Review

# A review on the advancements and challenges of artificial intelligence based models for predictive maintenance of water injection pumps in the oil and gas industry

Salama Mohamed Almazrouei[1] · Fikri Dweiri[1] · Ridvan Aydin[1] · Abdalla Alnaqbi[2]

© The Author(s) 2023     OPEN

## Abstract

This paper provides a comprehensive review on the Artificial Intelligence (AI) based models for predictive maintenance (PdM) of water injection pumps (WIPs) in the oil and gas industry (OGI). The review encompasses the selection of algorithms, data requirements, and optimization strategies, offering insights into advancements, challenges, and theoretical foundations including data pre-processing and feature selection. This review highlights AI-based PdM developments for WIPs, focusing on techniques and algorithms that enhance water injection pump performance and accurately predict maintenance needs. It emphasizes the effectiveness of algorithms in capturing pump data patterns and anomalies for proactive maintenance. Additionally, the review offers valuable insights for future research directions and practical implementations of AI in pump maintenance. This comprehensive assessment serves as a beacon for OGI experts in the selection of AI methods for pump maintenance, enabling them to refine their procedures, enhance efficiency, and reduce operational interruptions. A prominent highlight is the importance of data quality and interpretability, which play a pivotal role in facilitating well-informed decision-making during the integration of AI technologies into maintenance processes. This article focuses on the theoretical foundations of AI in the context of pump maintenance, providing a contribution to OGI industry. By integrating theoretical perspectives with real-world evidence, it offers insights for guiding future research and enhancing maintenance techniques. As a resource, it holds relevance for researchers, practitioners, and decision-makers within the OGI sector, contributing to the ongoing advancement of this field.

## Article Highlights

- Artificial Intelligence (AI) optimizes Oil & Gas pump maintenance for efficiency and reliability.

- Real-world case studies validate cost and time reductions, improving pump performance.
- Challenges in data management and ethical AI implementation require careful consideration.

✉ Salama Mohamed Almazrouei, U20105033@sharjah.ac.ae | [1]Department of Industrial Engineering and Engineering Management, College of Engineering, University of Sharjah, Sharjah, UAE. [2] ADNOC Offshore, Abu Dhabi, United Arab Emirates.

# 1 Introduction

The oil and gas industry (OGI) plays a crucial role in the global energy sector, providing the primary sources of fuel for various sectors, such as transportation, industry, and residential use. The OGI encompasses a wide array of activities, from the exploration, extraction, and refinement to the distribution of hydrocarbon resources, such as crude oil and natural gas [1]. Within this industry, the complexity of operations, vast infrastructure, and the presence of high-value assets underline the critical importance of efficient maintenance and reliability management. This is pivotal for achieving optimal production, ensuring safety, and enhancing cost-effectiveness [2]. Predictive maintenance (PdM) is the methodology employed in this context, incorporating diverse approaches like data analysis, machine learning (ML) algorithms, and sensor technologies to gather and analyze real-time operational and sensor data from crucial equipment [3]. Through the application of predictive models, it becomes feasible to discern intricate patterns and anomalies in equipment performance. This foresight enables the early identification of potential failures, thus adopting a proactive stance in mitigating the repercussions of unexpected downtime. PdM has proven to be a valuable asset to the oil and gas industry, helping to optimize its performance and cost-effectiveness [2]. PdM is crucial in the OGI, improving operational efficiency, reducing downtime, and optimizing maintenance strategies. Water Injection Pumps (WIPs) are vital for smooth operations and maximizing production concerning reservoir pressure and oil recovery [4]. PdM leverages advanced technologies and data analytics to predict failures and enable proactive maintenance based on real-time conditions [5]. This is especially critical for demanding equipment like Water Injection Pumps (WIPs) in the Oil and Gas Industry (OGI). Implementing AI-based PdM for these pumps offers enhanced strategies through cutting-edge AI techniques, including Machine Learning (ML) and deep learning (DL). These AI approaches can analyze extensive data to identify early warning indicators of potential failures, enhancing operational efficiency and reducing downtime [6]. Proactive maintenance planning, efficient resource allocation, and optimized spare parts inventory can be achieved based on real-time equipment conditions. Traditional maintenance approaches often result in unnecessary actions and costly disruptions, while reactive maintenance poses safety and environmental risks [7]. PdM addresses these challenges by leveraging advanced technologies and data analytics to monitor the real-time health and performance of upstream rotating equipment in the OGI

