estimation of wear and crack propagation, ensuring the continued operation of heavy industry and transport. Pumps and compressors, which find wide application in the oil and gas industry, have AI-assisted monitoring systems that avoid downtime that can be extremely expensive and failures that can be catastrophic. Even in the renewable energy domain, such as the hydro sector, mechanical PHM allows for longer operating life and improved economy, proving that predictive maintenance is not only reactive but also geared towards sustainable objectives [4, 33].

The decision to contemplate mechanical systems as a macro-area in this review is predicated on the premise that, firstly, as previously stated, they embody the industrial milieu with the most pervasive dissemination of PHM applications, thereby rendering it a conducive environment for the evaluation of methodologies, the comparison of approaches, and the assessment of the scalability of solutions. Conversely, they emphasise the challenges associated with the transition from controlled laboratory environments to noisy, heterogeneous, and frequently incomplete data collected in the field. In contrast to the aerospace sector, where sensors are more standardised, the mechanical domain imposes constraints on AI, compelling it to function in conditions characterised by uncertainty and variable data quality. It is precisely due to these challenges that advances in this area act as a testing ground for the robustness and adaptability of prognostic algorithms.

In contemplating the future, the integration of AI with mechanical PHM is anticipated to evolve through three distinct trajectories. The first concerns the progression towards hybrid models, wherein Machine Learning integrates and enhances engineering physics knowledge. The second pertains to interpretability and trust: industrial operators will increasingly demand systems that not only predict, but also explain the rationale behind their decisions. The third trajectory is scalability, with predictive maintenance extending from isolated machines to entire interconnected production ecosystems, requiring models capable of generalising across different contexts. These perspectives indicate that the mechanical sector, often perceived as traditional and conservative, is destined to become one of the main stages of the industrial revolution driven by Artificial Intelligence.

## 2.3 Manufacturing

The manufacturing sector is widely regarded as one of the most complex and dynamic domains for the application of Prognostics and Health Management (PHM). In contrast to highly specialised aerospace or mechanical systems, which are ubiquitous, manufacturing can be regarded as a true living ecosystem, in which machines, processes, materials, and human operators coexist in tightly interconnected networks. It is not merely a matter of ensuring the functionality of individual components; rather, it is a matter of orchestrating entire production lines, where the sudden stoppage of a machine can generate ripple effects of delays, inefficiencies, and significant economic losses. In this sense, manufacturing systems constitute a privileged testing ground for PHM, as they embody the real challenge of integrating predictive intelligence into heterogeneous, fast-paced, and constantly optimised environments for productivity [34, 35].

Historically, maintenance in manufacturing has been based on rigid schedules or reactive interventions, resulting in high costs due to both excessive maintenance and unplanned downtime. Nevertheless, the advent of data-driven methods has radically transformed the landscape. In contemporary industrial settings, a pervasive integration of sensors has become the norm. This is evident in the utilisation of accelerometers on robotic arms, acoustic sensors on Computer Numerical Control (CNC) machines, vision systems employed for quality inspection, and the proliferation of Internet of Things (IoT) devices that continuously monitor thermal and vibration parameters. This abundance of data is both an opportunity and a challenge: it provides the raw material for PHM, but requires algorithms capable of extracting useful signals from noisy, multidimensional flows. This is precisely where Artificial Intelligence (AI) has emerged, not as an accessory tool, but as an essential requirement [36, 37].

It is evident that deep learning has led to the emergence of novel prospects in the domain of predictive diagnostics. CNNs have been shown to be capable of interpreting vibration maps or thermal images, automatically learning discriminative features without the necessity of manual preprocessing. RNNs and LSTMs have been shown to be particularly effective in the analysis of temporal data, with the capacity to detect subtle variations in machine behaviour even over extended time periods. One such example is the utilisation of LSTM networks in the context of machine tool monitoring, particularly with regard to spindle motor vibrations. Their application enables the detection of tool wear well in advance of the conventional thresholds that would otherwise be applicable. Reinforcement Learning also holds particular promise, as it has the potential to facilitate the development of adaptive strategies, involving both the prediction of degradation and the dynamic adjustment of operating parameters with the aim of reducing stress and extending service life [38, 19].

Notwithstanding the advances that have been made, the manufacturing sector continues to present a number of distinctive challenges. Production environments are characterised by noise, datasets are frequently incomplete, and machine types exhibit significant variation across different manufacturing facilities. This heterogeneity indicates that models trained in one context may not be applicable to another. Moreover, the financial implications of false positives are of particular significance: the premature closure of a production line due to an erroneous prediction can result in losses that exceed those of an actual failure. For this reason, explainability, reliability, and human-machine collaboration are not secondary aspects, but central requirements for adoption. A growing body of research is proposing the integration of physics-based models with data-driven algorithms, with the objective of ensuring engineering robustness and leveraging the full potential of sensory data [1, 31].

Numerous and concrete examples of PHM applications in manufacturing are extant. In the context of additive manufacturing, the utilisation of predictive models facilitates the identification of defects during the printing process, thereby minimising scrap and material wastage. In the semiconductor industry, PHM algorithms detect micro-anomalies in etching or deposition processes, safeguarding yield in one of the most quality-sensitive sectors. Even in traditional machining, PHM helps reduce tool wear, optimize cutting parameters, and improve product quality, demonstrating how predictive intelligence has a direct impact on both efficiency and sustainability [36, 37].

The choice to include manufacturing as a macro-area in this review stems from strategic motivations. On the one hand, it represents the industrial context in which PHM demonstrates the greatest potential for scalability, but it also underscores the necessity to strike a balance between predictive accuracy, explainability, and integration with existing legacy systems. In addition, the manufacturing sector offers a unique perspective on the future of PHM as an enabling technology for the fourth industrial revolution, which is characterised by intelligent and interconnected systems that autonomously coordinate production flows. The analysis of this sector entails not only the evaluation of the performance of algorithms but also the broader question of how intelligence can be integrated into the very heart of production systems [35, 19].

Looking ahead, the path of PHM in manufacturing will likely follow three main directions. The first is the convergence with digital twins, virtual representations that constantly reflect the real-world conditions of machines and processes, in which AI will act as a predictive brain capable of providing real-time insights. The second issue pertains to scalability across factory networks: techniques such as federated learning and PHM cloud platforms will facilitate knowledge transfer between disparate sites without compromising data confidentiality. The third, and arguably most transformative, is the alignment of PHM with sustainability objectives: namely, the reduction of energy consumption, the minimisation of waste, and the extension of machine lifecycles.

The rationale underpinning the selection of manufacturing as the macro-area of study in this thesis is precisely this strategic dimension. While the aerospace and mechanical domains offer insights limited to highly specialised components or contexts, the manufacturing domain highlights the systemic challenge of integrating AI-based PHM into daily industrial operations. A comprehensive analysis of this sector is therefore recommended in order to gain insight into the potential reconfiguration of future production models by artificial intelligence. This reconfiguration has the capacity to render manufacturing processes not only smarter, but also more sustainable and human-centric.

## 2.4 Automotive and Wind Turbines

In the fourth and final macro-area, the SLR has opted to integrate automotive and wind turbines, which are conceptually disparate. On the one hand, there is the mobility of people and goods; on the other, the production of renewable energy. However, an examination through the lens of Prognostics and Health Management (PHM) reveals notable parallels between the two sectors. Both are characterised by reliance on intricate electromechanical systems, operation under conditions of persistent stress and environmental variability, and stringent requirements for safety, reliability, and efficiency. The rationale for the combined analysis of these two fields is rooted in the shared necessity to address the challenges posed by the utilisation of Artificial Intelligence (AI) in the context of predicting, preventing and mitigating failures in critical systems that are of paramount importance to contemporary society [39, 40].

In the automotive sector, the transition to electrification and the spread of connected vehicles have amplified the importance of PHM. The utilisation of predictive strategies has already become imperative within the domain of internal combustion engines, with the objective being the anticipation of component degradation. Such components include, but are not limited to, gearboxes, bearings, and thermal systems. The advent of electric vehicles has led to a significant increase in the complexity of the challenges faced. Lithium-ion batteries exhibit nonlinear degradation phenomena that are difficult to model, thermal management is crucial for safety, and electronic control units must maintain reliable performance over very long timescales. The advent of artificial intelligence has enabled the development of sophisticated prognostic tools that facilitate the estimation of the residual lifespan of batteries, the monitoring of the health status of motors and inverters, and the optimisation of maintenance procedures across entire vehicle fleets [41, 7].

Wind turbines present a range of challenges that are equally as significant. These systems are characterised by their substantial size and the fact that they function in challenging environments, including variable wind speeds, humidity, temperature fluctuations, and, in the case of offshore installations, exposure to salt spray. Failures in blades, gearboxes, or generators can result in significant economic losses, with long periods of downtime and high maintenance costs, compounded by the difficulty of servicing offshore. The PHM has been extensively implemented in this sector through the utilisation of vibration analysis techniques, acoustic emissions, and SCADA data. The integration of AI into these methodologies has been shown to enhance their precision in fault diagnosis, facilitate the estimation of the Remaining Useful Life (RUL) of components, and facilitate condition-based maintenance, thereby reducing the necessity for costly and unanticipated interventions [24, 42].

In the automotive sector, CNNs are utilised for multi-sensor analysis in autonomous vehicles, with the capacity to detect anomalies in real time across multiple subsystems, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs) have been demonstrated to be effective in predicting the progression of battery degradation, thereby capturing the intricate and temporal nature of electrochemical processes. Furthermore, reinforcement learning-based approaches are paving the way for adaptive fleet-level maintenance strategies, with the potential to balance costs with reliability [7]. In the field of wind energy, the integration of physical aerodynamics with data analytics has demonstrated efficacy, particularly in the domain of gearbox diagnostics. The utilisation of data-driven models in this context is often hindered by limited or imbalanced datasets [43].

However, data heterogeneity remains a significant impediment in both sectors. Within the context of vehicles, sensors exhibit variability across models and manufacturers. Similarly, in the domain of wind farms, turbines of varying ages and sizes generate inconsistent information flows. The issue of explainability and trust is also crucial: in the automotive industry, models must be transparent enough to be integrated into certification processes; in wind power, operators require systems that justify predictions before initiating costly maintenance interventions. These challenges serve to reinforce the notion that hybrid approaches, which are based on a synergy between physical models and data-driven algorithms, represent the most robust path [31, 40].

The existence of concrete examples demonstrates the direct impact of these technologies. In the context of electric vehicles, the accurate prediction of battery degradation not only enhances safety but also ensures more reliable range estimates, thereby directly influencing consumer trust. In the context of car-sharing fleets, the implementation of predictive PHM for braking and steering systems has been shown to reduce downtime and enhance operational efficiency. In the field of wind energy, the timely identification of blade cracks or generator anomalies enables the implementation of timely interventions, thus averting catastrophic failures and extending the operational lifespan of turbines. In both cases, PHM also contributes to sustainability goals by reducing waste, extending life cycles, and optimising energy efficiency [41, 42].

Moreover, these are the areas in which the efficacy of AI must be demonstrated, not within the confines of a laboratory setting, but rather in the dynamic and unpredictable conditions of the road and the open sky. From this standpoint, the comparison emphasises the universality of the challenge, which is evident in the common features of complex systems, regardless of whether these systems are batteries or wind turbines. The fundamental question that remains unanswered is how to enhance the intelligence, resilience, and reliability of complex systems.

Even when considering the future, the trajectory of PHM in these sectors appears to be converging. Digital twins are set to assume a pivotal function: vehicles and turbines will be supported by virtual replicas updated in real time, capable of reflecting the health of the systems. The utilisation of federated learning will facilitate the dissemination of knowledge among automotive manufacturers and wind farm operators without the disclosure of proprietary data, thereby expediting the development of algorithms. Finally, sustainability will remain a driving force: in the automotive industry, PHM will support circular economy strategies such as battery reuse; in wind power, it will help extend the life of systems and optimise renewable energy production.

For these reasons, the decision was taken to designate automotive and wind turbines as the joint macro-areas in this review. These initiatives not only symbolise technological advancements but also social imperatives, including safe mobility, clean energy, and sustainable development.

## 2.5 Cross-Sectoral Insights

The comparative analysis of the selected macro-areas provides a comprehensive overview of the heterogeneous contexts in which Prognostics and Health Management (PHM), supported by Artificial Intelligence (AI), is being actively developed and implemented. Each sector is characterised by a unique set of challenges and opportunities, which collectively demonstrate the maturity, versatility, and limitations of current approaches.

As previously stated, in the domain of aerospace, the critical nature of operations and the presence of extreme operating conditions render PHM a necessity. The utilisation of AI facilitates the provision of more accurate and timely prognostics for systems whose failure would result in catastrophic consequences. Mechanical systems, by contrast, represent ubiquity: they are omnipresent in industry and everyday life, offering a testing ground for the scalability of PHM methodologies and for their transition from controlled laboratory settings to noisy and incomplete real-world data. Manufacturing introduces an additional layer of complexity, with entire ecosystems of machines and processes interconnected; in this context, PHM is not only a matter of component-level prediction but also of ensuring the resilience and efficiency of entire production flows, thus positioning AI at the heart of Industry 4.0. Finally, the utilisation of artificial intelligence (AI) in automotive and wind turbines underscores its pivotal function in facilitating condition-based maintenance and sustainability-oriented strategies, encompassing battery reuse and the extension of turbine lifecycles.

When considered collectively, these macro-areas illustrate that PHM is not limited to a specific engineering discipline, but rather signifies a unifying paradigm across various industries. This reveals the dual nature of AI in prognostics: firstly, its capacity to enhance predictive accuracy and predict degradation phenomena; secondly, its propensity to raise novel inquiries concerning explainability, trust, and integration with physics-based knowledge. The result is a dynamic landscape in which different sectors act as complementary laboratories: the aerospace sector for safety and reliability, mechanical engineering for ubiquity and robustness, manufacturing for system integration, and automotive/wind energy for sustainability and social impact.

It is possible to draw two conclusions from this cross-sector perspective. Firstly, the adoption of PHM is no longer an option, but rather a strategic imperative, as it has the capacity to simultaneously reduce costs, improve safety, and contribute to sustainability. Secondly, AI-based PHM will not evolve along a single trajectory, but through hybrid and sector-specific paths, converging towards a future in which predictive intelligence is integrated into the fabric of industrial systems. This convergence underscores the rationale behind the analysis of these macro-areas: by examining them together, this review highlights both the diversity of the challenges and the universality of the goal of constructing intelligent, resilient, and reliable systems that support the technological and societal transitions of the 21st century.

# Chapter 3

# Methodological Framework

## 3.1 Guidelines

Any research that aspires to be scientifically credible must be based on a solid, transparent methodological framework. Simply collecting articles and reporting their results is not enough: the value of a review lies in the clarity of the process by which the sources were selected, filtered and analysed, not in the quantity of sources. In this sense, the methodological framework is the backbone of the entire thesis. It is here that the rules of the game are established: how research questions are formulated, what criteria are used to decide what to include and what to exclude, what tools are chosen to ensure replicability and bias reduction.

Over the past twenty years, the systematic approach to literature reviews has gained increasing acceptance within the scientific community. For example, Kitchenham and Charters (2007) helped introduce systematic literature reviews (SLRs) to software engineering as a more rigorous alternative to narrative reviews, and Tranfield et al. (2003) demonstrated their potential for management sciences. More recently, Carrera-Rivera et al. (2022) reiterated that an SLR should be understood not as a compilation, but rather as a true "methodological experiment," designed to be replicable, verifiable, and improvable by other researchers. [8] [9] [11]

This thesis followed the guidelines established by Carrera-Rivera et al. (2022), responding to the need to bring coherence to an extremely diverse field: the application of Artificial Intelligence to Prognostics and Health Management (PHM). The complexity of the topic requires a process that goes beyond simply listing studies, but rather allows for tracing connections, identifying convergences, and accurately distinguishing between consolidated evidence and hypotheses that are still being explored.

## 3.2 PICOT

One of the most challenging aspects of conducting a systematic review is establishing the methodological framework that will guide the entire process. As Carrera-Rivera et al. [9] point out, developing a protocol is not merely a technical exercise; it is also a declaration of intent that establishes the research's boundaries, criteria, and priorities. From this perspective, the PICOT (Population, Intervention, Comparison, Outcome, Time) model is an essential starting point, as it enables the transparent and replicable articulation of the key variables that inform document selection.

The PICOT elements were defined as follows:

**Table 3.1:** PICOT Framework Breakdown

| Element | Definition | Example |
|---|---|---|
| P (Population) | Industrial systems | Aerospace components<br>Wind turbines<br>Mechanical equipment |
| I (Intervention) | AI prognostic models | Artificial Intelligence<br>Deep Learning<br>Machine Learning |
| C (Comparison) | Traditional prognostics methods | Physics-Based models |
| O (Outcome) | Performance evaluation | RMSE (Root Mean Square Error) for RUL |
| T (Time) | Publication years (2000-2025) | Recent advances |

When applied to the topic of AI-supported PHM, PICOT first and foremost clarifies the **Population**, i.e. the set of systems and industrial sectors to which it refers. In this case, the population under consideration includes a number of heterogeneous fields, including aerospace, mechanical systems, manufacturing, automotive, and wind energy. These fields share the need to ensure reliability, efficiency, and safety through the implementation of predictive maintenance.

The term **Intervention** is employed to denote the application of artificial intelligence techniques, with a particular emphasis on machine learning and deep learning, to the domains of diagnosis and prognostics. The focus of this review is on these methodologies, which are intended to replace or complement traditional approaches based on physical models or statistical methods.

The **Comparison** dimension is defined as the process of comparing these established approaches. The protocol stipulates that selected contributions evaluate the advantages, limitations, and commonalities between different paradigms: data-driven, model-based, and hybrid solutions.

The **Outcome**, or expected outcome, concerns the methodologies' ability to accurately estimate the RUL of components, improve early detection of anomalies, reduce maintenance costs, and increase system safety. These objectives are not uniform across all sectors, but this very heterogeneity allows us to observe how the same techniques can produce different impacts depending on the context.

Finally, **Time** defines the time frame of interest. In this work, the focus is primarily on studies published in the last twentyfive years, a period in which AI has experienced exponential growth and sensor and data acquisition technologies have enabled the development of large-scale PHM applications.

The PICOT protocol, therefore, is not a bureaucratic exercise, but rather an epistemic filter: it establishes the subjects to be observed, the manner in which the results are to be compared, and the objectives to be considered central to answering research questions.

## 3.3 Research Questions (RQs)

Starting from the framework outlined by PICOT, the Research Questions (RQs) emerge, formulated as the compass that will guide the review. Far from being mere exploratory curiosities, the RQs represent true methodological constraints that define the type of evidence to be collected and how to interpret it.

1. **RQ1**: What are the predominant approaches, namely data-driven, model-based, and hybrid methods, employed for prognosis?

2. **RQ2**: What are the accuracy gains and limitations of AI methods compared to traditional prognostics approaches?

3. **RQ3**: Which sectors have most successfully deployed AI-based prognostics, and what barriers persist in real-world implementation?

These questions allow for a multi-dimensional exploration: methodological (RQ1), performance-oriented (RQ2), and application-specific (RQ3). Their formulation ensures that the review not only catalogues existing literature but also synthesises insights across disciplines, thus contributing to the advancement of AI-supported PHM.

The integration of PICOT and RQ facilitates the establishment of a robust methodological framework, thereby reducing subjectivity and enhancing transparency. The protocol, in fact, transforms a broad and multidisciplinary topic, such as the use of AI in PHM, into a set of narrow, verifiable, and reproducible questions. As Carrera-Rivera et al. [9] argue, the distinguishing feature of a systematic review is its clarity of purpose, which differentiates it from a narrative review. The former does not merely involve the reading and summarising of existing literature; rather, it involves the construction of a methodological experiment that has the potential to be replicated, expanded, or contested by other researchers.

## 3.4 Search Strategy and Selection Criteria

Translating research questions into a query string is a vital part of any systematic review. Without a robust and well-structured query, results may be incomplete or, conversely, excessively scattered. In this research project, the prompt was developed through an iterative process to maximise relevance and minimise false positives.

The final query string, which was used in major scientific databases such as Scopus, Web of Science and IEEE Xplore, is as follows:

```
("Industr*" OR "Iiot" OR "mechanic*" OR "engin*")
AND ("prognos*" OR "Remaining Useful Life" OR "PHM")
AND ("artificial intelligence" OR "Deep Learning" OR "Machine Learning")
AND NOT ("Medical" OR "Disease" OR "Agriculture“)
```

**Figure 3.1:** Prompt

This formulation is divided into three main blocks, each corresponding to one of the central dimensions of the analysis. The first block narrows the scope of application to industrial and engineering sectors, including mechanics, engineering, and the Industrial Internet of Things (IIoT). The second block introduces the topic of PHM, with explicit reference to the concept of RUL. Finally, the third block identifies the technological element, namely artificial intelligence in its machine learning and deep learning forms.