[6]. Through the continuous collection, harnessing, and analysis of real-time data sourced from a diverse array of outlets, encompassing sensors, control systems, and historical maintenance records, PdM models possess an inherent capability to identify anomalies, decipher patterns, and recognize early signals that may signify potential equipment failures [3]. By means of this extensive data scrutiny, PdM models facilitate the prompt detection of potential issues, thus enabling proactive maintenance interventions. This comprehensive data-driven approach equips organizations to preemptively anticipate and address impending challenges, effectively curtailing downtime, optimizing maintenance strategies, and augmenting operational efficiency [7]. The integration of PdM models instates a proactive maintenance paradigm that bolsters reliability, curtails costs, and extends the lifecycle of critical assets. The early identification and resolution of potential issues before they escalate into substantial failures empower operators to avert expensive breakdowns, mitigate production losses, and ensure the continuity of operations. Moreover, PdM offers the potential for more efficient resource planning and allocation [8, 9]. With the capability to accurately forecast maintenance needs, operators can optimize spare parts inventories, diminish the necessity for unplanned repairs, and streamline their maintenance schedules. PdM for upstream rotating equipment is increasingly vital due to industry complexity and criticality [4]. Digitalization and IoT provide abundant data for monitoring and optimizing equipment, enabling proactive maintenance and operational efficiency. Advanced analytics, ML, and AI in PdM unveil hidden patterns for more accurate maintenance recommendations [10]. This research aims to comprehensively review AI-based PdM of WIPs in the OGI. It provides insights into the theoretical foundations and applicability of AI-based models for WIPs. To achieve this objective, the following specific research objectives have been identified:

 (i). To examine the advancements in AI-based models for PdM of WIPs in the OGI.
 (ii). To identify the challenges associated with implementing AI-based models for PdM of WIPs.

Driven by the research objectives, the central inquiry that directs this paper from a theoretical standpoint is: What are the advancements and challenges encountered in the application of AI-based models for PdM of WIPs in OGI? This research question serves as a guiding framework for the literature review, theoretical analysis, and exploration of the identified objectives. By addressing this question, the research aims to contribute to understanding AI-based PdM in the context of WIPs, enabling researchers,

practitioners, and decision-makers to gain insights into the theoretical underpinnings and practical implications of implementing AI in maintenance strategies.

## 2 Overview of predictive maintenance techniques

Instead of hinging on fixed maintenance schedules or waiting for failures to manifest, PdM harnesses data-driven insights to optimize maintenance activities [11]. Gallab et al. [12] discusses an anticipation and decision-making approach (Prediction Model) in LPG sector maintenance activities with Multi-Agent Systems derived from AI. Their proposed approach aims to facilitate a better comprehension of maintenance activities within the LPG supply chain and ultimately lead to more effective safety analysis, enabling maintenance actors to make well-informed decisions for minimizing risks [5–9].

Several techniques are commonly employed in PdM to monitor and analyze equipment conditions [1–4, 13–15]. These techniques include:

(1) *Ongoing Monitoring of Condition* Condition monitoring entails the continuous or periodic measurement of various parameters like vibration, temperature, pressure, and oil analysis. Its purpose is to detect any abnormal behavior or changes in these parameters that may indicate potential equipment failures.

(2) *Analysis of Failure Modes and Effects (FMEA)* FMEA is a systematic method for identifying potential failure modes, their causes, and the effects these failures can have. It aids in prioritizing maintenance tasks and determining the criticality of different failure modes.

(3) *Utilization of Statistical Analysis* Statistical techniques, such as regression analysis, time series analysis, and pattern recognition, are employed to analyze historical data and identify trends, patterns, and anomalies. Statistical analysis aids in predicting future equipment behavior and identifying potential failure patterns.

(4) *Adoption of Reliability Centered Maintenance (RCM)* RCM is a structured maintenance approach that focuses on identifying the most critical maintenance tasks based on the equipment's function, failure consequences, and risk analysis. The aim of RCM is to optimize maintenance efforts by prioritizing tasks that have the greatest impact on reliability and safety.