The explicit exclusion of terms such as *"Medical"* or *"Disease"* was necessary to limit the corpus to industrial contexts and avoid the inclusion of biomedical literature, which, while sharing similar prognostic concepts, belongs to a radically different epistemic domain.

The overall structure of the prompt is therefore the result of a methodological compromise: on the one hand, the need to be sufficiently broad to include the variety of relevant application sectors; on the other, the rigor required to clearly define the review's boundaries.

The definition of selection criteria represents the second methodological pillar after the prompt's construction. As Kitchenham and Charters [8] emphasize, transparency at this stage is essential to ensure replicability and rigor, since every systematic review necessarily involves choices that can influence the final results.

In this study, the criteria were divided into several categories:

- **Time period**: contributions published from 2000 to the present were included. This threshold was chosen because it coincides with the spread of the first digital monitoring systems and the pioneering applications of machine learning in industry. Articles published before 2000 were excluded, as they were not relevant to the modern AI-based PHM paradigm.
- **Literature type**: included peer-reviewed articles published in internationally renowned journals and conference proceedings, as well as some notable doctoral theses. Technical reports, policy papers, working papers, newsletters, speeches, and contributions published solely in book chapters were excluded, as they lacked the same guarantee of peer review.
- **Language**: the selection was limited to contributions in English, to ensure terminological consistency and maximum scientific dissemination. Publications in other languages were excluded.
- **Accessibility**: only articles available in full-text were considered. Paywalled contributions for which institutional access was not available were excluded.
- **Content**: the most stringent and decisive criterion, which requires the inclusion only of studies in which artificial intelligence is explicitly applied to the domain of industrial PHM. Contributions involving non-AI-based methods, works focusing on irrelevant fields (e.g., medicine or agriculture), and articles dealing exclusively with diagnostics without prognostic elements were excluded.

This selection framework, combined with the search string, allowed us to build a bibliographic corpus consistent with the objectives of the thesis. The explicit criteria, in addition to ensuring transparency, allow other researchers to replicate the study, verify the choices made, or update them in the future with any new evidence.

**Table 3.2:** Inclusion and Exclusion Criteria

| **Criteria type** | **Inclusion** | **Exclusion** |
|---|---|---|
| Time period | Publications from 2000 to present | Publications before 2000 |
| Literature type | Peer-reviewed<br>Journal articles<br>Conference proceedings<br>PhD theses | Reports<br>Policy literature<br>Working papers<br>Newsletters<br>Speeches<br>Articles from books |
| Language | English | Non-English publications |
| Accessibility | Full-text available | Paywalled articles without institutional access |
| Content | AI applications in industrial PHM<br>Prognostics<br>RUL | Non-AI methods<br>Non-prognostic studies<br>Domains outside scope |

## 3.5 Screening process

After defining the search string and selection criteria, the next step was to screen the articles. This process aims to progressively reduce the set of documents initially identified, isolating those that are truly consistent with the review objectives. To ensure transparency and replicability, the process was represented using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) workflow, which is now considered the international standard for documenting the steps of a systematic review [13].

During the identification phase, searches of the Scopus, Web of Science and IEEE Xplore databases yielded a total of 519 contributions. An initial removal of duplicates excluded 83 entries, leaving 436 unique articles to be screened.

The next phase, screening, involved analysing titles and abstracts. At this stage, 245 articles were excluded because, despite containing similar keywords, they did not meet the established criteria (for example, studies in the medical or agricultural fields, or purely theoretical works without PHM applications). This left 191 contributions for further analysis, 12 of which could not be retrieved in full due to access limitations.

During the eligibility phase, articles were checked for relevance and systematically evaluated using a scoring grid, in line with recommendations on rigorous SLR protocols [11, 8]. Each contribution was assigned:

- a score from 1 to 5 for methodological quality (scientific rigour, clarity of objectives, appropriateness of methodology and robustness of results);
- a score from 1 to 5 for impact (relevance to the field of PHM, degree of innovation and potential for concrete industrial applications).

The sum of these two scores generated an overall score between 2 and 10. In line with the protocol, only articles with a final score higher than 8.5 were included in the final review corpus.

This evaluation phase enabled a transition from a basic inclusion/exclusion process to a comprehensive ranking system, ensuring that the final material not only aligned with the research questions, but also exemplified the highest standards of quality and scientific impact.

Figure 3.2 shows the distribution of scores, suggesting a bell-shaped curve: most articles are clustered between 7.5 and 9 points, while only a few reach the maximum score of 10.

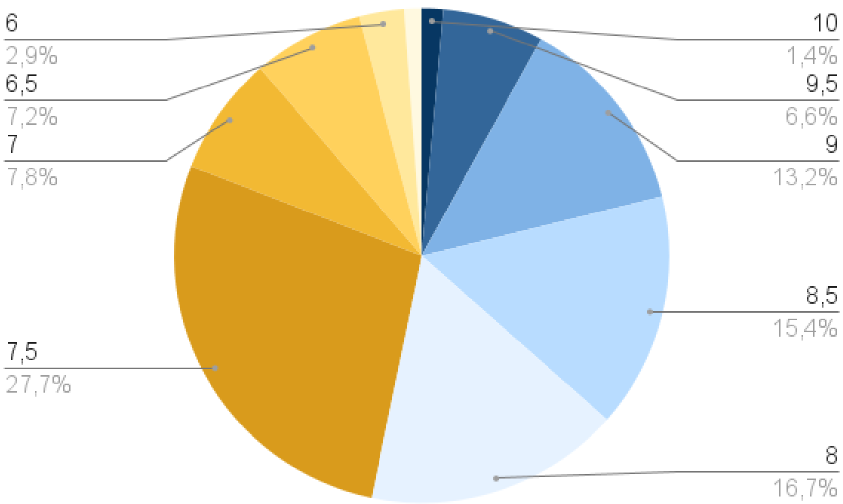


**Figure 3.2:** Distribution of quality and impact scores assigned to the articles.

The final phase, resulted in a final corpus of 54 articles, forming the empirical basis of the review. While these contributions varied in terms of application sector, methodology and level of technological maturity, they all met the quality and relevance criteria established in the protocol.

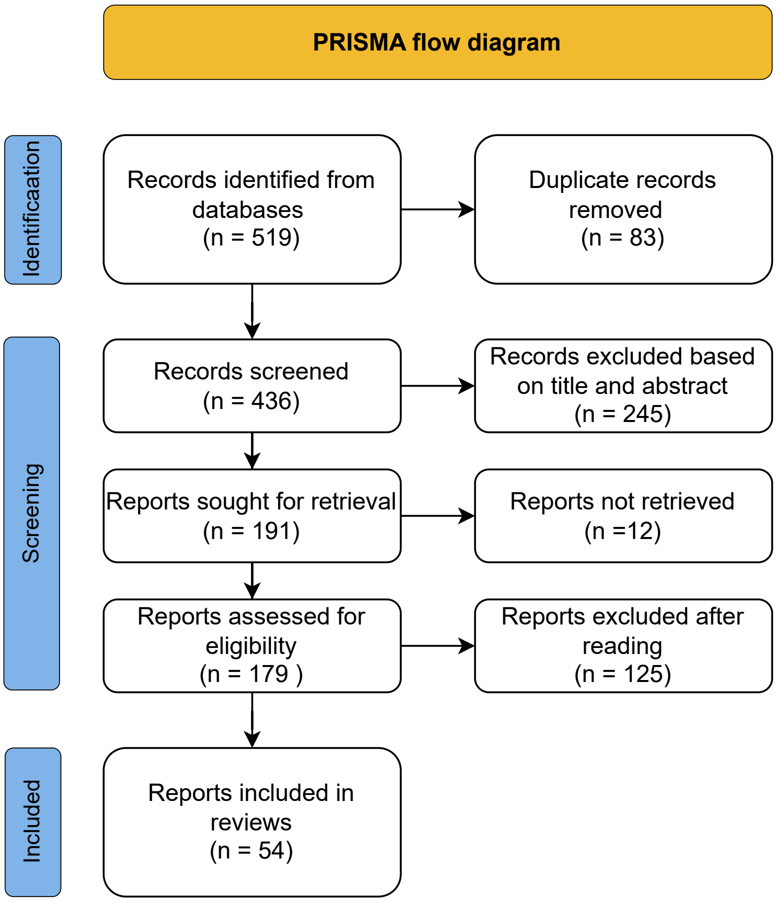


**Figure 3.3:** PRISMA flow diagram: the overall review process selected a total of 54 records [13].

The PRISMA diagram (Figure 3.3) summarises this process graphically: starting from over five hundred initial records, the selection process resulted in a small but robust set of works representing the state of the art in the application of artificial intelligence to prognostics and health management. This process is not only an exercise in formal rigour; it is also a fundamental methodological requirement, as it enables other researchers to replicate the procedure, verify its consistency and update it with new evidence in the future if necessary.

## 3.6 Final Articles and Macro-Area Distribution

The selection process encompassed all stages of the PRISMA protocol, from identification to qualitative assessment. This resulted in the inclusion of 54 articles. These form the core of the systematic review, the product of methodical filtering and evaluation that transformed an initial set of more than five hundred contributions into a coherent, analytically manageable body of work.

The distribution of these works, as shown in the macro-area graph (Figure 3.4), reveals a research geography that is far from random. The largest share belongs to *mechanical systems*, accounting for around one-third of the articles (33.3%). *Aerospace* follows with 27.8%, while *manufacturing* represents 22.2%. The smallest, yet still significant, macro-area is that integrating *automotive and wind turbines*, which together account for 16.7% of the total.

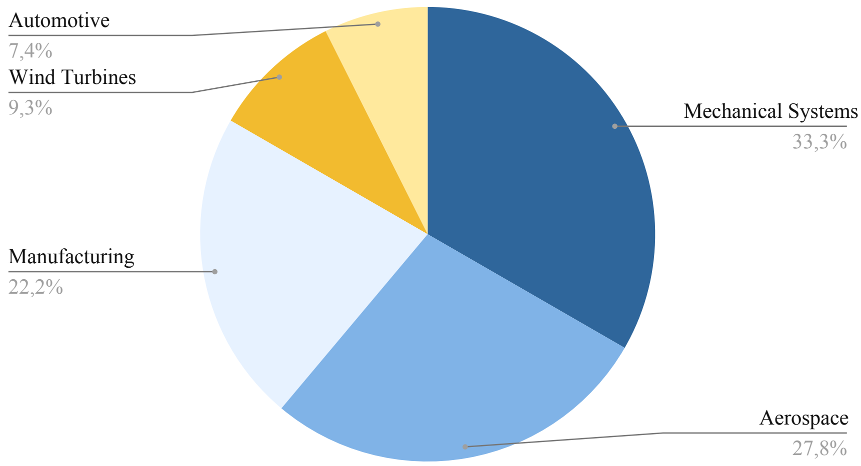


**Figure 3.4:** Distribution of the final corpus

As previously mentioned, the prevalence of mechanical systems can be explained by their ubiquity; they are present in every industrial sector and constitute an ideal testing ground for scalable PHM methodologies. Aerospace, on the other hand, appears to be the discipline's "pilot" sector, where the need for absolute safety and reliability has driven the immediate development of advanced prognostic tools. Manufacturing takes a different approach: here, PHM is systemic and extends beyond individual components to the entire production chain. The aim is to maintain the efficiency and resilience of industrial processes in line with the principles of Industry 4.0.

Finally, the integrated automotive–wind turbine macro-area is notable for its experimental and innovative nature. Although it is less well represented numerically, it contains some of the most interesting contributions in terms of future prospects, ranging from battery life prediction in electric vehicles to algorithms for fault prognosis using high-frequency SCADA data in wind turbines. In both cases, AI is required to operate in non-ideal contexts characterised by noisy, heterogeneous, and sometimes incomplete data, which makes these applications particularly valuable for testing the robustness of prognostic approaches.

Therefore, the distribution of areas should not be understood as a mere quantitative count, but rather as a reflection of the development trajectories of AI-supported PHM. The differences in weight between sectors demonstrate their varying stages of technological maturity, ranging from the excellence of the aerospace sector to the exploration of integrated processes in manufacturing and the sustainability frontier represented by emerging sectors.

From a methodological perspective, this chapter serves as a point of connection. It concludes the process of constructing the framework, from queries to filtering and screening to quality scoring, and paves the way for the subsequent analysis of results. Readers now have a well-defined corpus obtained through explicit, replicable, and transparent criteria. The adoption of evaluation thresholds (such as a minimum score of 8.5, resulting from the sum of quality and impact) has further strengthened the selection process by limiting subjectivity and clarifying decision-making.

**Table 3.3:** Summary of articles

| Citation | Authors | Year | Type | Area |
|---|---|---|---|---|
| [44] | Baptista & Henriques | 2022 | Article | Aerospace |
| [27] | Kim & Liu | 2020 | Article | Manufacturing |
| [45] | Asif et al. | 2022 | Article | Aerospace |
| [46] | Dhananjay et al. | 2024 | Article | Automotive |
| [47] | Nguyen & Medjaher | 2019 | Article | Manufacturing |
| [48] | Jiang et al. | 2024 | Article | Mechanical Systems |
| [49] | Chen | 2024 | Article | Aerospace |
| [50] | Bai & Zhao | 2023 | Article | Manufacturing |
| [51] | Caceres et al. | 2021 | Article | Manufacturing |
| [52] | Najdi et al. | 2025 | Article | Mechanical Systems |
| [53] | Krishna et al. | 2024 | Article | Automotive |
| [54] | Bala et al. | 2020 | Article | Aerospace |
| [55] | Zhang et al. | 2024 | Article | Mechanical Systems |
| [56] | Zhao & Zou | 2022 | Conference paper | Aerospace |
| [57] | Boujamza & Lissane | 2022 | Conference paper | Aerospace |
| [58] | Safavi et al. | 2024 | Article | Automotive |
| [59] | Mazaev et al. | 2021 | Article | Mechanical Systems |
| [60] | Ochella et al. | 2024 | Article | Mechanical Systems |
| [61] | Mishra et al. | 2021 | Article | Mechanical Systems |
| [62] | Pandit et al. | 2024 | Article | Wind Turbines |
| [63] | Zhao & Liu | 2022 | Article | Manufacturing |
| [64] | Wen et al. | 2021 | Article | Manufacturing |
| [25] | Liu et al. | 2022 | Article | Mechanical Systems |
| [65] | Wang et al. | 2021 | Article | Mechanical Systems |
| [66] | Deng et al. | 2021 | Article | Manufacturing |
| [67] | Berghout & Benbouzid | 2023 | Article | Aerospace |

Continued on next page

| Citation | Authors | Year | Type | Area |
|---|---|---|---|---|
| [68] | Liu et al. | 2020 | Conference paper | Mechanical Systems |
| [69] | Yuan et al. | 2024 | Article | Aerospace |
| [70] | Zhang et al. | 2025 | Article | Wind Turbines |
| [26] | Zhang et al. | 2020 | Article | Mechanical Systems |
| [71] | Boos et al. | 2024 | Conference paper | Mechanical Systems |
| [72] | Jaenal et al. | 2024 | Article | Manufacturing |
| [73] | Liu et al. | 2025 | Article | Mechanical Systems |
| [74] | Moros et al. | 2024 | Conference paper | Wind Turbines |
| [75] | Solís-Martín et al. | 2023 | Article | Manufacturing |
| [76] | Boujamza & Elhaq | 2024 | Conference paper | Aerospace |
| [77] | Berghout et al. | 2023 | Article | Aerospace |
| [78] | Desai et al. | 2020 | Conference paper | Wind Turbines |
| [79] | Mansouri et al. | 2017 | Conference paper | Mechanical Systems |
| [80] | Ma et al. | 2021 | Conference paper | Aerospace |
| [81] | Hu et al. | 2023 | Article | Manufacturing |
| [82] | Soualhi et al. | 2022 | Conference paper | Aerospace |
| [83] | Ensarioğlu et al. | 2023 | Article | Aerospace |
| [84] | Zhou et al. | 2023 | Article | Automotive |
| [3] | Wang et al. | 2023 | Article | Mechanical Systems |
| [85] | Wu et al. | 2023 | Article | Manufacturing |
| [86] | Listou et al. | 2019 | Article | Aerospace |
| [87] | Li et al. | 2023 | Article | Mechanical Systems |
| [88] | Magadán et al. | 2024 | Article | Mechanical Systems |
| [89] | Alomari & Andó | 2024 | Article | Aerospace |

Continued on next page

| Citation | Authors | Year | Type | Area |
|---|---|---|---|---|
| [90] | Chen et al. | 2022 | Article | Mechanical Systems |
| [91] | Schwendemann & Sikora | 2023 | Article | Mechanical Systems |
| [92] | De et al. | 2022 | Conference paper | Manufacturing |
| [93] | Verma et al. | 2022 | Conference paper | Wind Turbines |

# Chapter 4

# Interaction and comparison

The analysis conducted thus far has enabled us to outline the theoretical, methodological, and applicative aspects of Artificial Intelligence in the context of Prognostics and Health Management (PHM). Having explored the main techniques and use cases in various macro-areas, the next step is to correlate, compare, and critically discuss the findings.

The aim of this chapter is to intertwine the threads of analysis, observing how different approaches behave in specific contexts, assessing which metrics are truly comparable, and understanding the extent to which the choice of algorithm depends on factors such as accuracy, trust generation, computational costs, and data quality. The goal is to create a map of interactions where sectoral differences are not seen as obstacles, but as complementary viewpoints that enhance the broader conversation on PHM.

We will therefore transition from analysing macro-areas to identifying general trends (Macrotrends), and from comparing Deep Neural Networks with shallower approaches to reflecting on explainability, uncertainty, and reliability. Datasets, data quality, implementation limitations, and the most popular development platforms will also be discussed in a comparative exercise that aims to convey the complexity of the landscape without oversimplifying it.

## 4.1 Comparative Analysis of Macro-Areas

Comparing the various macro-areas of AI application in prognostics highlights the decisive influence of the industrial context on both methodological choices and research priorities. Each sector accentuates specific dimensions,safety, robustness, integration, sustainability, producing a mosaic of approaches that resists reduction to a single linear trajectory.

### 4.1.1 Aerospace

In this field, safety and certifiability remain the dominant imperatives. Studies employing the C-MAPSS benchmark to estimate the RUL of turbofans reveal a marked preference for deep architectures such as LSTM networks and CNNs [45, 76], often augmented with attention mechanisms [57] or transformer modules [71]. Alongside accuracy, researchers show increasing interest in uncertainty quantification, with probabilistic and ensemble models producing predictions framed by confidence intervals [67]. Interpretability also plays a crucial role: for example, [89] employ SHAP values to clarify links between sensor signals and degradation phenomena. The overall trend suggests a shift toward hybrid solutions where data-driven architectures and physics-based models converge to meet requirements of trust and transparency.

### 4.1.2 Manufacturing

In manufacturing, the main challenge lies in the complexity of multi-machine systems and their integration into the Industry 4.0 ecosystem. Multivariate Transformers have recently been applied to model intricate chemical processes and forecast multi-step failures [50], showing their ability to capture long-term dependencies among process variables. Other works propose GAN-based denoising to synthetically generate realistic data and mitigate the lack of large, high-quality datasets [44]. More than sparsity, the sector struggles with heterogeneity and uneven data quality, issues that stimulate the development of data-fusion pipelines and advanced normalization techniques.

### 4.1.3 Automotive

In the automotive domain, attention concentrates on energy storage and powertrain systems. Lithium-ion battery prognostics dominate research, with deep learning models estimating state of health and forecasting residual cycles under quantified uncertainty. Growing pressure for explainability reflects the needs of both manufacturers and end users, who demand not only accurate but also transparent models. The objective is a delicate balance: maintaining high predictive accuracy while addressing safety standards and sustainability goals.

### 4.1.4 Wind Energy

Wind turbines pose specific challenges: fluctuating operating conditions, harsh environments, and the difficulty of accessing remote plants. Research in this field often applies CNNs and autoencoders to detect anomalies in turbine signals [77]. Transfer learning emerges as a practical solution, allowing models trained in one wind farm to be adapted to others. The digital twin concept, integrating aerodynamic simulations with real-time sensor data, appears particularly promising in enhancing the robustness of degradation forecasts.

Ultimately, macro-areas differ not only in assets but also in their guiding values. Aerospace highlights safety and certification, mechanical systems emphasise robustness, manufacturing focuses on scalability and deployability, automotive research centres on sustainability, and wind power stresses long-term resilience. This diversity enriches the field, fostering methodological cross-pollination: SHAP applications in aerospace echo in battery diagnostics; transformers developed for industrial processes migrate to wind turbines; Echo State Networks originally used for bearings are repurposed to monitor lightweight embedded systems. Rather than fragmentation, this interplay reveals the creative exchange of methods across boundaries, driving the evolution of PHM towards a plural and adaptive discipline.