Proper maintenance is crucial for optimizing the performance and safety of oil and gas operations, offering benefits like improved safety, increased efficiency, and cost savings [3–6]. Comprehensive maintenance

programs, including regular inspections and preventative measures, along with training and acquiring necessary tools, are essential [11]. PdM, utilizing data and technology to anticipate failures, is particularly relevant in the OGI due to its serious consequences [8–10]. By implementing PdM with technologies like sensors and ML, companies can enhance safety, efficiency, and cost-effectiveness in their operations [7].

## 3 Methodology

This study follows the systematic review approach and adheres to the review phases protocol outlined by Khan et al. [10]. The selection process for included studies aligns with the guidelines proposed by Munim et al. [16]. Figure 1 represents the stages and steps encompassed by the review protocol, presenting a flowchart that outlines the selection process for the studies included in the analysis. Subsequent sections delve into comprehensive elucidations of the review protocol and elaborate on the meticulous search and selection procedures undertaken within the systematic literature review.
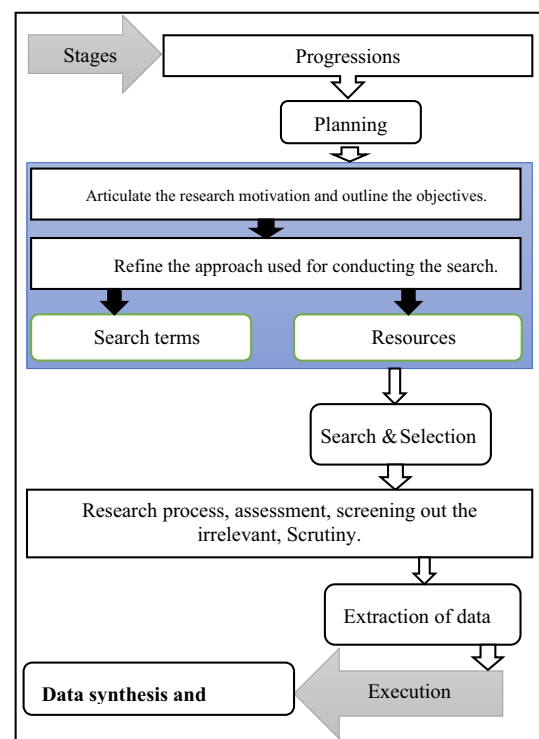


**Fig. 1** The sequential phases and steps of the review protocol

## 3.1 Inclusion and exclusion criteria for study selection

To ensure the meticulous selection of relevant studies, the review establishes precise inclusion and exclusion criteria. The inclusion criteria are specifically designed to encompass studies that delve into AI-based models or approaches for PdM of WIPs within the OGI, utilizing cutting-edge AI techniques such as ML or DL. Selected studies specifically discuss the PdM of WIPs, providing theoretical perspectives and methodologies related to AI-based models. Only peer-reviewed journal articles, conference papers, and scholarly publications are considered for inclusion to ensure high quality. Exclusion criteria eliminate studies that do not address AI-based models for PdM, focus on different equipment or industries, lack theoretical insights, or have redundant findings. Non-English studies are excluded due to resource limitations. By adhering to these criteria, the review selects relevant and high-quality studies to support the research objectives and address the research question effectively.

## 4 Research approach and strategy

This paper offers an in-depth examination through a systematic literature review of AI-based models utilized for PdM of WIPs within the OGI. Relevant studies were identified through systematic searches in academic databases such as Scopus, IEEE Xplore, ScienceDirect, and ACM Digital Library. The data extraction encompassed a range of elements, including research objectives, methodologies, theoretical frameworks, algorithms utilized, and discoveries concerning progressions and challenges. Employing thematic analysis on this data yielded a comprehensive portrayal of AI-based models for PdM, elucidated within the context of their theoretical foundations and frameworks. The findings were subsequently deliberated in the context of the research objectives, delivering theoretical implications and recommendations for both future research endeavors and practical applications. Systematic investigations occupy a prominent role in the realm of research, contributing valuable insights into the scholarly landscape and pinpointing areas of knowledge deficiency. These studies utilize quantitative techniques to scrutinize diverse facets of scientific publications, encompassing elements like citation patterns, collaborative networks, and prevailing research trends. By scrutinizing bibliographic data, researchers can attain a more profound comprehension of the existing corpus of literature, identify domains warranting further exploration, and gauge the influence and impact of specific research topics or disciplines. One application of bibliometric analysis lies in its ability to
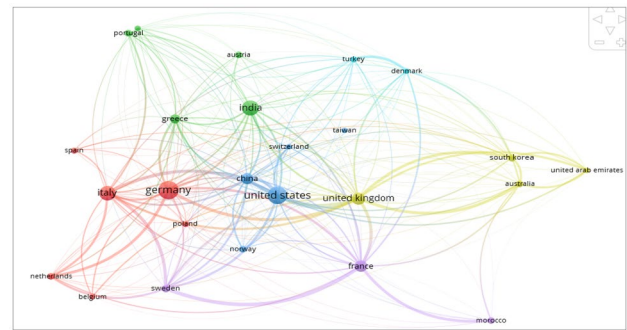