## 4.2 Macrotrends

When the collected works are considered in their totality, an evolutionary trajectory becomes evident. This trajectory concerns not only the technique but also the mindset with which artificial intelligence is applied to prognostics. In the early studies, the scene was dominated by models based on physical knowledge, built around thermodynamic equations or structural dynamics. The field of aerospace serves as a paradigmatic example. For an extended period, the estimation of the residual life of turbofans was predicated on engineering simulations to a considerable extent. The advent of datasets such as C-MAPSS, in conjunction with the seminal contributions of Asif et al. [45], which exemplified the superior performance of deep networks in comparison to conventional physics models, has precipitated a paradigm shift towards data-driven methodologies. This movement has not erased the past, but rather redefined it. Physics does not disappear; rather, it becomes a constraint, a framework, or a source of synthetic data. For instance, one may consider the utilisation of denoising GANs to generate degradation signals that are consistent with reality, yet based on simulations [44], or optimised Echo State Networks that balance engineering information and statistical learning [54].

Concurrent with this transition, the era of deep learning has emerged. In the space of a few years, CNNs, LSTMs, and, more recently, Transformers have become pervasive in almost all macro-areas. In the field of aerospace engineering, convolutions have demonstrated remarkable efficacy in the analysis of vibrational and acoustic signals [67], while recurrent architectures have been shown to be effective tools for capturing complex temporal dependencies, as evidenced by Boujamza and Elhaq [76]. It is evident that there is a plethora of work that extends beyond this scope: Boujamza and Elhaq [57] introduce attention mechanisms in LSTM models, while Boos et al. [71] explore alternative attention modules in Transformers to optimise RUL prediction. In the field of manufacturing, Bai and Zhao [50] demonstrate the capacity of a multivariate Transformer to manage intricate chemical processes, predicting both the present state and the subsequent progression of multiple variables in unison. It is noteworthy that the enhancement in accuracy is accompanied by the flexibility of these architectures, whereby each sector adapts them to its specific requirements. This adaptation involves the integration of convolutions and sequences, attention mechanisms and dimensional reduction, resulting in the emergence of authentic local dialects within the domain of deep learning.

However, while the enthusiasm for deep networks is palpable, there is a growing awareness of their limitations. The opacity of the models, the necessity for substantial amounts of data, and computational costs have directed the community towards a third direction: hybrid integration, or, to utilise the more prevalent label, physics-informed machine learning. In the field of aerospace engineering, Berghout et al. [77] have proposed ProgNet, a model that not only predicts degradation but also enforces its limits to feasible physical curves. In a similar vein, Alomari and Andó [89] have demonstrated the efficacy of SHAP techniques in enhancing transparency within complex models by assigning importance to sensory variables and thereby rendering the predictive process more comprehensible. In the field of wind energy, the integration of aerodynamic simulations with SCADA data has been shown to enhance the robustness of models that might otherwise be susceptible to overfitting limited datasets [67]. The underlying message is unambiguous: artificial intelligence should not be employed to substitute for physics, but rather to augment it, thereby expanding its scope.

Another strong signal that has emerged across the board is the growth of digital twins and, in parallel, federated learning. Digital twins are now more than just an evocative concept: in automotive batteries, they allow the simulation of extreme charge and discharge cycles without risking actual damage; in wind farms, they allow the testing of rare but potentially catastrophic failure scenarios. The idea that a model can exist in a parallel virtual environment, constantly fed by real data, has changed the way we think about prognostics, transforming it from a precise prediction to a continuous process of simulation and validation. At the same time, federated learning is emerging as a response to companies' growing reluctance to share sensitive data. Although still in its infancy, several studies discuss the possibility of training models distributed across vehicle fleets or turbine networks, maintaining the data at their source sites but exchanging only updated parameters. Looking ahead, this solution could become the cornerstone for overcoming the problem of accessing industrial data, a frequent topic in the literature yet a frustrating one for researchers.

The overarching question concerning all these trajectories is that of trust. The concepts of explainability and trustworthiness have become pivotal across the entire landscape. In order to be considered adequate, a model must not only predict accurately, but also provide a rationale for why it does so. Furthermore, it is essential that the model is capable of quantifying its uncertainty, demonstrating resilience to noise, and being replicable with external data. In the field of aerospace engineering, Alomari and Andó [89] emphasise the crucial role of explainability in predictive models, asserting that the absence of explainability renders predictions unfit for industrial application. In the field of manufacturing, Bai and Zhao [50] underscore the significance of process complexity in the context of algorithm utilisation. They contend that the utilisation of numerical output algorithms alone is insufficient, emphasising the necessity of providing legible justifications to ensure the effective adoption of such algorithms. As Berghout et al. [67] observe, in the absence of well-calibrated confidence intervals, a RUL model risks introducing uncertainty rather than reducing it. Similarly, Bala et al. [54] caution against the pitfall of overly complex models for edge scenarios, emphasising that confidence also hinges on computational efficiency.

A comprehensive analysis of the emerging macrotrends reveals a dynamic landscape. The discipline of prognostics is moving in multiple directions simultaneously, from physical models to neural networks, from CNNs to Transformer architectures, from real datasets to digital twins, and from "black box" algorithms to explainability practices. The fundamental question guiding research in this field is no longer simply "how accurate is the prediction?" but rather, "how much can it be trusted, what is the cost of its training and integration, and how compatible is it with the rules and constraints of my industry?" Consequently, the field of artificial intelligence for PHM cannot be regarded as a monolithic entity; rather, it is a dynamic domain in which technical innovations and industrial needs intertwine and transform each other.

### 4.2.1 From Physics-Based Modelling to Data-Driven Prognostics

One of the most obvious macrotrends to emerge from the literature review is the progressive abandonment of model-based approaches, which are rooted in a physical and deterministic description of phenomena, in favour of data-driven methods. To fully understand the significance of this shift, it is important to clarify the nature of the latter paradigm.

Data-driven methodologies rely on data rather than a predefined theoretical model to reveal regularities. Put simply, while the model-based approach starts from known physical laws (e.g. differential equations, thermodynamic and electrochemical models) and uses them to simulate a system's behaviour, the data-driven approach assumes that the observed data already contains the necessary information to predict a system's future state or estimate its remaining life. The algorithm does not know the degradation equations a priori, but rather learns them implicitly through statistical correlations and latent structures that emerge during training.

This approach is made possible by the growing availability of large amounts of sensory data: vibrations, temperatures, pressures and electrical signals collected from complex systems in operation. Machine learning and deep learning algorithms, from artificial neural networks to autoencoders, CNNs, LSTMs and transformers, build abstract representations of system behaviour, capable of capturing temporal dependencies and nonlinear relationships that escape classical modelling. Unlike a physical model, which necessarily simplifies reality, the underlying idea is that a data-driven model can adapt to the complexity of raw data and uncover hidden patterns that are not immediately apparent to engineers.

In critical contexts such as aerospace, physical models have long been the preferred solution. Turbofan thermodynamic equations and mechanical fatigue models, for example, were considered irreplaceable tools for making reliable predictions. However, benchmark datasets such as C-MAPSS have made it possible to train complex neural networks that can outperform traditional deterministic models. Asif et al. [45] demonstrate how a deep neural network can more accurately estimate the RUL of motors, while Boujamza and Elhaq [57, 76] use LSTM architectures enriched with attention mechanisms to improve their networks' ability to capture long-term dependencies in multichannel signals.

A similar shift is evident in rotating mechanics, where datasets such as PRONOSTIA and XJTU-SY provide a basis for comparison. Here, physical modelling of fatigue degradation still provides a framework, but its applicability is limited by variable operating conditions and high data noise. Data-driven approaches allow robust information to be extracted even in the presence of complex signals: Baptista and Henriques [44] use autoencoders and CNNs to detect latent vibration patterns, while Bala et al. [54] propose an optimised echo state network that combines predictability and low computational cost.

In manufacturing, this transition is even more pronounced: multivariate chemical processes cannot be fully described by deterministic equations because they involve complex interactions and emergent phenomena. Bai and Zhao [50] demonstrate how a multivariate transformer can model nonlinear dynamics and perform multi-step predictions with superior accuracy. They emphasise that the data-driven approach is not merely an alternative, but the only viable strategy for such complex systems.

A similar argument can be found in the automotive and wind energy sectors. While electrochemical models of lithium batteries are rigorous, they are inflexible when it comes to representing real-world degradation under variable operating conditions. Using LSTMs and networks with attention mechanisms, Berghout et al. [67] enable us to anticipate ageing phenomena that cannot be predicted by equations alone. Secondly, wind turbines present further challenges as traditional aerodynamic models cannot capture the variability of wind and weather conditions. However, the use of SCADA data has enabled the training of CNN and autoencoder models capable of estimating RUL with greater reliability, as shown by Berghout and Benbouzid [67].

The picture emerging from the fifty-four articles analysed shows that the data-driven approach is not a transitory trend, but a structural shift. Although physical models remain essential for providing theoretical constraints, generating synthetic data, and validating models, reliable prediction increasingly relies on learning from data. Examples such as ProgNet by Berghout et al. [77], which integrates physical curves within a deep network, and the explainability applications proposed by Alomari and Andó [89] confirm that physics will not be abandoned in the future, but rather that hybrid architectures combining scientific rigour and predictive capacity will be constructed.

### 4.2.2 Physics-informed hybridisation

The transition from deterministic, model-based approaches to the current era of purely data-driven methods in prognostics is neither linear nor definitive. Instead, the trajectory of research in recent years indicates a resurgence of physics, which is no longer perceived as a cage, but rather as an ally. This represents the paradigm of physics-informed hybridisation, whereby established laws, thermodynamic constraints and engineering knowledge are not disregarded, but rather integrated into neural networks. The concept is not a recent innovation; however, its execution has undergone a radical transformation, evolving from a rudimentary juxtaposition of independent components to a genuine integration of training methodologies.

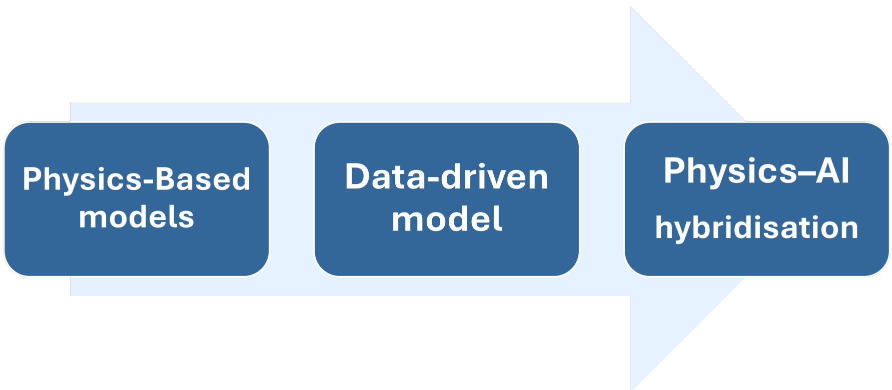


**Figure 4.1:** From physics to hybrid models

In the early stages of this story, hybrid approaches were rather rudimentary. A number of works have exploited simulations based on physics models in order to generate synthetic data with which to train machine learning algorithms. One example of this is aerospace applications, where simulators such as C-MAPSS were used as 'surrogates' to feed networks [87]. Hybridisation was therefore more of a statistical artefact than a genuine intertwining. The data was produced by a digital twin or a differential equation model, and the network passively learned from what it received.

However, as time progressed, a discernible shift in the prevailing literature became evident. As demonstrated by Baptista and Henriques [44], researchers operating within the domain of rotating systems have established that the mere existence of simulated data is insufficient in and of itself. Rather, it is imperative that physical constraints are incorporated directly into the cost function, thereby providing a framework for learning and ensuring predictions that are in accordance with conservation laws. In a similar vein, Zheng et al. [55], working in the field of batteries, have incorporated regularisation functions into their RUL models that penalise results that are incompatible with experimentally observed electrochemical cycles.

A particularly fruitful chapter concerns the use of Bayesian filters. In the present context, the fusion is characterised by a heightened degree of subtlety, which cannot be attributed merely to the imposition of constraints. Rather, it is imperative to recognise uncertainty as an integral component of the prediction process. In the work of Berghout and Benbouzid [67] on wind turbines, the combination of unscented Kalman filters with LSTM networks allows not only to estimate residual life but also to provide confidence intervals aligned with aerodynamic models. In a related study, Chen et al. [90] proposed the utilisation of regularised recurrent neural networks with particle filters, demonstrating their capacity to facilitate the dynamic updating of estimates based on probabilistic constraints derived from failure physics. This use of filters is indicative of a conceptual maturation, whereby hybridisation is no longer regarded as a mere sum, but rather as a dialogue between engineering epistemology and statistical learning.

In parallel, another trend has emerged: neural networks with integrated physical constraints. Berghout et al. [77] have demonstrated how to integrate theoretical degradation curves directly into the ProgNet model's layers so that it cannot learn implausible behaviours. Boujamza and Elhaq [76] extended the LSTM paradigm by introducing attention mechanisms that respect constraints derived from the mechanical properties of materials, thereby reducing false alarms and inconsistent predictions. In the context of complex manufacturing sectors, Bai and Zhao [50] demonstrated that the efficacy of multivariate transformers is contingent upon the incorporation of loss functions augmented with regularisations pertaining to energy balances.

Another element marking the transition to the current phase is the role of digital twins. Initially conceptualised as autonomous virtual replicas, digital twins are progressively integrating into hybridised architectures. Asif et al. [45] demonstrate how, in the aeronautical context, the digital twin provides rare or extreme scenarios, while the neural network bridges the gap with real operating conditions. In this sense, hybridisation is no longer static but dynamic: a continuous exchange of data between the physical model and the neural model.

The most recent literature, for instance that of Boos et al. [71], also signals a convergence towards forms of physics-guided attention. To illustrate this point, consider the Transformers model. These models are no longer trained as tabula rasa; their self-attention modules are constrained to respect predefined structural relationships, thereby improving the consistency of results. The focus has shifted from rectifying deep learning errors after they have occurred to the proactive shaping of the architecture to ensure adherence to the established principles from the very beginning.

A thorough examination of the 54 contributions reveals a discernible transition: initially, a "block-based" approach dominated, with physics on one side and network on the other; subsequently, a more mature phase characterised by physics being directly embedded in the algorithms, the objective functions, and the layers themselves. This transformation also reflects an epistemic tension: the need to reconcile the predictive power of deep models with the need to maintain engineering plausibility, trust, and transparency.

What is emerging today is therefore not a convenient compromise, but a new grammar of prognostics. In aeronautical systems, CNN–LSTM fusion with thermodynamic constraints is becoming a de facto standard. In rotating systems, autoencoders and CNNs are regularised with fatigue laws [44]. In the context of batteries, attention models are trained within the physical constraints of electrochemical cycles [55]. In manufacturing processes, transformers constrained by energy balances are showing their first convincing applications [50]. In wind turbines, the combination of Bayesian filters with neural networks has been shown to produce not only values, but reliable probabilistic predictions [67].

Ultimately, physics-informed hybridisation represents the third major evolutionary phase of prognostics. After the eras of purely physical models and entirely data-driven methods, a new paradigm is emerging in which laws and data coexist within hybrid architectures. This is not a return to the past, but a leap forward: physical constraints no longer suffocate the network; they guide it. Rather than ignoring consolidated science, deep learning incorporates it to offer predictions that are accurate, plausible and credible. The most recent contributions clearly signal this direction, which appears destined to set the standard for the next generation of prognostic systems.

### 4.2.3 Deep Learning

A second macrotrend that clearly emerges from the literature is the increasingly widespread adoption of deep learning approaches. While the first turning point involved the gradual abandonment of purely physical models in favour of data-driven approaches or, most recently, of hybrid approaches, the second relates to the growing prevalence of architectures such as CNNs, LSTMs and, more recently, transformers. Not only does the literature demonstrate the adoption of these models, it also shows their ongoing hybridisation. This is in contrast to what are commonly referred to as shallow methods that have been the standard for years.

The early applications of PHM were fundamentally based on shallow models, which included support vector machines, random forests and one or two layer neural networks. They ensured robustness, reduced computation times and offered a degree of interpretability. In work on bearing failures, for instance, methods such as SVM or advanced linear regression produced comparable results when the datasets were small or the features had been manually extracted (e.g. vibration indices, Fourier transforms, statistical moments). However, with the increasing complexity of systems and the amount of available data, these approaches have shown clear limitations, including excessive reliance on feature engineering, poor adaptability to multichannel signals and difficulty representing complex nonlinear phenomena [26] [87].

At the same time, LSTM networks and their derivatives (such as Gated Recurrent Units (GRUs)) have transformed tasks where temporal dependencies are crucial. For example, predicting the remaining useful life of aircraft engines (C-MAPSS dataset) or estimating the degradation of lithium-ion batteries greatly benefits from the ability of LSTMs to store long-term information. Studies by Wu et al. [85] and Kim et al. [27] demonstrate that LSTMs outperform simpler recurrent networks. Furthermore, variants such as bidirectional or dilated RNNs [76] [90] have been shown to further extend predictive capabilities.

In recent years, the field has been enriched by the arrival of transformers, which have introduced self-attention mechanisms already established in natural language to prognostics. In contrast to recurrent networks, transformers are capable of capturing dependencies even over very long sequences without the necessity of temporal loops. In their 2023 article, Bai and Zhao [50] demonstrate the potential of multivariate transformers in modelling complex chemical processes, thereby achieving multi-step predictions of a higher degree of accuracy than is typical of traditional approaches. In contrast, Boos et al. [71] experimented with customised attention modules for RUL prediction, achieving results that surpassed those of CNNs and LSTMs. This observation suggests that transformers are not merely a passing fad, but rather a foundational element that is poised to consolidate its position, particularly in contexts characterised by heterogeneous and high-dimensional data.

The comparison between shallow and deep approaches is therefore a matter of balancing predictive power and cost. In scenarios where data is limited and critical variables are well understood, shallow models have been shown to be advantageous [54]. Conversely, deep approaches predominate in contexts characterised by an abundance of datasets and the presence of intricate failure phenomena. However, this augmentation of power is accompanied by certain drawbacks, namely protracted training periods, elevated resource utilisation and diminished transparency.

It is important to note that a number of authors have emphasised that the frontier is not characterised by total replacement, but rather by integration. Hybrid pipelines have been developed in which a CNN extracts features from the signals and a shallow regressor (e.g. random forest or gradient boosting) performs the final prediction [63]. In other studies, unsupervised autoencoders have been utilised for the purpose of compressing the data. Thereafter, the compact representation is passed to traditional models. This finding serves to substantiate the notion that the so-called “shallow vs. deep” dichotomy is not without its ambiguities, but rather represents a methodological continuum.

The existing literature also demonstrates how deep learning is being adapted in divergent ways according to the sector. In the aerospace sector, for instance, hybrid CNN–LSTM architectures dominate [45] [67]. In rotating mechanics, one-dimensional CNNs and autoencoders prevail [44]. In the field of batteries, recurrent and attention-based approaches are the most promising [55]. In manufacturing, transformers for complex multivariate processes are beginning to gain ground [50]. This adaptability, rather than accuracy alone, is indicative of the sophistication of deep learning: rather than a monolithic building block, it is a set of tools tailored to the specific requirements of the domain.

In summary, the emergence of deep learning signifies not only a technological advancement, but also a cultural transformation. The field of prognostics has evolved from the utilisation of manual descriptors and rudimentary models to a paradigm in which networks autonomously learn multilayer structures. CNNs, LSTMs and transformers are no longer considered experimental tools, but rather they have become well-established methods within the field of PHM. In the context of specialised applications where efficiency and interpretability are paramount, shallow approaches will persist. However, the prevailing trend in mainstream applications is towards greater complexity and depth. The imminent challenge is not merely one of precision, but rather, the capacity to harmonise predictive efficacy with transparency, trust, and industrial applicability.

## 4.3 Approches

The world of AI-based prognostic methodologies is not uniform, but a diverse range of approaches that have become popular at various times and in different contexts. These approaches often reflect the complexity of the available data and the needs of different industrial sectors. This section takes you on a journey through the main architectural families that have emerged in recent literature, including recurrent neural networks (RNNs and LSTMs), convolutional neural networks (CNNs), autoencoders, echo state networks, transformers and hybrid models. Each class has its own advantages and limitations and is used for specific applications in areas ranging from aerospace and wind energy to lithium-ion batteries and industrial machinery.

### 4.3.1 Recurrent Neural Networks

Recurrent neural networks were among the first tools introduced in AI-based prognostics due to their ability to model temporal sequences. The basic principle is to update a hidden state at every instant that incorporates past information, allowing us to capture dependencies along the temporal evolution of the data. This mechanism makes them particularly suited to processing signals that characterise industrial and aerospace systems, such as vibrations, pressure or temperature. However, classical RNNs quickly revealed their limitations; the difficulty of propagating the gradient over long sequences results in a loss of long-term memory, reducing predictive capacity in complex scenarios. Consequently, they are now more commonly used as an experimental baseline than as a practical solution, as evidenced by numerous comparative studies with more advanced architectures [45, 83, 87].

The real breakthrough came with the introduction of Long Short-Term Memory (LSTM) architecture, which features a gating mechanism that can regulate the input, retention and output of information in a memory cell. This architecture overcomes the vanishing gradient problem and enables the learning of long-term correlations in data. The rapid and widespread adoption of these models in prognostics and health management is evident. In turbofans, for instance, Asif et al. [45] demonstrated that a deep LSTM model, applied to the NASA C-MAPSS dataset, surpasses conventional physics models in terms of accuracy. Similarly, Boujamza and Lissane Elhaq [57] enhanced LSTMs with attention mechanisms, enabling dynamic weighting of contributions from diverse sensory channels. Boujamza and Elhaq [76] subsequently proposed a further evolution, introducing dilated recurrent networks that can extend the time horizon without incurring excessive computational costs.

Applications are not limited to aerospace. In the lithium-ion battery sector, for example, Dhananjay Rao et al. [46] developed an optimised multilayer LSTM with hyperparameter tuning techniques to predict the RUL of batteries under variable operating conditions. Meanwhile, Zhou et al. [84] demonstrated how multiscale health indicators can be integrated with sequential models to reliably estimate both RUL and health status. In bearings, Najdi et al. [52] proposed a carefully tuned Res-LSTM capable of adapting to changing operating conditions. Li et al. [87], meanwhile, combined multi-branch CNNs and bidirectional long short-term memory (BiLSTM) networks to simultaneously capture local patterns and long-term dynamics. LSTMs have also found widespread use in the wind energy sector. Verma et al. [93] used recurrent networks to analyse high-frequency SCADA data, and Zhang et al. [70] employed LSTMs in integrated models for pump prognostics in offshore turbines.

A distinctive aspect emerging from the literature is the tendency to not use LSTMs as isolated models, but rather to integrate them into hybrid architectures. The CNN-LSTM combination, in particular, has become almost standard in domains where integrating the CNNs' ability to extract local patterns from time-frequency representations with the LSTMs' sequential memory is crucial. Examples of this approach are found in Wang et al. [65, 3] and Liu et al. [25], who demonstrated how merging the two techniques improves performance in estimating the RUL of bearings. In other cases, such as in the work of Chen et al. [90, 49], LSTMs have been combined with hierarchical attention modules and transformative components, demonstrating their ability to integrate into increasingly complex architectures.

Equally relevant is the issue of scarcity of labelled data and variability of operating conditions. To address these limitations, several studies have experimented with semi-supervised and transfer learning approaches. Listou Ellefsen et al. [86] showed how semi-supervised training improves LSTM performance by reducing the need for complete labels, while Schwendemann and Sikora [91] demonstrated the effectiveness of transfer learning in bearings by adapting CNN-LSTM models to different types with limited data. Wen et al. [64] and Zhao and Liu [63] instead adopted domain adaptation strategies, which allow for maintaining predictive accuracy across platforms and conditions.

The growing maturity of LSTMs is accompanied by interest in uncertainty quantification, which is crucial in critical fields such as aerospace and energy. Kim and Liu [27] developed a Bayesian framework based on LSTMs to estimate confidence intervals, while Caceres et al. [51] introduced probabilistic RNNs capable of modelling the randomness of degradation processes. Other works, such as those by Mazaev et al. [59] and Ochella et al. [60], have expanded the discussion to the entire Bayesian ecosystem, emphasising the need to combine predictive accuracy with reliability measures.

The picture that emerges from the fifty-four articles considered is that of a technology that has now reached a state of consolidation: LSTMs have established themselves as one of the most robust and versatile tools for data-driven prognostics, with applications ranging from aerospace to batteries, from bearings to wind power, and even complex industrial processes. They are not without limitations: high computational costs, sensitivity to hyperparameters, and the need for explainability strategies [89, 75] remain critical factors. Nevertheless, even in the Transformer era, LSTMs remain central, serving as both a baseline and an integral component of hybrid architectures combining temporal memory, local feature extraction, and global attention. In other words, they are a necessary step in the evolutionary trajectory of AI-based prognostics: a memory grammar that has enabled the modelling of complex phenomena and which continues to be invaluable in real-world and industrial scenarios.

### 4.3.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs), developed for image recognition, are based on a simple yet powerful principle: rather than connecting each neuron to all inputs as in a fully connected network, they utilise convolutional filters that run locally over the data. Each filter is a small matrix of weights that calculates local linear combinations and produces activation maps when applied to portions of the signal. These filters are learnt during training and become specialised in detecting characteristic patterns, such as regular oscillations, sudden changes and degradation transients.

In the case of one-dimensional signals, typical of mechanical or aerospace systems, the convolution runs over time and captures local patterns in the time series. This means that a CNN can recognise, for example, the characteristic oscillation of a faulty bearing or an anomalous variation in the airflow of a turbofan. When the data is transformed into two-dimensional representations, such as spectrograms or scalograms, CNNs work exactly as they do with images, identifying geometric structures in the time-frequency planes.

Alongside convolutional layers, the architecture includes pooling layers, which reduce dimensionality while retaining essential information, and fully connected layers, which translate the learned features into final outputs, such as the estimated RUL. A crucial aspect of this approach is weight sharing, whereby the same filter is applied at all positions. This has two main benefits. Firstly, it drastically reduces the number of parameters. Secondly, it makes the network less prone to overfitting than dense networks. Moreover, the hierarchical configuration of CNNs enables the initial layers to acquire rudimentary features (e.g. minor variations, local harmonics), while progressively more sophisticated layers develop abstract representations of intricate phenomena.

This layered capability is evident in many prognostic applications. In turbofans, Baptista and Henriques [44] demonstrated that 1D CNNs can isolate hidden vibrational patterns and clean data from noise thanks to an integration with GAN denoising. In the field of bearings, Wang et al. [65, 3] employed CNNs to capture intricate spatiotemporal patterns, even introducing convolutional attention mechanisms that enable them to weight the most pertinent contributions from diverse sensors. In the domain of manufacturing, Deng et al. [66] developed dilated multiscale CNNs, which have been shown to expand the receptive field of filters and capture both rapid details and slow trends in signals.

The efficacy of CNNs is not solely attributable to their precision; their robustness is equally noteworthy. Local filters have been shown to acquire the capability to identify structures that remain valid despite the presence of noise or moderate variations in operating conditions. In the context of high-speed aeronautical systems, Berghout and Benbouzid [67] demonstrated that combining CNNs with collaborative feature selection strategies enhances prognostics when compared to purely recurrent networks. In the field of battery research, Zhou et al. [84] utilised CNNs to extract multiscale indicators from charge and discharge cycles. These indicators are then fed into more complex machine learning models.

Of course, there are limitations: CNNs excel at identifying local patterns, but lack intrinsic temporal memory. For this reason, they are often combined with LSTMs or GRUs, which add the ability to model long-term dynamics. An example is provided by Liu et al. [25], who integrated CNNs and LSTMs with time-frequency representations to improve RUL estimates of bearings, or the CALAP model by Wu et al. [85], which fuses CNNs and LSTMs with dual attention modules. Adaptation to new conditions has also been addressed using CNNs: Wen et al. [64] and Zhao and Liu [63] used convolution-based domain adaptation approaches to transfer knowledge from one platform to another or from one operating regime to another.

In short, CNNs function as hierarchical local pattern extractors, capable of transforming raw, noisy data into stable, predictive representations. In prognostics, this translates into the ability to identify early signs of degradation hidden in vibrations, pressures, or electrochemical signals. It is precisely this modular and interpretable nature that makes them one of the fundamental building blocks of almost all PHM architectures today, rarely used alone but almost always as part of a hybrid ecosystem with recurrent networks, attention mechanisms, or physical constraints.

### 4.3.3 Autoencoders

Autoencoder networks have gradually gained significant traction in prognostics, not so much for their ability to directly predict a component's RUL, but for their role as tools for unsupervised learning and the construction of robust latent representations. The idea of compressing data into a low-dimensional space and then reconstructing it dates back to established traditions in signal processing, but with the advent of deep learning, this principle has found new vitality. Pioneering work in PHM, such as that of Listou Ellefsen et al. [86] on semi-supervised architectures for turbofans, highlighted the use of latent encoders to reduce label dependence. In parallel, comparative studies on data-driven methodologies [47] highlighted the need for algorithms capable of extracting robust features even when degradation is not yet evident.

An autoencoder consists of two parts: an encoder, which progressively reduces data until it reaches a bottleneck, and a decoder, which attempts to reconstruct the original input. The model is trained by minimising the difference between input and reconstruction, and the quality of the resulting latent code depends on the trade-off between representational capacity and reduction constraints. In engineering systems, this mechanism is particularly useful because it allows for the automatic distinction between variations due to noise and patterns that correspond to actual degradation trajectories. It is no coincidence that, already in early studies on batteries [26], it was observed that compressing electrochemical signals through dimensionality reduction allowed for the highlighting of latent ageing patterns, difficult to capture with traditional techniques.

Baptista and Henriques [44] provided a concrete example of how autoencoders can be used in conjunction with generative techniques in turbofans. Their 1D-DGAN model combines a denoising autoencoder with a generative adversarial network in order to isolate hidden vibrational signals and produce more robust predictions in the presence of noise. This approach highlights an important point: the autoencoder is not only useful for compression, but can also act as an intelligent filter to improve the quality of downstream features. In the bearing domain, Magadán et al. [88] used denoising variational autoencoders on multiple datasets (IMS, FEMTO and XJTU-SY) and obtained models that could maintain high performance even in cross-domain conditions. De Beaulieu et al. [92], on the other hand, emphasised the usefulness of AEs in unsupervised contexts. They used an LSTM-based encoder-decoder to estimate a long-term health index, from which the RUL was subsequently inferred.

More recent developments have led to regularised versions of autoencoders, in which the latent code is constrained or filtered. Zhang et al. [55] introduced a residual shrinkage encoder to construct robust health indices in bearings, limiting the impact of noise and producing a smoother, more interpretable degradation curve. In the battery sector, Zhou et al. [84] also exploited AEs integrated with multiscale indicators. They showed that using health indicators derived from compressed electrochemical signals allows for more reliable predictions of both the state of health and the RUL.

The role of autoencoders is not limited to diagnostics. They are increasingly being applied in the generation of synthetic data, a crucial aspect when real datasets are unbalanced or lacking in fault examples. Variational autoencoder

(VAE) models and hybrid GANs are moving in this direction, allowing for the simulation of plausible degradation signals, improving the training of predictive networks. This theme also appears in more general works on model robustness and transferability, such as those by Wen et al. [64] and Zhao and Liu [63], who show how a good latent representation facilitates domain adaptation and reduces dependence on the original dataset.

These approaches also demonstrate their potential when integrated with hybrid architectures. For example, CNNs and LSTMs benefit from features preprocessed by autoencoders. Liu et al. [25] combined convolutions and latent representations to estimate the RUL of bearings, and Li et al. [87] combined BiLSTMs and multibranch CNNs with dimensionality reduction modules. This demonstrates that creating well-structured latent codes is key to achieving high performance. In more advanced contexts, such as those described by Chen [49] and Boos et al. [71], Transformers can be fed features generated by an autoencoder, reducing the risk of overfitting and making the network more stable.

Naturally, limitations and cautions remain. Reconstruction error, while useful, does not always directly correlate with the probability of imminent failure, and the interpretation of latent representations often remains opaque. The need for greater transparency has led to the integration of AEs with explainability tools: Alomari and Andó [89], for example, discussed the use of SHAP to interpret temporal features in aerospace PHM models, highlighting how even compressed representations must be made readable to increase operator confidence.

Looking at the fifty-four articles considered overall, a clear trajectory emerges. It begins with simple autoencoders, useful for reducing data dimensionality; it moves on to variational autoencoders and denoising models, capable of improving robustness; and it ends with penalised encoders and hybrid GANs, which expand their use to synthetic data generation and cross-domain adaptation. In all cases, AEs emerge as intermediate, not final, tools: microscopes capable of delving into data and producing representations that make subsequent predictive steps more reliable. This is what makes them an integral part of the PHM ecosystem today, on a par with CNNs, LSTMs, and Transformers, playing a complementary yet crucial role in building reliable and transferable predictive models.

### 4.3.4 Echo State Networks

Echo State Networks (ESNs), although less widespread than CNNs, LSTMs, or Transformers, represent a significant step in the evolution of architectures for prognostics and predictive maintenance. The underlying idea stems from the reservoir computing paradigm, developed in the early 2000s and revived in recent years in light AI applications. In a field like PHM, where increasingly complex and deep models have taken centre stage [45, 57, 50], ESNs offer a counter-current approach, which focuses not on extreme depth but on computational parsimony. This characteristic has emerged strongly when compared to more complex architectures: comparative studies demonstrate that, despite slightly lower accuracies, ESNs maintain an ideal efficiency profile for distributed computing scenarios and embedded applications [62, 58].

An ESN is composed of three fundamental blocks: the input layer that receives the data, the reservoir, a recurrent network with sparse connections and randomly initialised weights, and the output layer, the only one actually trained. Unlike RNNs or LSTMs, in which gradient backpropagation involves the entire network (resulting in high costs and the risk of a vanishing gradient), in ESNs the reservoir remains static and acts as a high-dimensional nonlinear map. Its function is to transform the input sequence into internal dynamic trajectories, from which the output layer extracts the prediction using linear regression methods. This means that the training cost is dramatically reduced, a crucial aspect for real-world industrial applications, where complex models such as Transformers [49, 71] or multi-branch CNNs [87] can be too costly to implement on edge devices.

The work of Bala et al. [54] represents a paradigmatic demonstration of the potential of ESNs in aeronautical contexts. The authors used an ESN enriched with a metaheuristic optimisation algorithm, the Grasshopper Optimisation Algorithm, to improve the reservoir configuration and better fit it to the prognostic data. This approach reduced the variability of results, often related to the random choice of initial parameters, and achieved competitive performance in failure prediction. This is a particularly interesting result when compared to more traditional models such as LSTMs [45, 83] or hybrid CNN-LSTMs [65, 85], which, while more accurate, require much greater training time and computational resources.

ESNs have primarily been applied in domains where computational lightness is a strategic advantage. In the field of aerospace systems, datasets such as C-MAPSS provide a valuable basis for comparison, and the work of Bala et al. [54] demonstrates the effectiveness of ESNs even in scenarios that are usually dominated by LSTMs and Transformers. In rotating systems where vibration data requires real-time processing, ESNs show particular promise: reservoir computing can be implemented directly on board, reducing the need to send large amounts of data to central servers [3, 48]. In the wind energy sector and distributed energy applications, having lightweight models that can be periodically updated in edge computing provides a competitive advantage, as discussed in comparisons between deep models and lighter solutions [62, 74].

Another attractive feature of ESNs is their ability to work with small datasets. Unlike Transformers, which require large amounts of data to generalise without overtraining [80, 50], ESNs exploit the dynamic richness of the reservoir to extract useful information from relatively short sequences. This is a clear advantage in sectors such as batteries, where obtaining comprehensive end-of-life degradation data is expensive and time-consuming [84, 58]. Of course, there are limitations. The reservoir's initial configuration — its size, degree of sparsity, and the so-called spectral radius that governs its stability — is crucial to performance. Poorly chosen parameters can turn the reservoir into a noise generator, negating the architecture's benefits. For this reason, evolutionary optimisation methods have been introduced in recent years to improve performance, such as the aforementioned Grasshopper Optimisation Algorithm [54] or similar strategies adopted in other meta-optimisation contexts [46, 69].

A growing trend in the field is the utilisation of ESNs in hybrid architectures. These modules can be integrated with convolutional modules for local feature extraction, or with LSTM and GRU models to incorporate sequential long-term memory. In consideration of future prospects, these findings position them as optimal candidates for industrial edge computing scenarios, wherein the challenges encompass not only accuracy but also the sustainability of models in terms of latency and resources. A comparison with more compute-intensive approaches, such as attention-based CNNs [90, 3] and Transformer models [71, 49], is particularly instructive. Although ESNs do not always achieve the same accuracy metrics, their energy efficiency and ease of implementation can prove decisive in real-world scenarios.

In conclusion, Echo State Networks can be regarded as a concrete and promising alternative to the unstoppable growth of deep models. As demonstrated in the literature [54, 62, 58], the findings indicate that, while they cannot substitute CNNs or LSTMs in scenarios where maximum accuracy is paramount, they constitute an optimal solution when practical constraints necessitate lightweight, rapid, and adaptable systems that can function with limited data. In a landscape that is increasingly oriented towards edge computing and the Industrial Internet of Things, reservoir computing has the potential to once again assume a leading role, thereby serving as a crucial bridge between cutting-edge academic research and practical field implementation.

### 4.3.5 Transformers

Transformers represent a significant turning point for AI-based prognostics methods. In contradistinction to recurrent networks, which process data sequentially, these models rely on the self-attention mechanism, which facilitates the calculation of interactions between each element of the sequence and all the others, irrespective of temporal distance. Mathematically, the representation of each instant is updated through a weighted combination of information from the entire sequence, with weights learned during training. This eliminates the step-by-step propagation constraint typical of LSTMs and mitigates vanishing gradient problems [80].

The consequence of this is an enhanced ability to capture long-term dependencies, a crucial feature in engineering systems subject to progressive degradation. In the field of aerospace engineering, Ma et al. [80] demonstrated that a Transformer architecture applied to the C-MAPSS dataset outperforms LSTM networks in estimating the RUL of turbofans. This is due to the Transformer's ability to model patterns distributed over large time intervals. In a similar vein, Chen et al. [90] proposed a spatial attention-based convolutional Transformer model for bearings, integrating local convolutions and global attention mechanisms, which are advantageous for processing high-dimensional and noisy signals.

This innovation is not confined to the aeronautics sector. In complex chemical processes, where state variables are numerous and interconnected, Bai and Zhao [50] proposed a multivariate Transformer for multi-step prediction. This model is capable of capturing nonlinear correlations between dozens of variables, and has been shown to outperform convolutional networks and LSTMs. In the field of energy, Transformers have been employed in lithium-ion batteries, with models demonstrating the capacity to estimate State of Health (SoH) and RUL with greater precision than conventional electrochemical methodologies [84, 69]. In the field of wind energy, research conducted by Pandit et al. [62] and Zhang et al. [70] has demonstrated that attention-based architectures possess the capacity to more accurately model high-frequency SCADA signals. These architectures are able to capture temporal and seasonal patterns that prove elusive to more conventional networks.

In recent years, interest has grown in attention variants designed to reduce computational complexity without sacrificing performance. Boos et al. [71] analysed several alternative attention modules, while Chen [49] proposed positional and hierarchical solutions. These attempts address one of the most critical issues: the quadratic cost $O(L^2)$ of attention with respect to the length of the sequence. In real-world scenarios, such as aircraft engines or wind turbines, where signals can contain millions of samples, the computational burden becomes difficult to support for direct industrial use.

Another limitation concerns temporal representation. Self-attention is, by construction, permutation-invariant: it does not possess an intrinsic notion of order. To address this deficit, positional encoding is introduced, but in several cases it has been observed that, without preliminary convolutional modules, the model struggles to effectively capture local dynamics [90, 50].

The third issue is data dependence. To fully exploit their potential, Transformers require large, balanced datasets. In public benchmarks such as C-MAPSS, this is possible, but in more specific domains, the data is often sparse or noisy. Listou Ellefsen et al. [86] showed that, under semi-supervised conditions, the predictive performance of deep networks drops significantly, while Liu et al. [73] highlighted the difficulties in cross-condition generalisation. To mitigate these limitations, the literature proposes transfer learning [91] and domain adaptation [64, 63] strategies, aimed at transferring knowledge from rich domains to poorer contexts.

A further issue that must be addressed is that of interpretability. Transformers are networks characterised by millions of parameters, which pose significant challenges in terms of transparency. In critical applications such as aerospace, accuracy alone is insufficient; it is necessary to understand the underlying reasons behind a predictive decision. Attempts such as that of Alomari and Andó [89], who used SHAP, an Explainable AI (XAI) method that serves to quantify the contribution of each input variable to a complex prediction, to highlight the temporal importance of features, represent steps forward, but the problem of trustworthiness remains unresolved.

In the field of PHM, Alomari and Andó [89] applied SHAP to turbofan models, highlighting how the contribution of features varies over time and changes with operating regime. This means that not only can we know how much a variable matters on average, but we can also observe when it becomes decisive. For example, vibration on a specific channel may be of little importance in the early stages of a component's life, but become increasingly important as degradation progresses.

Finally, the issue of industrial scalability remains to be addressed. The implementation of a Transformer in a laboratory setting or on public datasets differs from its integration within embedded systems, remote turbines, or operational engines. Edge computing imposes lightweight and upgradeability requirements that these architectures, in their original form, do not always satisfy [82, 85].

In conclusion, Transformers have created new opportunities in the field of prognostics, enabling the capture of complex and nonlinear relationships between heterogeneous signals. However, extant literature emphasises that the judicious use of these techniques requires a balance between accuracy and computational cost, between predictive power and interpretive confidence. Consequently, the most robust perspectives do not appear to indicate the exclusive utilisation of Transformers. Instead, there is a tendency towards hybrid architectures, in which CNN modules extract local features, Transformer blocks capture global relationships, autoencoders compress information, and Bayesian filters quantify uncertainty [51, 60]. It is in the coexistence of these approaches, rather than in a single model, that we discern the future trajectory of PHM.