**Fig. 2** Visualization of bibliographic coupling and country relationships
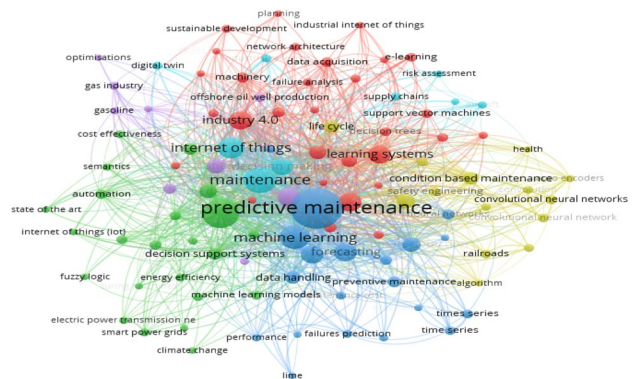


**Fig. 3** The relationship between co-occurrence and index keywords

explore the relationship between bibliographic coupling and countries. This analytical approach enables researchers to delve into collaborative patterns and knowledge exchange between scholars from different nations.

By studying the shared references among articles, researchers can identify clusters of countries that are closely connected in terms of their research output and collaboration networks. Figure 2 portrays the correlation between bibliographic coupling and countries, showcasing the interrelationship observed when employing the keywords "Predictive Maintenance" and "AI" in a search query. A total of 345 relevant document results pertaining to the topic were obtained from the search, specifically focusing on studies conducted between 2013 and 2023. Figure 2 highlights a substantial number of research studies focusing on this area, indicating the significance and interest in PdM using AI across multiple countries. Figure 3 presents the relationship between co-occurrence and index keywords. Notably, it is evident that PdM has garnered substantial attention in recent research. The figure showcases the frequency of co-occurrence of keywords related to PdM, highlighting its prominence, and indicating a strong research focus in
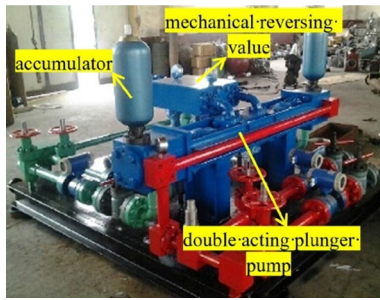
**Fig. 4** Dominant keyword research in OGI
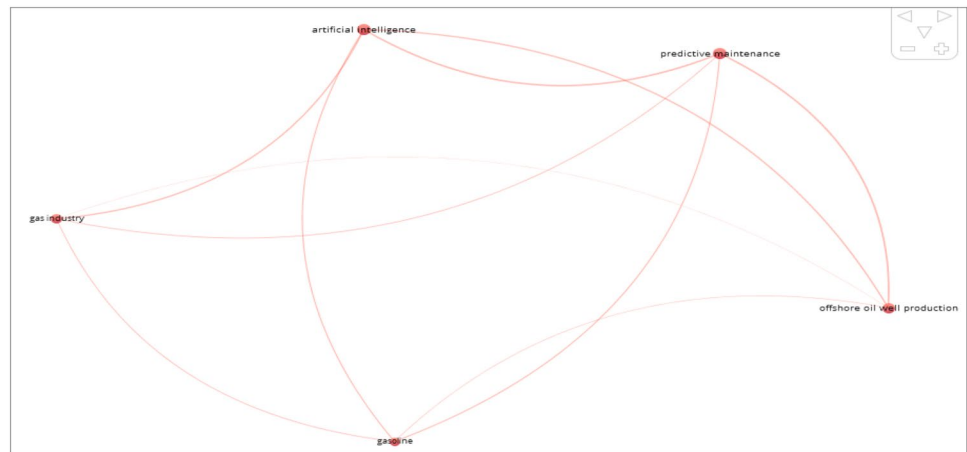