## 4.4 Explainability

In recent years, the use of deep neural networks in PHM has led to a leap in performance in terms of accuracy and predictive capacity. However, this progress has also generated a downside: models have become increasingly complex, opaque, and difficult to interpret. While in the early stages of the discipline's development, RUL prediction was based on transparent physical models, grounded in known equations and laws, the advent of neural architectures has made it difficult to understand the underlying reasons behind a prediction. In critical sectors such as aerospace or energy, accuracy alone is not enough: trust, justification, and accountability of algorithmic decisions are required [75].

This is where the issue of explainability comes in, namely the set of techniques that seek to make the functioning of models otherwise perceived as black boxes more understandable. The goal is not merely academic: a maintenance engineer must be able to justify a decision based on an AI algorithm, a regulatory agency must be able to certify the system's reliability, and an industrial operator must be confident that the predictions are consistent with the physics of component degradation.

Physics-based models, however simplified, always allow for the tracing of a cause-and-effect chain: a temperature change, an anomalous vibration, an accelerated fuel consumption, all elements that can be traced back to known equations. With the advent of purely data-driven models, this connection has weakened. An LSTM or Transformer network, while providing highly accurate estimates of RUL, operates in a space of internal representations that are difficult for a technician to interpret. This opacity breeds mistrust, especially when a predictive decision has economic or safety implications.

A first line of research focuses on feature attribution methods and, more generally, on making the functioning of attention modules visible. Alomari and Andó [89] provide the most obvious example: the use of SHAP to analyse prognostic models on turbofans, producing temporal maps of importance and dependence between variables. The idea is to transform the opaque behaviour of a neural network into a sequence of signals readable by the engineer: which sensors are crucial in the early stages of the life cycle, which become more relevant as degradation accelerates. In parallel, numerous studies exploit attention itself as an explanatory tool. In bearings, Chen et al. [90] show that the fusion of local convolutions and Transformer modules with spatial attention not only improves accuracy, but also allows attention weights to be read as indicators of informative regions in vibrational signals. Similarly, Wang et al. [3] integrate convolutional attention mechanisms to highlight which portions of temporal data contribute most to the estimation of residual life. In the aeronautical field, Zhao and Zou [56] use dual-channel networks with parallel attention, allowing them to distinguish the contribution of short-term components from long-term trends. In all these cases, explainability coincides with the ability to give an engineering sense to the weights of a network.

A second development axis concerns health indices and latent representations. Here, explainability is not entrusted to post-processing, but to the construction of interpretable intermediate variables. De Beaulieu et al. [92] propose a health index calculated in an unsupervised manner using autoencoders and LSTMs, and use it as a synthetic representation of degradation. The added value is not only predictive, but also communicative: the HI becomes a tool readable by a maintenance engineer. Zhang et al. [55] push in the same direction by introducing an encoder with residual shrinkage, capable of generating noise-robust indices that regularly track the component's decay. In batteries, Zhou et al. [84] use multiscale indicators derived from current and voltage signals: these are not mere inputs to the grid, but indicators capable of linking the prediction to the physics of electrochemical ageing. The explanation, in these cases, is obtained by anchoring the learned representation to observable and known phenomena.

A third, increasingly relevant, approach consists in quantifying uncertainty as a form of practical explainability. The logic is simple: a model that provides a predictive point without indicating how confident it is risks being useless in critical contexts. Kim and Liu [27] develop a Bayesian framework capable of producing credible intervals on the RUL; Cáceres et al. [51] introduce Bayesian RNNs that distinguish between epistemic and aleatory uncertainty; Mazaev et al. [59] use Bayesian CNNs to estimate the residual life of solenoid valves, offering not only a prediction but also a confidence bar; Ochella et al. [60] consolidate this perspective by showing the effectiveness of Bayesian networks in quantifying uncertainty in industrial applications. In all these cases, explainability manifests itself not so much as the openness of the model, but as a communication of the confidence that the model itself places in its prediction. It is a form of "algorithmic honesty" that integrates and complements feature attribution techniques.

No less important is the issue of cross-domain robustness, which can be viewed as an indirect form of explainability. If a model works only in a specific context, its interpretation is fragile; if, on the other hand, it maintains predictive stability across different domains, it means it has captured genuine regularities. Schwendemann and Sikora [91] demonstrate that transfer learning, applied to bearings of different types, reduces dependence on data from a single test rig and makes the network's decisions more interpretable. Similarly, Wen et al. [64] and Zhao and Liu [63] develop domain adaptation techniques to estimate the residual life of heterogeneous systems: the goal is not only to improve accuracy, but also to eliminate spurious dependencies that would make predictions unintelligible. This is a different, more subtle idea of explainability: not directly showing "why" a network makes decisions, but ensuring that its reasons are not tied to irrelevant details or artefacts.

In many cases, explainability is achieved by combining different approaches. In turbofans, Asif et al. [45] use deep networks on C-MAPSS and report sensitivity analyses to highlight which variables dominate in different phases of the cycle; Ma et al. [80], in proposing a Transformer for the same dataset, emphasise that attention allows for the highlighting of otherwise invisible long-term correlations. Boujamza and Lissane Elhaq [57] and Boujamza and Elhaq [76] adopt LSTMs with dilated attention modules, allowing for the observation of the relative weight of instants and variables along very long sequences. In rotating mechanical systems, Ensarioğlu et al. [83] present change-point detection mechanisms that identify transitions that can be interpreted as maintenance events. In chemical processes, Bai and Zhao [50] demonstrate that multivariate Transformers can separate the contributions of different state variables, providing greater transparency into the complex interactions between quantities. In the context of batteries, Dhananjay Rao et al. [46] and Krishna et al. [53] have developed predictive architectures that can also explain degradation in terms consistent with electrochemical phenomena. In the context of wind power, Verma et al. [93] and Pandit et al. [62] utilise SCADA data to identify the sensors and environmental conditions that are critical for gearbox failures. Alongside these positive practices, however, limitations emerge. Solís-Martín et al. [75] emphasise the problem of the soundness of explanations: not all attention maps or SHAP estimates truly reflect the internal workings of the model, and they are not always stable under small data perturbations. In a critical field like aerospace, this risk is unacceptable. This opens a new research avenue aimed not only at producing explanations, but also at verifying their validity, for example through sanity checks, weight randomisation, and controlled sensitivity studies. This is an important conceptual shift: explainability is no longer an optional extra, but a design requirement, subject to quality standards equal to accuracy metrics.

Looking ahead, it seems that the literature is converging towards a hybrid model comprising explanations based on feature attribution and attention, enriched with interpretable indicators (health index), supported by uncertainty measures and constrained by physical rules. It is in this combination that the most promising trajectory is seen. The work of Berghout et al. [77] on ProgNet, which integrates physical curves into neural models, and that of Liu et al. [68] on digital twins and domain adaptation, demonstrate that explainability can be incorporated into the model structure as well as being applied post-hoc. Similarly, using denoising autoencoders [44] or variational autoencoders [88] enables the creation of compact representations in which distance from the healthy manifold becomes a readable anomaly index.

Ultimately, analysis of the fifty-four articles clearly demonstrates that explainability is necessary for turning accurate models into reliable tools. This can involve using SHAPs to highlight temporal contributions [89]; attention maps in bearings [90, 3]; robust health indices [92, 55]; uncertainty quantification [27, 51, 60]; or integrated physical constraints [77, 68]. The direction is clear: to build predictive models that are powerful, understandable, and justifiable.

**Table 4.1:** Explainability approaches in PHM literature

| Approach | Example paper | Contribution to Explainability |
|---|---|---|
| Feature attribution (SHAP) | Alomari & Andó (2024) [89] | Temporal importance maps for turbofan prognostics, showing variable relevance across life cycle stages. |
| Attention mechanisms | Chen et al. (2022) [90]; Wang et al. (2023) [3] | Attention weights interpreted as indicators of informative regions in vibration/time-series signals. |
| Health indices (HI) | De Beaulieu et al. (2022) [92]; Zhang et al. (2024) [55] | Construction of interpretable degradation curves and robust synthetic health indices. |
| Uncertainty quantification | Kim & Liu (2020) [27]; Cáceres et al. (2021) [51]; Ochella et al. (2024) [60] | Bayesian methods providing confidence intervals and distinction between epistemic and aleatory uncertainty. |
| Domain adaptation | Wen et al. (2021) [64]; Zhao & Liu (2022) [63] | Increased robustness across domains, reducing reliance on context-specific artefacts. |
| Physics-informed models | Berghout et al. (2023) [77]; Liu et al. (2020) [68] | Integration of physical laws and digital twins to constrain models and improve interpretability. |

## 4.5 Metrics for Prognostics Models Evaluation

In any field of applied science, measurement is the process by which intuitive ideas are distinguished from empirical evidence. In Prognostics and Health Management (PHM), metrics are not a marginal detail but fundamental instruments: they establish the usefulness of a model, enable comparison across approaches, and define the distance between algorithmic outputs and the decisions of maintenance engineers. The discourse surrounding metrics is inevitably linked to the notion of credibility.

As Saxena et al. [28] observed, "prognostic technologies lack standard definitions, and therefore performance evaluation suffers from ambiguity and inconsistency." This lack of consensus has complicated the comparison of prognostic methods for years, producing results that are difficult to interpret on a larger scale.

In this context, the selection of metrics assumes a strategic dimension, as it not only determines the evaluation of a model but also defines the nature of the variables being measured. Consequently, this choice influences both optimisation processes and operational outcomes.

Metrics in PHM are stratified across multiple levels depending on the type of evaluation. Some studies focus on the quantification of numerical error, others on the timeliness of prediction, while more recent works measure the probabilistic reliability of models that output distributions rather than point estimates. In addition, cost-sensitive approaches highlight pragmatic considerations, such as the financial implications of false alarms or the operational impact of overestimating the RUL. The intricacies of this subject are best clarified by examining these categories in detail.

## Pointwise Error Metrics

The most widely used category comprises pointwise error metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$$

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

These measures dominate in benchmark studies, such as those based on the C-MAPSS dataset for turbofan engines [45, 83]. Their main advantage is simplicity and comparability, but they remain blind to aspects such as uncertainty, variance, and decision-making implications [27]. In bearings and gearboxes, for instance, RMSE and MAE are often combined with temporal performance metrics to better capture degradation dynamics [90, 93].

## Time-Dependent Metrics

PHM has a distinctive temporal nature: the value of a prediction depends on *when* it becomes accurate. For this reason, metrics such as Prognostic Horizon (PH), Relative Accuracy (RA), and Cumulative Relative Accuracy (CRA) were introduced [94]. PH, for example, defines the earliest point at which predictions remain within an acceptable error bound until failure. These indices are particularly relevant in contexts such as wind energy, where early and stable predictions can yield significant operational savings [62, 70]. Other works confirm that temporal stability is as critical as raw accuracy, especially when planning maintenance across distributed assets [82, 85].

## Probabilistic Metrics and Uncertainty Quantification

Recent studies have emphasised probabilistic assessment, in which predictions are expressed as distributions rather than single values. One prominent metric is the Continuous Ranked Probability Score (CRPS):

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} \left[F(z) - \mathbf{1}\{y \leq z\}\right]^2 dz$$

where $F(z)$ is the predictive cumulative distribution and $y$ the observed value.

Another widely used index is the Prediction Interval Coverage Probability (PICP):

$$\mathrm{PICP} = \frac{1}{N}\sum_{i=1}^{N} c_i, \quad c_i = \begin{cases} 1, & \text{if } y_i \in [L_i, U_i] \\ 0, & \text{otherwise} \end{cases}$$

These metrics, applied in turbofan and battery studies [51, 60, 84], highlight that reliability depends not only on predictive accuracy but also on the confidence intervals surrounding estimates. This is crucial in aerospace and energy applications, where decision-makers need to know both the expected RUL and the likelihood of failure within a given time frame.

## Asymmetric Scoring Functions

In safety-critical systems, the costs of early and late predictions are not equivalent. To capture this asymmetry, the NASA scoring function is widely adopted:

$$S = \sum_{i=1}^{N} \begin{cases} e^{-\frac{\hat{y}_i - y_i}{\alpha}} - 1, & \text{if } \hat{y}_i \geq y_i \\ e^{\frac{\hat{y}_i - y_i}{\beta}} - 1, & \text{if } \hat{y}_i < y_i \end{cases}$$

where $\alpha$ and $\beta$ regulate the severity of penalties. This function, extensively used in NASA benchmarks and in subsequent aerospace studies [89], strongly penalises late predictions, reflecting the low tolerance for unexpected failures in real-world scenarios.

## Cost-Oriented and Operational Metrics

Beyond accuracy, modern PHM requires cost-sensitive evaluation. Metrics such as expected maintenance cost, avoided downtime, and false alarm rate are increasingly reported, particularly in Industry 4.0 and IoT-enabled maintenance contexts [53, 72]. Wu et al. [85] also stress that in edge-computing environments, computational overhead must be included in the evaluation, making lightweight models preferable even when slightly less accurate.

## Composite and Recent Metrics: $\alpha$–$\lambda$, NMSE, and XAI-Linked Scores

While pointwise and probabilistic scores cover fundamental aspects, recent PHM literature increasingly adopts composite metrics that better reflect operational needs and governance requirements[89, 75]. Three families stand out:

- the $\alpha$–$\lambda$ performance metric for time-to-failure utility
- the Normalized Mean Square Error (NMSE) for scale-robust comparison
- explainability-linked metrics (fidelity and stability) to assess the consistency and robustness of interpretability tools

The $\alpha$–$\lambda$ metric formalises the intuition that a RUL predictor is only useful if it becomes accurate early enough and stays accurate until end-of-life. Let $e(t) = \hat{r}(t) - r(t)$ denote the RUL error at time $t$ for a run-to-failure trajectory, with $r(t)$ the true remaining life. Given a relative error bound $\alpha \in (0,1)$ and a persistence window $\lambda > 0$, define the earliest time $t^\star$ such that

$$|e(\tau)| \leq \alpha\, r(\tau) \quad \text{for all } \tau \in [t^\star,\, t_f],$$

where $t_f$ is the failure time. The prognostic horizon is $PH_{\alpha,\lambda} = t^\star - (t_f - \lambda)$, and a prediction is considered operationally valid if $PH_{\alpha,\lambda} \geq 0$. In practice one reports the proportion of runs meeting the criterion for selected $(\alpha, \lambda)$ pairs, or aggregates $PH_{\alpha,\lambda}$ across runs [94]. This metric complements RMSE/MAE by rewarding models that stabilise early and remain within a tolerable cone as the asset approaches failure; it is particularly relevant in aerospace and wind applications where late accuracy is of limited value [62, 70].

To compare performance across units, operating conditions, or datasets with different scales and variances, the NMSE normalises the MSE by a scale of the target. Two common definitions are:

$$\mathrm{NMSE}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\mathrm{Var}(y)}, \qquad \mathrm{NMSE}_2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}.$$

Both yield a dimensionless quantity; $\mathrm{NMSE}_2 = 1 - R^2$ for linear least-squares. NMSE is increasingly reported in comparative studies on bearings and wind SCADA where cross-dataset benchmarking is central [87, 62, 3], and in deep architectures for RUL under multiple regimes [80, 90]. Its advantage over RMSE/MAE is scale invariance; its limitation is the dependence on the chosen normaliser (variance vs. total sum of squares), which must be stated explicitly to ensure comparability.

Finally, as explainability becomes a prerequisite in safety-critical PHM, the quality of explanations must itself be evaluated. Two practical, model-agnostic notions are widely used:

- Fidelity measures how well an explanation (e.g., SHAP attributions, attention weights, or a local surrogate) preserves the predictive behaviour of the original model in the neighbourhood of an instance. A simple fidelity score is the local $R^2$ or correlation between the model output under perturbations and a constrained surrogate fitted in the same neighbourhood; in turbofan settings, SHAP-based analyses stress the need for faithful, additive decompositions aligned with the model output [89].

- Stability assesses whether explanations are robust to small, label-preserving perturbations or to re-sampling. Low stability flags brittle explanations. This concern is central in soundness analyses of XAI for PHM [75], and is particularly pertinent for attention maps in CNN/Transformer pipelines [90, 71, 3].

The review of the 54 analysed articles reveals three macrotrends. First, pointwise error metrics remain ubiquitous but are progressively integrated with temporal and probabilistic indices. Second, standardisation is lacking: PH, CRPS, and PICP are applied inconsistently, limiting comparability across studies [86, 73]. Third, a persistent gap remains between metrics and actual operational value. A model with lower RMSE is not necessarily the one that produces more useful predictions for maintenance decision-making.

In addition, the growing adoption of explainability and trust related indices [89, 75] marks a shift: metrics are no longer confined to performance evaluation but also function as instruments of credibility. This reflects the evolution of PHM from a purely predictive science to a decision-support discipline.

Metrics are no longer secondary tools. In support of this, the literature indicates a clear transition: from simple error indices towards uncertainty-aware, time-sensitive, cost-oriented, and explainability-driven measures. Yet a lack of harmonisation persists, and no universally accepted benchmark framework exists. Future research must integrate predictive accuracy with operational impact to ensure that metrics reflect the realities of industrial deployment.

## 4.6 Uncertainty

One of the fundamental issues to address when discussing prognostics is uncertainty. While in the previous chapter, metrics provided the toolbox for evaluating models and performance, here the focus shifts to what these metrics seek to measure in depth: a model's ability to state not only the "how much is left" before failure, but also the level of doubt that accompanies this statement. Uncertainty thus becomes a structural dimension of prognostics, an indicator of epistemic as well as operational reliability.

Recent literature demonstrates that deterministic models, irrespective of their sophistication, are no longer adequate. In their 2020 publication, Kim and Liu [27] presented a Bayesian deep learning framework for interval estimation of Remaining Useful Life (RUL). The authors emphasised that point-by-point prediction can induce a false sense of security [27]. The architecture of these systems provides predictive distributions, thereby paving the way for decisions based not only on the expected value but also on the probability that the component will exceed certain thresholds. In a similar vein, Mazaev et al. [59] demonstrated that the implementation of Bayesian CNNs for solenoid valves engenders estimates of RUL that, when accompanied by uncertainty bands, exhibit enhanced robustness to noise and unobserved conditions. This aspect is also evident in turbofans, where Ochella et al. [60] demonstrated that the distinction between epistemic and aleatory uncertainty becomes essential: the former is reduced with more data and better models, while the latter remains unavoidable and must be managed as an intrinsic characteristic of the system.

It is therefore not surprising that in reference benchmarks, model evaluation is no longer based solely on point metrics (RMSE, MAE) but also on probabilistic indicators such as the Continuous Ranked Probability Score (CRPS) and the Prediction Interval Coverage Probability (PICP). Ensarioğlu et al. [83] have shown, for example, that in turbofans, the combination of deep networks and change point detection techniques improves the coverage of the prediction ranges, as measured by PICP, more than it reduces the mean error. Similarly, Zhou et al. [84], working on lithium-ion batteries, have shown that CRPS offers a more sensitive assessment of the predictive distribution, rewarding models that not only center the mean but also accurately describe the variability of electrochemical degradation. This is a clear shift: the focus is no longer simply on "how close" the predicted value is, but on "how well-designed" its probability distribution is.

A recurring theme in the most cited works concerns data sparsity and variability. Listou Ellefsen et al. [86] have observed that under semi-supervised conditions, deep models tend to lose not only accuracy but also consistency in quantifying uncertainty. In other words, the predictive range becomes uninformative precisely when it is most needed. To address this limitation, several studies have explored transfer learning and domain adaptation strategies. Schwendemann and Sikora [91], for example, transferred knowledge between different bearing types, reducing not only the mean error but also the predictive spread. Similarly, Wen et al. [64] and Zhao and Liu [63] showed how domain adaptation techniques allow for better uncertainty control when a model trained on a data-rich domain is applied to real-world scenarios with few observations.

The link between uncertainty and explainability is perhaps the most relevant development in recent years. Alomari and Andó [89], using SHAP to analyze turbofans, showed that the stability of importance maps is closely linked to the epistemic variance of the model: inconsistent explanations often coincide with high levels of uncertainty. On a more theoretical level, Solís-Martín et al. [75] have warned about the risk of "fragile" explanations, proposing the concept of soundness as a criterion for assessing the robustness of explanations themselves. From this perspective, new metrics such as the fidelity and stability of attention maps or SHAP attributions become tools for measuring uncertainty, broadening the perspective that traditionally limited it to prediction intervals.

The application implications are transversal. For instance, in the context of turbofans, Transformer models, such as the one proposed by Ma et al. [80], have been demonstrated to generate more consistent predictive distributions over extended sequences in comparison to classical LSTMs, thereby enhancing forecast stability. In the field of bearings, Chen et al. [90] pioneered a novel approach by integrating spatial attention and convolutions, thereby demonstrating that noise reduction leads to not only reduced average errors but also more precise and informative uncertainty ranges. In the wind sector, Pandit et al. [62] and Zhang et al. [70] highlighted that the use of hybrid architectures allows for forecast stabilisation even under extreme weather conditions, reducing false alarms. For batteries, studies such as those by Safavi et al. [58] and Dhananjay Rao et al. [46] showed that probabilistic approaches can distinguish between physiological degradation and anomalous phenomena precisely through a careful use of uncertainty.

The impression that emerges from the analysis of all fifty-four articles is that uncertainty has become the true language through which models communicate the degree of reliability of their predictions. While traditional metrics (RMSE, MAE, NASA score) provide useful numbers for comparing performance, quantifying uncertainty adds an epistemic and operational layer without which prognostics would risk remaining confined to academia. Today, however, the focus is on models that do not pretend to be infallible, but rather declare their own margin of doubt, thus offering decision makers not just a number, but a probability map within which to plan costs, safety, and operational continuity. In this sense, uncertainty is no longer a flaw to be minimized, but rather an attribute to be cultivated and made explicit, an integral part of the trust we can place in artificial intelligence systems for predictive maintenance.

## 4.7 Trustability

If AI-based prognostics wants to break free from the reassuring confines of public datasets and research laboratories, it must address a more subtle and complex obstacle than mean error or predictive variance: trustworthiness. The notion of trustability in a model encompasses not only the question of its functionality, but also the extent to which its users, such as an engineer can place their trust in its predictions. This concept encompasses accuracy, uncertainty, explainability and robustness, and is situated at the intersection of technology and human perception. It is important to note that a model may exhibit excellent RMSE, calibrated prediction intervals, and even plausible explanations; however, if it fails to inspire trust, it remains confined to publication.

The literature reviewed clearly shows that trust in PHM models has become a research field of its own. Alomari and Andó [89], for example, tackled the issue by integrating SHAP with prognostic networks for turbofans, showing that transparency into the temporal trends of features can increase the model's credibility in the eyes of practitioners. Solís-Martín et al. [75] openly discussed the "soundness" of XAI, introducing the problem that simply providing an attribution is not enough: it must be ensured that the attribution is stable and consistent. In other words, trustability is not synonymous with explainability, but includes it along with other dimensions.

The first component is robustness. Reliable models must be able to operate not only on clean data, but also in the presence of noise and under variable operating conditions and in situations that were not observed during training. Berghout and Benbouzid [77] illustrated in their study of high-speed aircraft bearings that collaborative feature selection can enhance the reliability of predictions in the presence of noisy signals. Similarly, Soualhi et al. [82] showed that adapting to different operating conditions in turbofans requires resilience to sudden changes as well as accuracy, which directly impacts operator confidence. In the field of wind energy, Verma et al. [93] and Pandit et al. [62] highlighted that models able to maintain stable performance in extreme weather conditions are considered more reliable, regardless of average error.

A second pillar of trustability is the quantification of uncertainty. Previous chapters have demonstrated how metrics such as CRPS and PICP [51, 60] enable us to assess how accurately a model represents variability. The key here is that uncertainty becomes an act of honesty: declaring how much we "do not know" is paradoxically what makes a model more trustworthy. Not surprisingly, in critical applications such as lithium-ion batteries, Zhou et al. [84] and Safavi et al. [58] underlined the need for models that provide reliable ranges, because an imprecise estimate accompanied by a calibrated interval is often more useful than an unjustified point prediction.

trustability is also linked to a model's ability to generalise beyond the domain in which it was trained. This is where transfer learning and domain adaptation strategies become essential. Schwendemann and Sikora [91] showed that predictive models can be transferred between different bearing types while maintaining robust performance and consistent confidence intervals. Similarly, Wen et al. [64] and Zhao and Liu [63] demonstrated the transferability of models across platforms and operating conditions using adaptation approaches, thereby reducing the risk of catastrophic errors resulting from misplaced overconfidence. Increasingly, the literature emphasises that trustability is not a static attribute of the model, but rather a property that builds over time through generalisation tests, stress tests and validation in real domains.

Another key element is the understandability of decisions. If a model predicts that an engine will fail after 25 cycles, confidence in this prediction increases if it is possible to identify the variables that drove this conclusion, and if these variables are consistent with engineering knowledge. Alomari and Andó [89] showed that SHAP-based explanations applied to turbofans can generate temporal importance maps that align with vibration and temperature signals recognised by experts. This type of consistency fosters trust by creating a bridge between the algorithm and the engineer. Conversely, when explanations are based on obscure or implausible features, the model's perceived reliability plummets, even if the error metrics are strong.

trustability is also intertwined with the issues of costs and implementability. As emphasised by Wu et al. [85], in edge computing and distributed maintenance scenarios, trust concerns not only the accuracy of predictions but also the computational sustainability of the model. An algorithm that demands excessive computing power or long response times may be perceived as unusable, irrespective of its precision. Jaenal et al. [72] highlighted that trust in machine learning systems can be enhanced by ensuring a balance between accuracy and implementability. This is achieved through the integration of models into real-world contexts without technical friction.

Finally, a consideration of social and organisational perception is warranted.

The requirement for a model to be accurate, uncertain in a calibrated way, and explainable is insufficient in itself; it must also be accepted by those who use it. This point is illustrated by studies on batteries and aeronautical systems, where the adoption of PHM models is often contingent on the ability to transparently communicate their limitations and advantages [53, 89]. trustability, therefore, is not merely a technical attribute, but also a matter of communication between those who develop models and those who must make operational decisions.

A review of the literature shows that trustability has become the true currency with which prognostics gains acceptance in industry. Traditional metrics measure accuracy, uncertainty quantifies the margins of doubt, and explainability offers a window into the "black box." Trustability, on the other hand, is the synthesis, that is the ability of a model to be trusted, accepted, and used. It is a concept that, far from being purely technical, touches on the epistemic and cultural dimensions of predictive maintenance. Ultimately, a reliable model is not one that makes the fewest mistakes, but one that can build a relationship of trust with the engineer who must decide whether to stop an engine, replace a bearing, or postpone a repair. And it is this trust, even more than mathematical accuracy, that determines whether AI-based prognostics will remain confined to the pages of scientific journals or become an everyday engineering tool.

## 4.8 Datasets in AI-Based Prognostics

The topic of datasets represents one of the most delicate yet, at the same time, most overlooked issues in the debate on AI-based prognostics. A model, no matter how sophisticated, cannot ignore its fuelling factor: data. Data is never neutral, but reflects acquisition conditions, sensor limitations, the availability of failure scenarios, and industry's willingness or otherwise to share it. Recent literature clearly confirms that the availability of adequate datasets is not a technical detail, but rather a factor that influences the entire development process of predictive systems [45, 44].

A first element to consider concerns the type of dataset. In the aerospace sector, the almost obligatory reference is the renowned C-MAPSS (Commercial Modular Aero-Propulsion System Simulation), a tool developed by NASA for the simulation of realistic large commercial turbofan engine data, which constitutes the preferred testbed for most studies on turbofan engines. Asif et al. [45] demonstrate in their work that deep neural networks trained on C-MAPSS can outperform traditional physics models when it comes to estimating RUL. However, several authors point out that, although valuable, this dataset has significant limitations, including simplified scenarios, simulated flight conditions and reduced noise levels compared to the real operating environment [83, 82]. This paradox highlights a constant tension between the need for standardisation, which C-MAPSS guarantees, and adherence to real-world complexity, which it cannot fully capture.

A similar scenario can be observed in the context of rotating bearings, where datasets such as PRONOSTIA or XJTU-SY have emerged as de facto standards, thereby establishing themselves as the accepted benchmark for research and analysis in this field. Here, studies such as those by Liu et al. [25, 73] and De Beaulieu et al. [92] underscore a critical aspect: the duration of the sequences is frequently brief, yet the operating conditions replicated in the laboratory exhibit substantial divergence from their industrial counterparts. Consequently, a model that has been trained on these datasets may exhibit excellent performance in a controlled environment, but rapidly degrade in the field. The emergence of unsupervised approaches based on latent health indices, as exemplified by the methodology proposed by De Beaulieu et al. [92], can be attributed to the necessity to extract robust information from data that is characterised by noise and incompleteness.

The scenario of wind turbines is even more complex, where the availability of data is abundant, also due to SCADA systems, but problematic from a qualitative point of view. Desai et al. [78] and Verma et al. [93] show that high-frequency signals are characterised by strong noise and a very low incidence of real failures, which makes it difficult to train deep networks without resorting to balancing or simulation strategies. Subsequent studies, such as those by Pandit et al. [62] and Zhang et al. [70], demonstrated that attention-based architectures are able to partially handle the heterogeneity of SCADA data, but even in this case the problem remains structural: abundance does not coincide with quality.

Lithium-ion batteries offer another interesting insight. Here, the availability of public datasets, such as those from NASA, is less than in aerospace, and most studies rely on experimental data collected in the laboratory. Mansouri et al. [79] and Zhou et al. [84] highlighted the high variability of charge–discharge cycles, while Zhang et al. [26], using impedance spectroscopy, demonstrated how laboratory datasets can offer high-quality signals but are difficult to transfer to real-world scenarios. The main challenge, as noted by Krishna et al. [53], is therefore representativeness: a model that operates under controlled conditions struggles to replicate the same performance in complex operating environments, subject to variable temperatures and unpredictable stress conditions.

The issue of data quality, however, is not limited to sparsity or noise, since there is also the structural problem of imbalance: in real datasets, data relating to normal conditions significantly outweigh fault data. Listou Ellefsen et al. [86] showed how, in semi-supervised contexts, the scarcity of fault labels compromises the generalisation ability of deep networks. Precisely for this reason, interest has begun to spread in data augmentation techniques, such as the GAN-based approach proposed by Baptista and Henriques [44], capable of generating synthetic signals consistent with the underlying physics.

A number of strategies have been proposed in the extant literature as a means of mitigating the limitations of datasets. These include domain adaptation and transfer learning approaches, which are emerging as essential tools for transferring models from data-rich to data-poor domains. As posited by Wen et al. [64] and Zhao and Liu [63], models have been proposed that are capable of learning representations that remain invariant under operating conditions. In contrast, Schwendemann and Sikora [91] have experimented with transfer learning techniques to generalise from one bearing type to another. In parallel, physics-informed models [77, 60] offer the possibility of integrating synthetic data generated by physics simulations with observed data, thus increasing the variety and robustness of available datasets. A more recent, but rapidly growing, trajectory concerns federated learning, discussed for industrial PHM pipelines by Jaenal et al. [72], which allows models to be trained on distributed datasets without centralising data, preserving industrial privacy and expanding the information base.

The common thread that emerges is that dataset quality is not an ancillary aspect, but a truly critical design variable. Seemingly brilliant performance metrics can prove misleading if the reference dataset is overly simplified or unrepresentative. What emerges is that data quality analysis is inseparable from metrics analysis (see the Metrics chapter) and, in turn, affects the very ability to ensure explainability and trustability in predictive models.

In conclusion, the datasets available today have been a fundamental catalyst for the evolution of AI-based prognostics, but their partial and often artificial nature requires caution. The literature converges on one point: without rich, heterogeneous, and high-quality data, even the most refined algorithm risks remaining confined to the academic realm. Future prospects seem to focus on a combination of techniques ranging from augmentation to transfer learning, from physical simulations to federated learning, to overcome the chronic scarcity of real-world failure data and bring prognostic models closer to the most challenging yet most relevant testbed: that of the operational industrial world.

## 4.9 Implementability and Computational Cost

When moving from academic prototypes to operational implementation, AI-based prognostics encounters a less fascinating but much more concrete challenge than theory: computational cost. Accurately predicting the Remaining Useful Life is not enough, nor is providing an elegant explanation or a well-calibrated uncertainty range. In industrial settings, AI must be lightweight, fast, stable, and integrable into systems that often have limited resources. It is in this transition from research to field deployment that many sophisticated architectures lose ground.

The reviewed literature highlights a recurring phenomenon. While many models that emerge as “state of the art” on datasets such as C-MAPSS or PRONOSTIA work well in theory, they often become impractical when implemented on edge devices, embedded systems, or distributed SCADA platforms. Soualhi et al. [82] observed, in turbofan applications, that training deep architectures was not only computationally expensive, but also inflexible when updating the model with new data. The dynamic nature of industrial systems requires algorithms that can be retrained or adapted without disrupting the infrastructure.

The situation is even more evident with Transformers, currently a major focus of the literature but often at odds with real-world hardware limitations. Boos et al. [71] show that the self-attention operator has a quadratic cost with respect to sequence length. On aeronautical or wind-turbine data, where signals can exceed millions of samples, this implies hundreds of megabytes of memory and inference times incompatible with online maintenance. To mitigate the problem, their study proposes lighter attention modules that reduce the number of operations while maintaining performance close to standard Transformers. Chen [90] introduces a variant with positional and hierarchical attention, where signals are first compressed by convolutional layers, reducing dimensionality before global analysis: an effective trade-off between accuracy and computational sustainability.

Another recurring strategy in the literature is reducing reliance on full retraining. Ensarioğlu et al. [83] combine deep networks with change-point detection so that the model is updated only when significant changes in system behaviour emerge. This avoids continuous reprocessing of the entire dataset; the model intervenes only when necessary, saving resources and accelerating deployment. In wind energy, Pandit et al. [62] demonstrate that lightweight models with incremental updates enable execution on SCADA platforms without transferring large data volumes to central servers.

Edge computing is a constant topic. Wu et al. [85] developed the CALAP fusion model specifically to perform inference and real-time updates on low-power devices and avoid cloud latency. Jaenal et al. [72] took a similar approach with MachNet, a unified architecture for industrial predictive maintenance designed with computational limitations and ease of deployment in mind. The result is reduced inference time and improved compatibility with legacy infrastructures—an aspect often overlooked in purely theoretical literature.

Some works attack the problem from the opposite side: reducing the data rather than the model. Baptista and Henriques [44] demonstrate how GAN-based denoising can generate clean synthetic degradation signals prior to analysis, resulting in more compact features and faster training. De Beaulieu et al. [92] exploit autoencoders to build a long-term health index that compresses thousands of vibration samples into a compact vector; less data also implies faster inference. In turbofans, Berghout et al. [77] introduce ProgNet, a hybrid model incorporating physical knowledge to reduce the required number of parameters and improve training and execution efficiency.

Regarding batteries, Safavi et al. [58] and Dhananjay Rao et al. [46] highlight that, while offline training can be expensive, inference becomes rapid if the model is designed with compressed layers or pruning. Using optimised LSTMs or reduced-attention variants enables response times compatible with IoT and battery management systems, where available power is minimal.

Computational cost is closely linked to industrial scalability. Verma et al. [93] and Zhang et al. [70] show that, for wind turbines, the challenge is not only model accuracy, but the ability to analyse terabytes of SCADA data without collapsing the infrastructure. Hybrid solutions-convolution to filter noise, attention for global correlation-have become common practice. A similar idea emerges from Liu et al. [73], who propose a mixture-of-encoders architecture: instead of a single massive network, multiple specialised encoders handle different regimes, reducing computational load and the risk of overfitting.

In this context, it is clear that implementability and computational cost are not only a matter of algorithm choice, but also of digital infrastructure. Federated learning, discussed by Jaenal et al. [72], is emerging as a natural response: models learn from multiple sites without centralising data, reducing transfer costs and privacy concerns. Bayesian models, on the other hand, are more expensive to train [51, 60], but they offer uncertainty quantification that can drastically reduce the cost of incorrect or premature maintenance. A slower but more reliable model may be preferable in safety-critical sectors such as aerospace or rail.

The emerging picture is that accuracy is no longer the only important factor. Recent papers converge on concepts such as deployability, real-time feasibility, and computational sustainability. Research is moving toward hybrid and adaptive architectures capable of balancing predictive power with operational lightness. If the first wave of AI-based prognostics models was built to show that "it was possible," the current wave aims to show that "it is actually usable."

Industrial systems do not require perfect models; they require reliable, stable, and lightweight models. Prognostics should not exist in the laboratory: it must function in the field, within the limits of network, memory, and real time. The direction indicated by the literature is clear: fewer monolithic architectures, more modular pipelines; less centralised computing, more edge; less complexity for its own sake, more engineering robustness. If the first revolution in prognostics was mathematical, the second will be architectural.

## 4.10 Software Platforms

In the context of AI-based prognostics, the focus is predominantly on models, signals, and algorithms. The role of the software platforms on which these models are developed and, more importantly, on which they are intended to be deployed once they leave the laboratory is less obvious. However, the selection of the platform has a significant impact on the reproducibility, transferability, upgradeability, and compatibility of the model with existing infrastructure. In summary, the platform serves as the conduit between the realm of modelling and the tangible world.

A comprehensive review of the existing literature provides a fairly clear picture. Python dominates the academic scene, particularly due to the machine learning ecosystem centred on TensorFlow, PyTorch, Keras and Scikit-learn. Asif et al. [45], working with C-MAPSS to predict the RUL of turbofans, use TensorFlow/Keras to implement deep networks and compare them with physical models. Berghout et al. [77] build ProgNet in PyTorch, leveraging the flexibility of hybrid models and the ability to easily integrate physical and data-driven layers. Chen et al. [90] and Bai and Zhao [50] also implement transformers and multivariable convolutional networks with Python frameworks, demonstrating that code portability and the availability of specialized libraries make Python a natural environment for experimenting with new models.

The situation changes, however, when looking at industrial contexts. While Python is ideal for research, prototyping, and experimentation, the transition to production is not always straightforward. Some studies show that directly integrating Python models into SCADA environments or embedded systems is non-trivial. Verma et al. [93], in the wind energy sector, emphasize the need to migrate part of the model to lighter platforms to avoid latency and compatibility issues. For this reason, research still relies on MATLAB in several cases, especially when combining control algorithms, real-world signals, and rapid prototyping. Zhang et al. [26], for example, develop models for lithium-ion batteries based on impedance spectroscopy using MATLAB, leveraging its robust signal processing and engineering prototyping. Despite its perceived conservatism, the aerospace industry continues to utilise MATLAB, acknowledging its role in ensuring certification, traceability, and pipeline stability.

Alongside Python and MATLAB, the literature identifies a third, less prominent but significant area: proprietary platforms. Wu et al. [85], with the CALAP model, highlight the use of custom frameworks to perform online inference and reduce cloud dependence. Jaenal et al. [72] also propose MachNet as an architecture designed for direct integration into industrial systems, without having to use complex external libraries. This is indicative of a clear signal: not all models are intended to exist in open-source environments, with many companies preferring proprietary pipelines for reasons pertaining to privacy, security, or certification.

The issue of platforms also relates to hardware resources. A Transformer may be elegant on paper, but if it requires dedicated GPUs or remote servers to operate, its industrial adoption becomes unlikely. Boos et al. [71] clearly show that the self-attention operator has a quadratic cost with respect to the sequence length. Chen [90] introduces a variant with positional and hierarchical attention, in which signals are first compressed by convolutional layers, reducing dimensionality before global analysis: an effective tradeoff between accuracy and computational sustainability.

Some works take a different approach to the problem, focusing on reducing the data rather than the model. Baptista and Henriques [44] demonstrate how GANs can be utilised to generate degradation signals and 'clean' noisy data prior to analysis. This results in more compact features and faster models for training. De Beaulieu et al. [92] exploit autoencoders to build a long-term health index that synthesises thousands of vibration samples into a compact vector; less data also means faster inference.

Edge computing is a constant topic of discussion. Wu et al. [85] developed the CALAP fusion model for this very purpose: to perform inference and real-time updates on low-power devices and avoid cloud latency. Jaenal et al. [72] took a similar approach with MachNet, a unified architecture for industrial predictive maintenance that was designed with computational limitations, model transfer, and ease of implementation in mind. The result is reduced inference times and improved compatibility with legacy infrastructures.

Some works address the challenge of retraining. Ensarioğlu et al. [83] combine deep networks with change-point detection to update the model only when significant shifts in behaviour appear. In wind energy, Pandit et al. [62] also demonstrate how lightweight models and incremental updates enable execution on SCADA platforms without transferring large data volumes.

Regarding batteries, Safavi et al. [58] and Dhananjay Rao et al. [46] highlight that, while offline training is costly, inference can be rapid if the model is designed with compressed layers or pruning. Furthermore, using optimised LSTMs or lighter versions with reduced attention enables response times compatible with IoT and Battery Management Systems (BMS), where available power is minimal. crps Computational cost is closely linked to industrial scalability. Verma et al. [93] and Zhang et al. [70] show that, in wind turbines, the challenge is not only model accuracy, but the ability to analyse terabytes of SCADA data without collapsing the infrastructure. Adopting hybrid models, such as convolution to filter noise and attention for global correlation, is a popular solution: it reduces complexity without sacrificing predictive power.

Federated learning, discussed by Jaenal et al. [72], is emerging as a natural response to the need to distribute the load: models learn from multiple sites without centralizing data, reducing transfer costs and privacy concerns.