this area. When the keyword combination of TITLE-ABS-KEY ("Predictive Maintenance" AND "AI" AND "OIL") was used, the search yielded a total of 16 document results. Furthermore, Fig. 4 provides a visualization of the dominant keyword research in the OGI in the context of the search query. It offers insights into the collaborative patterns and knowledge exchange among different countries. This figure showcases the frequency and associations between different keywords, providing valuable information about the research landscape and areas of focus in the field. After refining our search criteria using the keywords "Advancements" OR "Challenges" AND "AI-Based Model" AND "Predictive Maintenance" AND "Water Injection Pumps" AND "Oil and Gas," and subsequently focusing the search to "AI-Based Model" AND "Predictive Maintenance" AND "Water Injection Pumps" AND "Oil and Gas," we successfully identified relevant literature. Furthermore, when we further narrowed our search by utilizing the keywords "Predictive Maintenance" AND "Water Injection Pumps," we uncovered a deeper layer of related research. This extensive body of related work has been meticulously reviewed and critiqued in this study.

The limited availability of research on PdM for WIPs in the OGI highlights a significant research gap. To date, no studies have been found that specifically investigate the prediction of maintenance requirements, considering causes such as equipment type, operating conditions, age of equipment, quality of maintenance, environmental factors, type of oil or gas, and safety considerations, as well as effects including wear and tear, corrosion, clogging, electrical issues, leaks, and performance improvements.

## 5 Water injection pump in the oil and gas industry sector

Water injection pumps are essential equipment in the OGI, used to increase pressure and enhance production by injecting water into wells. They play a crucial role in maintaining production levels and improving recovery rates [17]. Data mining techniques are utilized in the industry to address challenges and support decision-making, including the maintenance of production levels through methods like water injection. Lubricant analysis is crucial for monitoring pump condition and reliability [11–13]. While previous studies have utilized data mining techniques for lubricant analysis, they often rely on laboratory data. This research focuses on predicting the lubricant service life of WIPs using real-time field data from a crude oil production facility. It considers data pre-processing and the choice of regressor to assess prediction accuracy [18]. Data mining predictions' accuracy is influenced by factors such as algorithm choice, data volume, and data pre-processing [19]. This study focuses on the impact of data pre-processing on predicting lubricant service life for WIPs in a crude oil recovery facility. The dataset is divided into five categories based on different combinations of motor and pump data to identify factors affecting lubricant service life. Electric motor-driven WIPs in Sumatra, Indonesia, were examined using condition monitoring techniques like vibration monitoring, infrared scanning, and lubricating
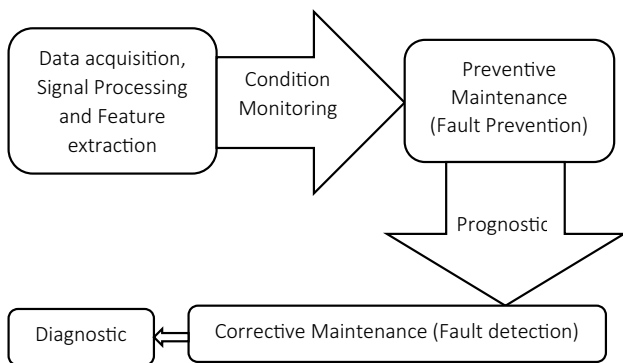
**Fig. 5** The lubricant sampling points of WIP are used to predict the service life of the oil [20]



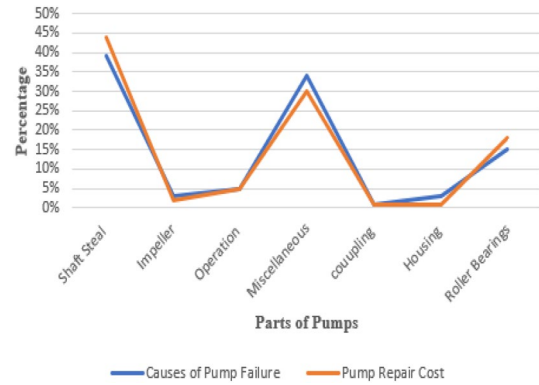**Fig. 7** Analysis of distribution in causes of failure and associated repair costs [23]

oil analysis. Various approaches exist for data pre-processing and lubricant service life prediction in the OGI, including ML, statistical analysis, expert systems, and physical modeling [7]. Statistical techniques like regression analysis can predict maintenance needs by modeling the relationship between lubricant condition and equipment performance. Expert systems provide guidance based on expert knowledge, and physical models utilize known equipment and lubricant properties for prediction. The choice of data pre-processing and prediction method depends on organizational needs and resources [20]. Regular lubricant analysis, depicted in Fig. 5, helps identify issues like contamination, degradation, or wear, enabling timely action to prevent equipment damage or failure.