The emerging picture is that accuracy is no longer the only important factor. Recent papers recur on concepts such as deployability, real-time feasibility, and computational sustainability. Research is converging toward hybrid and adaptive architectures, capable of balancing predictive power with operational lightness.

In short, industrial systems do not require perfect models; they require reliable, stable, and lightweight models. Prognostics must function in the field, within the limits of network, memory, and real-time. The direction indicated by the literature is unequivocal: fewer monolithic algorithms, more modular solutions; less centralized processing, more edge; less complexity for its own sake, more engineering robustness.

**Table 4.2:** Comparison of software platforms for AI-based prognostics

| **Platform** | **Advantages** | **Limitations** |
|---|---|---|
| Python | Open-source and flexible ecosystem with rich machine and deep learning libraries (TensorFlow, PyTorch, Keras); ideal for prototyping and research; strong community support and cross-platform compatibility. | Not always suitable for real-time or embedded applications; deployment on industrial systems often requires conversion or optimisation (e.g. ONNX, C++). |
| MATLAB | Comprehensive signal processing and control toolboxes; certified and stable environment widely used in aerospace and manufacturing; strong integration with data acquisition and hardware. | Closed-source and costly licensing; limited support for large-scale deep learning training; less adaptable for distributed or edge environments. |
| Proprietary / Custom Platforms | High compatibility with industrial assets; real-time execution; enhanced cybersecurity and intellectual property protection; tailored integration with existing infrastructures | Low transparency; poor reproducibility of academic results; high development and maintenance effort. |
| Cloud and Distributed Frameworks | High scalability and computational power; suitable for big data processing and continuous monitoring;simplifies model updates and version control. | High latency and bandwidth consumption; potential privacy issues; requires reliable networkand cybersecurity management. |
| Edge / Federated Environments | Low latency and reduced data transmission;processing close to the sensor; supports privacy-preserving learning; scalable across device fleets. | Resource-constrained hardware; model compression and optimisation required; limited by energy and memory constraints. |

# Chapter 5

# Conclusions

The dynamics of industrial maintenance are undergoing a historic transition: from the concept of failure as a sudden and inevitable event to the ability to anticipate, measure, and manage it. In this process, PHM is assuming a central role, and this thesis aimed to investigate how the introduction of Artificial Intelligence is effecting a transformation in the methods by which the reliability of engineering systems is predicted. Through a systematic literature review based on fifty-four recent studies, the work analyzed approaches, results, evaluation metrics, and real-world application cases, with the aim of understanding the maturity of the transition to AI and the challenges that remain.

The first question (RQ1) addressed concerns which methodologies currently dominate industrial prognostics. A clear trend emerges from the works analyzed: the use of data-driven methods is no longer an experimental element, but an established practice. Architectures such as Convolutional Neural Networks, LSTMs, autoencoders, and, more recently, Transformers, allow degradation patterns to be learned directly from sensory data, even in noisy and non-stationary conditions [41, 6, 80]. At the other end of the spectrum, model-based approaches continue to play a significant role when the physics of failure is known and modelable, especially through Bayesian filters, thermodynamic and mechanical models, or numerical simulators. However, the literature reveals an increasingly clear convergence: hybrid architectures, in which neural networks, physical constraints, and degradation models coexist, represent a concrete and promising direction for development. They integrate the flexibility of machine learning with the engineering consistency of modeling, mitigating limitations such as data scarcity and the risk of overfitting.

The second question (RQ2) concerns the actual advantage of AI over traditional methods. The picture that emerges is clear: in most cases, machine learning-based techniques achieve superior accuracy in Remaining Useful Life estimates, fault classification, and the ability to generalize to variable operating conditions. However, these improvements come at a price: increased computational complexity, the need for large, high-quality datasets, and a loss of decision-making transparency. This has led the scientific community to develop techniques for explainability, uncertainty quantification, and predictive reliability, topics that now appear essential for mature industrial use. In other words, AI is measured not only in terms of accuracy, but also in terms of its technical "credibility."

The third question (RQ3) of the thesis focuses on the areas in which AI for PHM has actually been applied. The most mature sectors are aerospace, manufacturing, mechanical systems, wind turbines, and the automotive sector [93, 70]. In turbofans, data-driven PHM is now an integral part of maintenance strategies and shows consolidated results thanks to datasets such as C-MAPSS [28]. In wind turbines, SCADA data analysis has allowed us to anticipate gearbox failures and optimize interventions on field-installed machinery. Adoption is also growing in manufacturing, primarily thanks to Industry 4.0 and the increased availability of sensors.

These results allow us to draw some critical conclusions. AI-based PHM is no longer just a scientific possibility, but a concrete technological path. The integration of data, physical models, edge computing, and cloud solutions opens up scenarios that make maintenance truly predictive, rather than preventive. However, industrial adoption cannot ignore issues such as the reliability of forecasts, the explainability of algorithmic decisions, robustness to noisy data, and the economic impact of deploying complex networks. In other words, AI is proving to work, but it must prove to be reliable.

The work is not without limitations, and recognizing them is an essential part of a scientific thesis. The choice to focus on recent years allows us to grasp the state of the art, but excludes historical techniques that laid the theoretical foundations of the field. Furthermore, some industrial sectors remain poorly documented, and several studies report results on experimental datasets rather than on real machines, making it difficult to assess the transferability of the methods.

Future prospects show a clear evolutionary path: hybrid physics-AI models, digital twins, federated learning to preserve data privacy, explainability integrated into diagnostic and prognostic tools, lightweight networks for edge computing, and standard metrics shared by the scientific community. It is likely that the industrial maturity of AI-based PHM will depend not only on predictive accuracy, but also on its integration with operational processes, risk management, and asset economics.

In conclusion, the analysis conducted shows that Artificial Intelligence is not replacing traditional prognostics: it is transforming it. Maintenance is no longer a sequence of scheduled interventions, but an informed, anticipatory, and adaptive system. The vision we are moving toward is that of industrial infrastructures capable of knowing themselves, predicting their operational future, and optimizing their decisions. This is a technological, but above all a cultural, shift, and represents one of the most significant steps towards the maintenance of the future.

The systematic mapping we have conducted, while defining a solid state of the art, is by its very nature a starting point. The critical challenges resisting the industrial assimilation of AI in PHM outline an extremely fertile research horizon, structured around three main directions that deserve immediate and targeted exploration.

The future cannot be satisfied with post-hoc hybrid models, where physics acts as a mere corrective filter. The next frontier lies in the development of intrinsically physics-informed architectures, where the fundamental degradation equations are not external constraints, but an integral part of the neural network's learning structure. This requires deeper research into how sensor data and thermodynamic or mechanical knowledge can be fused at the feature and loss function levels with greater refinement, ensuring not only more accurate prediction but also domain-based explainability (XAI).

Furthermore, while LSTMs have been the workhorse for sequences, the latent potential of Transformers for multivariable and multiscale systems, such as a turbofan, requires rigorous validation. Systematic benchmarking is imperative to assess whether the self-attention mechanism can provide a sufficient computational and explanatory advantage to justify its complexity of edge implementation.

The problem of failure data scarcity cannot be solved by simply waiting. Research must focus on the development of synthetic data generation techniques (using Generative Adversarial Networks or Variational Autoencoders) that go beyond statistical replication and realistically model the final degradation trajectories, which are often not represented in real datasets.

At the same time, true implementation at the Industrial IoT level is undermined by computational costs. Future research should focus on developing quantization and pruning techniques for deep models to make them compatible with the Edge Computing environment, where latency is critical and resources are limited. This includes exploring advanced lightweight architectures, such as binarized neural networks, that can ensure trustworthiness without overloading the hardware.

Finally, widespread adoption is inextricably linked to standardization. The scientific community is challenged to define universal metrics and validation protocols for trustworthiness. These standards should include not only UQ metrics (CRPS, PICP) but also evaluation of the explanatory robustness of XAI techniques [89, 60].

The culmination of these research directions is their functional integration into the concept of the Digital Twin. The AI-based PHM model, armed with explainability and uncertainty awareness, is poised to become the "predictive brain" of every digital twin. The final challenge is to ensure that RUL prediction translates into self-regulating decision input for the twin, enabling true real-time system lifecycle management, thus closing the loop between predictive science and operational practice.

# Bibliography


[1] Jay Lee, Fei Wu, Weibin Zhao, Masoud Ghaffari, Lin Liao, and David Siegel. «Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications». In: *Mechanical Systems and Signal Processing* 42.1-2 (2014), pp. 314–334 (cit. on pp. 2, 23, 26).

[2] Jin Liu, Nagi Gebraeel, Jing Shi, and Chuan Zhou. «Remaining useful life prediction of aircraft engines using a hybrid approach». In: *Reliability Engineering & System Safety* 179 (2018), pp. 1–11 (cit. on pp. 2, 16, 22).

[3] Haitao Wang, Jie Yang, Ruihua Wang, and Lichen Shi. «Remaining Useful Life Prediction of Bearings Based on Convolution Attention Mechanism and Temporal Convolution Network». In: *IEEE Access* 11 (2023), pp. 24407–24419. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3255891. (Visited on 10/06/2025) (cit. on pp. 2, 43, 57, 59, 62, 63, 66, 68, 69, 73).

[4] Yaguo Lei, Naipeng Li, Liang Guo, Ning Li, Tingting Yan, and Jing Lin. «Machinery health prognostics: A systematic review from data acquisition to RUL prediction». In: *Mechanical Systems and Signal Processing* 104 (2018), pp. 799–834 (cit. on pp. 2, 4, 17, 23, 24).

[5] Kai Goebel, Bhaskar Saha, Abhinav Saxena, Indranil Roychoudhury, Mark Walker, and Jose Celaya. «Prognostics: the science of making predictions». In: *2008 IEEE Aerospace Conference*. IEEE. 2008, pp. 1–7 (cit. on pp. 2, 21, 22).

[6] Shuwen Zhang, Shuning Zhang, Bin Wang, and Thomas G Habetler. «Deep learning algorithms for bearing fault diagnostics—A comprehensive review». In: *IEEE Access* 6 (2018), pp. 63503–63516 (cit. on pp. 2, 14, 22, 24, 87).

[7] J. Tian et al. «AI-enabled prognostics and health management for electric vehicle fleets». In: *IEEE Transactions on Transportation Electrification* 7.3 (2021), pp. 1425–1438 (cit. on pp. 3, 28).

[8] Barbara Kitchenham and Stuart Charters. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Tech. rep. EBSE Technical Report, 2007 (cit. on pp. 3, 4, 32, 36, 38).

[9] Angela Carrera-Rivera, William Ochoa, Felix Larrinaga, and Ganix Lasa. «How-to Conduct a Systematic Literature Review: A Quick Guide for Computer Science Research». In: *MethodsX* 9 (Jan. 2022), p. 101895. ISSN: 2215-0161. DOI: `10.1016/j.mex.2022.101895`. (Visited on 07/28/2024) (cit. on pp. 3, 32, 33, 35).

[10] Pearl Brereton, Barbara Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. «Lessons from applying the systematic literature review process within the software engineering domain». In: *Journal of Systems and Software* 80.4 (2007), pp. 571–583 (cit. on p. 4).

[11] David Tranfield, David Denyer, and Palminder Smart. «Towards a methodology for developing evidence-informed management knowledge by means of systematic review». In: *British Journal of Management* 14.3 (2003), pp. 207–222 (cit. on pp. 4, 32, 38).

[12] Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, and Robert X Gao. «Deep learning and its applications to machine health monitoring». In: *Mechanical Systems and Signal Processing* 115 (2017), pp. 213–237 (cit. on pp. 4, 8, 22, 24).

[13] David Moher, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G Altman. «Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement». In: *PLoS medicine* 6.7 (2009), e1000097 (cit. on pp. 5, 37, 39).

[14] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Third edition, Global edition. Prentice Hall Series in Artificial Intelligence. Boston Columbus Indianapolis: Pearson, 2016. ISBN: 978-0-13-604259-4 978-1-292-15397-1 (cit. on pp. 5, 6).

[15] Alan M. Turing. «Computing Machinery and Intelligence». In: *Mind* 59.236 (1950), pp. 433–460 (cit. on p. 6).

[16] John McCarthy. *What is Artificial Intelligence?* `http://www-formal.stanford.edu/jmc/whatisai/`. Online report, Stanford University. 2007 (cit. on p. 6).

[17] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, 1997 (cit. on pp. 6, 9).

[18] Allen Newell and Herbert A. Simon. «Computer Science as Empirical Inquiry: Symbols and Search». In: *Communications of the ACM* 19.3 (1976), pp. 113–126 (cit. on p. 6).

[19] Feng Jia, Yaguo Lei, Jing Lin, Xiaodong Zhou, and Na Lu. «A machine learning approach for machine fault diagnostics and prognostics». In: *Mechanical Systems and Signal Processing* 138 (2019), p. 106613 (cit. on pp. 8, 26, 27).

[20] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. 2nd. Cambridge, MA: MIT Press, 2018. ISBN: 9780262039246 (cit. on pp. 8, 12).

[21] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. «Deep learning». In: *Nature* 521.7553 (2015), pp. 436–444 (cit. on p. 11).

[22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016 (cit. on pp. 11, 12).

[23] David Silver et al. «Mastering the game of Go with deep neural networks and tree search». In: *Nature* 529.7587 (2016), pp. 484–489. DOI: `10.1038/nature16961` (cit. on p. 13).

[24] Yaguo Lei et al. «PHM in wind energy: Current progress and future prospects». In: *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 1917–1934 (cit. on pp. 13, 14, 28).

[25] Bingguo Liu, Zhuo Gao, Binghui Lu, Hangcheng Dong, and Zeru An. «Deep Learning-Based Remaining Useful Life Estimation of Bearings with Time-Frequency Information». In: *Sensors* 22.19 (Jan. 2022), p. 7402. ISSN: 1424-8220. DOI: `10.3390/s22197402`. (Visited on 10/07/2025) (cit. on pp. 15, 42, 57, 59, 61, 79).

[26] Yunwei Zhang, Qiaochu Tang, Yao Zhang, Jiabin Wang, Ulrich Stimming, and Alpha A. Lee. «Identifying Degradation Patterns of Lithium Ion Batteries from Impedance Spectroscopy Using Machine Learning». In: *Nature Communications* 11.1 (Apr. 2020), p. 1706. ISSN: 2041-1723. DOI: `10.1038/s41467-020-15235-7`. (Visited on 10/07/2025) (cit. on pp. 16, 43, 54, 60, 79, 83).

[27] Minhee Kim and Kaibo Liu. «A Bayesian Deep Learning Framework for Interval Estimation of Remaining Useful Life in Complex Systems by Incorporating General Degradation Characteristics». In: *IISE Transactions* 53.3 (Dec. 2020), pp. 326–340. ISSN: 2472-5854. DOI: `10.1080/24725854.2020.1766729`. (Visited on 10/07/2025) (cit. on pp. 17, 42, 54, 57, 67–70, 74).

[28] Abhinav Saxena, Kai Goebel, Don Simon, and Neil Eklund. «Metrics for evaluating performance of prognostic techniques». In: *2008 International Conference on Prognostics and Health Management*. IEEE. 2008, pp. 1–17 (cit. on pp. 21, 22, 70, 88).

[29] Peter Kalgren, Carl Byington, Michael Roemer, and Matthew Watson. «Application of prognostic health management in digital electronic systems». In: *2007 IEEE Aerospace Conference*. IEEE. 2007, pp. 1–9 (cit. on p. 21).

[30] Andrew Heng, Shuzhi Sam Zhang, Andy CK Tan, and Joseph Mathew. «Rotating machinery prognostics: State of the art, challenges and opportunities». In: *Mechanical Systems and Signal Processing* 23.3 (2009), pp. 724–739 (cit. on pp. 21, 23).

[31] Andrea Coraddu, Luca Oneto, Andrea Ghio, Silvia Savio, Davide Anguita, and Massimo Figari. «Machine learning approaches for condition-based maintenance». In: *Naval Engineers Journal* 128.3 (2016), pp. 115–125 (cit. on pp. 22, 24, 26, 29).

[32] Sudhakar Srinivasan, Palaniappan Ramasamy, and Muthukumar Subramaniam. «Aircraft health monitoring system with predictive analytics». In: *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE. 2016, pp. 1–8 (cit. on p. 22).

[33] Naipeng Li, Yaguo Lei, Jing Lin, Shiming X Ding, and Jun Zhang. «Remaining useful life prediction based on a general expression of stochastic degradation processes». In: *Mechanical Systems and Signal Processing* 116 (2019), pp. 251–270 (cit. on pp. 23, 24).

[34] Jay Lee, Jun Ni, Dragan Djurdjanovic, Hailing Qiu, and Huaiqing Liao. «Reconfigurable manufacturing systems: From design to implementation». In: *CIRP Annals* 55.1 (2006), pp. 481–494 (cit. on p. 25).

[35] Dominik Thomas and Patrick Klein. «Industry 4.0: The future of productivity and growth in manufacturing industries». In: *Journal of Manufacturing Science and Engineering* (2016) (cit. on pp. 25, 27).

[36] Jay Lee, Behrad Bagheri, and Hung-An Kao. «Cyber physical systems and manufacturing industry». In: *Manufacturing Letters* 3 (2015), pp. 18–23 (cit. on p. 26).

[37] Xun Xu, Yongxin Lu, Birgit Vogel-Heuser, and Lihui Wang. «Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues». In: *Robotics and Computer-Integrated Manufacturing* 61 (2020), p. 101837 (cit. on p. 26).

[38] Rui Zhao, Dazhong Wang, Ruqiang Yan, Kezhi Mao, Fei Shen, and Robert X Gao. «Deep learning and its applications to machine health monitoring». In: *Mechanical Systems and Signal Processing* 115 (2019), pp. 213–237 (cit. on p. 26).

[39] Cristian Ceclu et al. «Prognostics and health management for connected and electric vehicles: A review». In: *IEEE Access* 10 (2022), pp. 12045–12062 (cit. on p. 28).

[40] Juan Serradilla et al. «AI and hybrid modelling for prognostics in wind energy: Challenges and future directions». In: *Renewable Energy* 179 (2021), pp. 1423–1438 (cit. on pp. 28, 29).

[41] Y. Li et al. «Battery health management system for electric vehicles: A data-driven review». In: *Applied Energy* 236 (2019), pp. 108–124 (cit. on pp. 28, 29, 87).

[42] Andrew Kusiak and Anoop Verma. «Review of SCADA data-driven approaches for wind turbine condition monitoring». In: *Renewable and Sustainable Energy Reviews* 13.9 (2011), pp. 2739–2747 (cit. on pp. 28, 29).

[43] Wei Zhang et al. «Hybrid physics-based and data-driven prognostics for wind turbine gearboxes». In: *Renewable Energy* 145 (2020), pp. 2293–2303 (cit. on p. 28).

[44] Marcia L. Baptista and Elsa M.P. Henriques. «1D-DGAN-PHM: A 1-D Denoising GAN for Prognostics and Health Management with an Application to Turbofan». In: *Applied Soft Computing* 131 (Dec. 2022), p. 109785. ISSN: 15684946. DOI: 10.1016/j.asoc.2022.109785. (Visited on 10/03/2025) (cit. on pp. 42, 46, 47, 50, 52, 53, 55, 58, 60, 68, 78, 79, 81, 84).

[45] Owais Asif, Sajjad Ali Haider, Syed Rameez Naqvi, John F. W. Zaki, Kyung-Sup Kwak, and S. M. Riazul Islam. «A Deep Learning Model for Remaining Useful Life Prediction of Aircraft Turbofan Engine on C-MAPSS Dataset». In: *IEEE Access* 10 (2022), pp. 95425–95440. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3203406. (Visited on 10/06/2025) (cit. on pp. 42, 46, 47, 50, 53, 55, 56, 62, 67, 70, 78, 83).

[46] K. Dhananjay Rao, A. Ramakrishna, M. Ramesh, Pallanti Koushik, Subhojit Dawn, P. Pavani, Taha Selim Ustun, and Umit Cali. «A Hyperparameter-Tuned LSTM Technique-Based Battery Remaining Useful Life Estimation Considering Incremental Capacity Curves». In: *IEEE Access* 12 (2024), pp. 127259–127271. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3450871. (Visited on 10/05/2025) (cit. on pp. 42, 57, 63, 68, 75, 82, 84).

[47] Khanh T.P. Nguyen and Kamal Medjaher. «A New Dynamic Predictive Maintenance Framework Using Deep Learning for Failure Prognostics». In: *Reliability Engineering & System Safety* 188 (Aug. 2019), pp. 251–262. ISSN: 09518320. DOI: 10.1016/j.ress.2019.03.018. (Visited on 10/07/2025) (cit. on pp. 42, 60).

[48] Li Jiang, Biaobiao Cao, Xin Zhang, Bingyang Chen, Lei Wang, and Yibing Li. «A Novel Spatio-Temporal Characteristic Extraction Network for Bearing Remaining Useful Life Prediction». In: *Measurement Science and Technology* 35.11 (Aug. 2024), p. 116142. ISSN: 0957-0233. DOI: 10.1088/1361-6501/ad6f37. (Visited on 10/05/2025) (cit. on pp. 42, 62).

[49] Xinping Chen. «A Novel Transformer-Based DL Model Enhanced by Position-Sensitive Attention and Gated Hierarchical LSTM for Aero-Engine RUL Prediction». In: *Scientific Reports* 14.1 (May 2024), p. 10061. ISSN: 2045-2322. DOI: 10.1038/s41598-024-59095-3. (Visited on 10/05/2025) (cit. on pp. 42, 57, 61–64).

[50] Yiming Bai and Jinsong Zhao. «A Novel Transformer-Based Multi-Variable Multi-Step Prediction Method for Chemical Process Fault Prognosis». In: *Process Safety and Environmental Protection* 169 (Jan. 2023), pp. 937–947. ISSN: 09575820. DOI: 10.1016/j.psep.2022.11.062. (Visited on 10/08/2025) (cit. on pp. 42, 46, 48–50, 52–55, 62, 64, 68, 83).

[51] Jose Caceres, Danilo Gonzalez, Taotao Zhou, and Enrique Lopez Droguett. «A Probabilistic Bayesian Recurrent Neural Network for Remaining Useful Life Prognostics Considering Epistemic and Aleatory Uncertainties». In: *Structural Control and Health Monitoring* 28.10 (2021), e2811. ISSN: 1545-2263. DOI: 10.1002/stc.2811. (Visited on 10/07/2025) (cit. on pp. 42, 57, 65, 67–69, 71, 77, 82).

[52] Boubker Najdi, Mohammed Benbrahim, and Mohammed Nabil Kabbaj. «Adaptive Res-LSTM Attention-based Remaining Useful Lifetime Prognosis of Rolling Bearings». In: *International Journal of Prognostics and Health Management* 16.1 (2025). ISSN: 2153-2648. DOI: 10.36001/ijphm.2025.v16i1.4171 (cit. on pp. 42, 57).

[53] Gopal Krishna, Rajesh Singh, Anita Gehlot, Ahmad Almogren, Ayman Altameem, Ateeq Ur Rehman, and Seada Hussen. «Advanced Battery Management System Enhancement Using IoT and ML for Predicting Remaining Useful Life in Li-ion Batteries». In: *Scientific Reports* 14.1 (Dec. 2024), p. 30394. ISSN: 2045-2322. DOI: 10.1038/s41598-024-80719-1. (Visited on 10/08/2025) (cit. on pp. 42, 68, 72, 78, 79).

[54] Abubakar Bala, Idris Ismail, Rosdiazli Ibrahim, Sadiq M. Sait, and Diego Oliva. «An Improved Grasshopper Optimization Algorithm Based Echo State Network for Predicting Faults in Airplane Engines». In: *IEEE Access* 8 (2020), pp. 159773–159789. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.3020356. (Visited on 10/05/2025) (cit. on pp. 42, 47, 49, 50, 55, 62, 63).

[55] Baobao Zhang, Jianjie Zhang, Peibo Yu, Jianhui Cao, and Yihang Peng. «Asymmetric-Based Residual Shrinkage Encoder Bearing Health Index Construction and Remaining Life Prediction». In: *Sensors* 24.20 (Jan. 2024), p. 6510. ISSN: 1424-8220. DOI: 10.3390/s24206510. (Visited on 10/06/2025) (cit. on pp. 42, 52, 53, 55, 60, 67–69).

[56] Zijian Zhao and Pengyuan Zou. «Attention-Based Dual-channel Deep Neural Network For Aero-engine RUL Prediction Under Time-varying Operating Conditions». In: *Journal of Physics: Conference Series* 2386.1 (Dec. 2022), p. 012027. ISSN: 1742-6596. DOI: 10.1088/1742-6596/2386/1/012027. (Visited on 10/05/2025) (cit. on pp. 42, 66).

[57] Abdeltif Boujamza and Saad Lissane Elhaq. «Attention-Based LSTM for Remaining Useful Life Estimation of Aircraft Engines». In: *IFAC-PapersOnLine* 55.12 (2022), pp. 450–455. ISSN: 24058963. DOI: 10.1016/j.ifacol.2022.07.353. (Visited on 10/05/2025) (cit. on pp. 42, 46, 47, 50, 56, 62, 67).

[58] Vahid Safavi, Arash Mohammadi Vaniar, Najmeh Bazmohammadi, Juan C. Vasquez, and Josep M. Guerrero. «Battery Remaining Useful Life Prediction Using Machine Learning Models: A Comparative Study». In: *Information* 15.3 (Mar. 2024), p. 124. ISSN: 2078-2489. DOI: 10.3390/info15030124. (Visited on 10/05/2025) (cit. on pp. 42, 62, 63, 75, 77, 82, 84).

[59] Ganjour Mazaev, Guillaume Crevecoeur, and Sofie Van Hoecke. «Bayesian Convolutional Neural Networks for Remaining Useful Life Prognostics of Solenoid Valves With Uncertainty Estimations». In: *IEEE Transactions on Industrial Informatics* 17.12 (Dec. 2021), pp. 8418–8428. ISSN: 1941-0050. DOI: 10.1109/TII.2021.3078193. (Visited on 10/05/2025) (cit. on pp. 42, 57, 67, 74).

[60] Sunday Ochella, Fateme Dinmohammadi, and Mahmood Shafiee. «Bayesian Neural Networks for Uncertainty Quantification in Remaining Useful Life Prediction of Systems with Sensor Monitoring». In: *Advances in Mechanical Engineering* 16.7 (July 2024), p. 16878132241239802. ISSN: 1687-8132, 1687-8140. DOI: 10.1177/16878132241239802. (Visited on 10/08/2025) (cit. on pp. 42, 57, 65, 67–69, 71, 74, 77, 80, 82, 89).

[61] Madhav Mishra, Jesper Martinsson, Kai Goebel, and Matti Rantatalo. «Bearing Life Prediction With Informed Hyperprior Distribution: A Bayesian Hierarchical and Machine Learning Approach». In: *IEEE Access* 9 (2021), pp. 157002–157011. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3130157. (Visited on 10/05/2025) (cit. on p. 42).

[62] Ravi Pandit, Matilde Santos, and Jesus Enrique Sierra-García. «Comparative Analysis of Novel Data-Driven Techniques for Remaining Useful Life Estimation of Wind Turbine High-Speed Shaft Bearings». In: *Energy Science & Engineering* 12.10 (2024), pp. 4613–4623. ISSN: 2050-0505. DOI: 10.1002/ese3.1911. (Visited on 10/08/2025) (cit. on pp. 42, 62–64, 68, 71–73, 75, 76, 79, 81, 84).

[63] Dongdong Zhao and Feng Liu. «Cross-Condition and Cross-Platform Remaining Useful Life Estimation via Adversarial-Based Domain Adaptation». In: *Scientific Reports* 12.1 (Jan. 2022), p. 878. ISSN: 2045-2322. DOI: 10.1038/s41598-021-03835-2. (Visited on 10/08/2025) (cit. on pp. 42, 55, 57, 59, 61, 65, 67, 69, 75, 77, 80).

[64] Bin cheng Wen, Ming qing Xiao, Xue qi Wang, Xin Zhao, Jian feng Li, and Xin Chen. «Data-Driven Remaining Useful Life Prediction Based on Domain Adaptation». In: *PeerJ Computer Science* 7 (Sept. 2021), e690. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.690. (Visited on 10/05/2025) (cit. on pp. 42, 57, 59, 61, 65, 67, 69, 75, 77, 80).

[65] Xu Wang, Tianyang Wang, Anbo Ming, Qinkai Han, Fulei Chu, Wei Zhang, and Aihua Li. «Deep Spatiotemporal Convolutional-Neural-Network-Based Remaining Useful Life Estimation of Bearings». In: *Chinese Journal of Mechanical Engineering* 34.1 (June 2021), p. 62. ISSN: 2192-8258. DOI: 10.1186/s10033-021-00576-1. (Visited on 10/05/2025) (cit. on pp. 42, 57, 59, 62).

[66] Feiyue Deng, Yan Bi, Yongqiang Liu, and Shaopu Yang. «Deep-Learning-Based Remaining Useful Life Prediction Based on a Multi-Scale Dilated Convolution Network». In: *Mathematics* 9.23 (Jan. 2021), p. 3035. ISSN: 2227-7390. DOI: 10.3390/math9233035. (Visited on 10/06/2025) (cit. on pp. 42, 59).

[67] Tarek Berghout and Mohamed Benbouzid. «Diagnosis and Prognosis of Faults in High-Speed Aeronautical Bearings with a Collaborative Selection Incremental Deep Transfer Learning Approach». In: *Applied Sciences* 13.19 (Jan. 2023), p. 10916. ISSN: 2076-3417. DOI: 10.3390/app131910916. (Visited on 10/06/2025) (cit. on pp. 42, 46–53, 55, 59).

[68] Chenyu Liu, Alexandre Mauricio, Junyu Qi, Dandan Peng, and Konstantinos Gryllias. «Domain Adaptation Digital Twin for Rolling Element Bearing Prognostics». In: *Annual Conference of the PHM Society* 12.1 (Nov. 2020), pp. 10–10. ISSN: 2325-0178. DOI: 10.36001/phmconf.2020.v12i1.1294. (Visited on 10/06/2025) (cit. on pp. 43, 68, 69).

[69] Wei Yuan, Xinlong Li, Hongbin Gu, Faye Zhang, and Fei Miao. «Engine Remaining Useful Life Prediction Based on PSO Optimized Multi-Layer Long Short-Term Memory and Multi-Source Information Fusion». In: *Measurement and Control* 57.5 (May 2024), pp. 638–649. ISSN: 0020-2940. DOI: 10.1177/00202940231214868. (Visited on 10/08/2025) (cit. on pp. 43, 63, 64).

[70] Wanwan Zhang, Jørn Vatn, and Adil Rasheed. «Gearbox Pump Failure Prognostics in Offshore Wind Turbine by an Integrated Data-Driven Model». In: *Applied Energy* 380 (Feb. 2025), p. 124829. ISSN: 03062619. DOI: 10.1016/

j.apenergy.2024.124829. (Visited on 10/06/2025) (cit. on pp. 43, 57, 64, 71, 72, 75, 79, 82, 84, 88).

[71] Eugen Boos, Jan Zimmermann, Hajo Wiemer, and Steffen Ihlenfeldt. «Investigation of Alternative Attention Modules in Transformer Models for Remaining Useful Life Predictions: Addressing Challenges in High-Frequency Time-Series Data». In: *Procedia CIRP* 122 (2024), pp. 85–90. ISSN: 22128271. DOI: 10.1016/j.procir.2024.01.012. (Visited on 10/06/2025) (cit. on pp. 43, 46, 48, 53, 54, 61–64, 73, 81, 83).

[72] Alberto Jaenal, Jose-Raul Ruiz-Sarmiento, and Javier Gonzalez-Jimenez. «MachNet, a General Deep Learning Architecture for Predictive Maintenance within the Industry 4.0 Paradigm». In: *Engineering Applications of Artificial Intelligence* 127 (Jan. 2024), p. 107365. ISSN: 09521976. DOI: 10.1016/j.engappai.2023.107365. (Visited on 10/06/2025) (cit. on pp. 43, 72, 77, 80–84).

[73] Yang Liu, Bihe Xu, and Yangli-ao Geng. «Multi-Condition Remaining Useful Life Prediction Based on Mixture of Encoders». In: *Entropy* 27.1 (Jan. 2025), p. 79. ISSN: 1099-4300. DOI: 10.3390/e27010079. (Visited on 10/06/2025) (cit. on pp. 43, 64, 73, 79, 82).

[74] D. Moros, N. Berrabah, and I. Ashton. «Neural Networks for Offshore Wind Turbine Converter Failure Prognosis». In: *IFAC-PapersOnLine* 58.4 (2024), pp. 592–597. ISSN: 24058963. DOI: 10.1016/j.ifacol.2024.07.283. (Visited on 10/06/2025) (cit. on pp. 43, 62).

[75] David Solís-Martín, Juan Galán-Páez, and Joaquín Borrego-Díaz. «On the Soundness of XAI in Prognostics and Health Management (PHM)». In: *Information* 14.5 (May 2023), p. 256. ISSN: 2078-2489. DOI: 10.3390/info14050256. (Visited on 10/08/2025) (cit. on pp. 43, 58, 66, 68, 72, 73, 75, 76).

[76] Abdeltif Boujamza and Saâd Lissane Elhaq. «Optimizing Remaining Useful Life Predictions for Aircraft Engines: A Dilated Recurrent Neural Network Approach». In: *IFAC-PapersOnLine* 58.13 (2024), pp. 811–816. ISSN: 24058963. DOI: 10.1016/j.ifacol.2024.07.582. (Visited on 10/06/2025) (cit. on pp. 43, 46, 47, 50, 52, 54, 56, 67).

[77] Tarek Berghout, Mohamed-Djamel Mouss, Leïla-Hayet Mouss, and Mohamed Benbouzid. «ProgNet: A Transferable Deep Network for Aircraft Engine Damage Propagation Prognosis under Real Flight Conditions». In: *Aerospace* 10.1 (Jan. 2023), p. 10. ISSN: 2226-4310. DOI: 10.3390/aerospace10010010. (Visited on 10/06/2025) (cit. on pp. 43, 46, 48, 51, 52, 68, 69, 76, 80, 81, 83).

[78] Arch Desai, Yi Guo, Shawn Sheng, Caleb Phillips, and Lindy Williams. «Prognosis of Wind Turbine Gearbox Bearing Failures Using SCADA and Modeled Data». In: *Annual Conference of the PHM Society* 12.1 (Nov. 2020), pp. 10–10. ISSN: 2325-0178. DOI: 10.36001/phmconf.2020.v12i1.1292. (Visited on 10/08/2025) (cit. on pp. 43, 79).

[79] Sina Sharif Mansouri, Petros Karvelis, George Georgoulas, and George Nikolakopoulos. «Remaining Useful Battery Life Prediction for UAVs Based on Machine Learning». In: *IFAC-PapersOnLine* 50.1 (July 2017), pp. 4727–4732. ISSN: 24058963. DOI: 10.1016/j.ifacol.2017.08.863. (Visited on 10/07/2025) (cit. on pp. 43, 79).

[80] Qianxia Ma, Ming Zhang, Yuchun Xu, Jingyan Song, and Tao Zhang. «Remaining Useful Life Estimation for Turbofan Engine with Transformer-based Deep Architecture». In: *2021 26th International Conference on Automation and Computing (ICAC)*. Sept. 2021, pp. 1–6. DOI: 10.23919/ICAC50006.2021.9594150. (Visited on 10/07/2025) (cit. on pp. 43, 62–64, 67, 73, 75, 87).

[81] Qiankun Hu, Yongping Zhao, Yuqiang Wang, Pei Peng, and Lihua Ren. «Remaining Useful Life Estimation in Prognostics Using Deep Reinforcement Learning». In: *IEEE Access* 11 (2023), pp. 32919–32934. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3263196. (Visited on 10/06/2025) (cit. on p. 43).

[82] Moncef Soualhi, K. Nguyen, Kamal Medjaher, and N. Zerhouni. «Remaining Useful Life Estimation of Turbofan Engines Using Adaptive Fault Detection Learning». In: *Annual Conference of the PHM Society* 14.1 (Oct. 2022). ISSN: 2325-0178. DOI: 10.36001/phmconf.2022.v14i1.3261. (Visited on 10/06/2025) (cit. on pp. 43, 65, 71, 76, 78, 81).

[83] Kıymet Ensarioğlu, Tülin İnkaya, and Erdal Emel. «Remaining Useful Life Estimation of Turbofan Engines with Deep Learning Using Change-Point Detection Based Labeling and Feature Engineering». In: *Applied Sciences* 13.21 (Jan. 2023), p. 11893. ISSN: 2076-3417. DOI: 10.3390/app132111893. (Visited on 10/08/2025) (cit. on pp. 43, 56, 62, 68, 70, 74, 78, 81, 84).

[84] Yifei Zhou, Shunli Wang, Yanxing Xie, Xianfeng Shen, and Carlos Fernandez. «Remaining Useful Life Prediction and State of Health Diagnosis for Lithium-Ion Batteries Based on Improved Grey Wolf Optimization Algorithm-Deep Extreme Learning Machine Algorithm». In: *Energy* 285 (Dec. 2023), p. 128761. ISSN: 03605442. DOI: 10.1016/j.energy.2023.128761. (Visited on 10/07/2025) (cit. on pp. 43, 57, 59, 60, 63, 64, 67, 71, 74, 77, 79).

[85] Mingyan Wu, Qing Ye, Jianxin Mu, Zuyu Fu, and Yilin Han. «Remaining Useful Life Prediction via a Data-Driven Deep Learning Fusion Model-CALAP». In: *IEEE Access* 11 (2023), pp. 112085–112096. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3322733. (Visited on 10/06/2025) (cit. on pp. 43, 54, 59, 62, 65, 71, 72, 77, 81, 83, 84).

[86] André Listou Ellefsen, Emil Bjørlykhaug, Vilmar Æsøy, Sergey Ushakov, and Houxiang Zhang. «Remaining Useful Life Predictions for Turbofan Engine Degradation Using Semi-Supervised Deep Architecture». In: *Reliability Engineering & System Safety* 183 (Mar. 2019), pp. 240–251. ISSN: 09518320. DOI: 10.1016/j.ress.2018.11.027. (Visited on 10/07/2025) (cit. on pp. 43, 57, 60, 64, 73, 75, 79).

[87] Jian Li, Faguo Huang, Haihua Qin, and Jiafang Pan. «Research on Remaining Useful Life Prediction of Bearings Based on MBCNN-BiLSTM». In: *Applied Sciences* 13.13 (Jan. 2023), p. 7706. ISSN: 2076-3417. DOI: 10.3390/app13137706. (Visited on 10/08/2025) (cit. on pp. 43, 52, 54, 56, 57, 61, 62, 73).

[88] L. Magadán, J.C. Granda, and F.J. Suárez. «Robust Prediction of Remaining Useful Lifetime of Bearings Using Deep Learning». In: *Engineering Applications of Artificial Intelligence* 130 (Apr. 2024), p. 107690. ISSN: 09521976. DOI: 10.1016/j.engappai.2023.107690. (Visited on 10/06/2025) (cit. on pp. 43, 60, 68).

[89] Yazan Alomari and Mátyás Andó. «SHAP-based Insights for Aerospace PHM: Temporal Feature Importance, Dependencies, Robustness, and Interaction Analysis». In: *Results in Engineering* 21 (Mar. 2024), p. 101834. ISSN: 25901230. DOI: 10.1016/j.rineng.2024.101834. (Visited on 10/06/2025) (cit. on pp. 43, 46, 48, 49, 51, 58, 61, 65, 66, 68, 69, 71–73, 75–78, 89).

[90] Chong Chen, Tao Wang, Ying Liu, Lianglun Cheng, and Jian Qin. «Spatial Attention-Based Convolutional Transformer for Bearing Remaining Useful Life Prediction». In: *Measurement Science and Technology* 33.11 (Aug. 2022), p. 114001. ISSN: 0957-0233. DOI: 10.1088/1361-6501/ac7c5b. (Visited on 10/07/2025) (cit. on pp. 44, 52, 54, 57, 63, 64, 66, 68–70, 73, 75, 81, 83, 84).

[91] Sebastian Schwendemann and Axel Sikora. «Transfer-Learning-Based Estimation of the Remaining Useful Life of Heterogeneous Bearing Types Using Low-Frequency Accelerometers». In: *Journal of Imaging* 9.2 (Feb. 2023), p. 34. ISSN: 2313-433X. DOI: 10.3390/jimaging9020034. (Visited on 10/06/2025) (cit. on pp. 44, 57, 65, 67, 75, 77, 80).

[92] Martin Hervé De Beaulieu, Mayank Shekhar Jha, Hugues Garnier, and Farid Cerbah. «Unsupervised Remaining Useful Life Prediction through Long Range Health Index Estimation Based on Encoders-Decoders». In: *IFAC-PapersOnLine* 55.6 (2022), pp. 718–723. ISSN: 24058963. DOI: 10.1016/

j.ifacol.2022.07.212. (Visited on 10/07/2025) (cit. on pp. 44, 60, 66, 68, 69, 79, 81, 84).

[93] Ayush Verma, Donatella Zappalá, Shawn Sheng, and Simon J. Watson. «Wind Turbine Gearbox Fault Prognosis Using High-Frequency SCADA Data». In: *Journal of Physics: Conference Series* 2265.3 (May 2022), p. 032067. ISSN: 1742-6596. DOI: 10.1088/1742-6596/2265/3/032067. (Visited on 10/08/2025) (cit. on pp. 44, 57, 68, 70, 76, 79, 82–84, 88).

[94] Abhinav Saxena, Kai Goebel, Don Simon, and Neil Eklund. «Prognostics performance: Evaluation metrics and scoring methods». In: *Proceedings of the IEEE Aerospace Conference*. 2009 (cit. on pp. 71, 72).