## 5.1 Analysis and anticipation of pump failures

Prognostics constitutes a substantial area of inquiry due to its foundational role in advanced predictive technologies [21]. This field not only assists in appraising the performance of equipment but also plays a pivotal role in forecasting the timing of potential failures and mitigating the impact of unexpected breakdowns [21]. The central

focus of prognostics and health management (PHM) applications lies in the realm of prediction, encompassing the determination of a system's Remaining Useful Life (RUL) and the formulation of a contemporaneous maintenance strategy [21]. It is essential to distinguish between diagnosis and prognosis, with the pivotal distinction residing in their temporal orientation. Diagnosis is concerned with identifying the nature and cause of a specific issue, typically applied after a system failure has transpired [21]. In contrast, prognosis, rooted in the Greek concept denoting foreknowledge and fore-sensing, is geared toward predicting faults before they manifest. The primary aim of prognosis is to anticipate events before their occurrence, thereby emphasizing the critical element of timing in this context, as opposed to diagnosis [15–17]. Figure 6 provides an overview of Condition-Based Maintenance (CBM), encompassing diagnostic and prognostic maintenance, in accordance with the framework by Tchakoua et al. [22]. CBM is orchestrated through three fundamental processes, as illustrated in Fig. 7, which encompass data collection through sensors, signal processing utilizing diverse data techniques, and the extraction of features comprising characteristics facilitating the assessment of the equipment's current state [23].

Figure 6 demonstrates how information from the system's current and past statuses, drawn from the collected data, can be harnessed to identify or predict faults in pumps. Upon diagnosis of a defect, corrective maintenance is employed to rectify the issues. Conversely, if a failure is foreseen, preventative maintenance is executed in advance of the impending fault occurrence. It is noteworthy that faults in pumps can emanate from operational causes, system-related issues, mechanical malfunctions, or a combination thereof [20, 21]. Mechanical issues in pumps transpire due to problems with components like bent rotors, misalignments, and bearing complications. System-related flaws entail improper installation and



**Fig. 6** A summary of CBM [23]

leakage. Operational faults, primarily arising from obstructions, cavitation, and flow-related matters, manifest during the active operation of the pump [24].

Per Tiwari et al. [25] findings, cavitation, attributed to pressures falling below the vapor pressure at the suction, stands out as the most prevalent and recurrent issue encountered in pump systems, a consensus supported by several studies [21–23, 25]. In addition to cavitation, another frequently recurring challenge in pumps relates to damage occurring in the casing and impeller, which can be attributed to the pulsation of pressure linked to internal recirculation. Furthermore, blockages in the suction and discharge pipes caused by solid particulates and impurities within the pumped liquid represent a recurring issue [23–25]. It is worth noting that these faults can manifest at any stage of the pump's operation. A study conducted by Grundfos Research and Technology [26] provides a detailed breakdown of the root causes of pump failures, along with descriptions of the associated repair costs, as illustrated in Fig. 7. Historically, the process of diagnosing faults has predominantly relied on physical assessments of equipment health. This approach introduces labor-intensive procedures that can, in turn, affect the precision of fault diagnosis. At its core, fault diagnosis methodologies can be broadly categorized as either model-based or data-driven. Model-based fault diagnosis methods involve the utilization of a mathematical model representing the observed system [26]. By employing such a model, estimations of system or process outputs are generated and subsequently compared against actual process outputs to yield a residual signal or innovation. These residual signals, in turn, provide insights into potential fault conditions based on a comparison between the model-generated outputs and the actual system outputs [26].

### 5.2 Factors affecting the maintenance prediction of water injection pump

WIPs are commonly used in OGI operations to increase reservoir pressure and enhance oil recovery. Several factors can influence the maintenance needs of these pumps. The type of oil or gas being injected can impact corrosion resistance and maintenance frequency [13–16]. Operating conditions, including temperature, pressure, and flow rate, also affect maintenance requirements [8, 10–13]. Older pumps are generally more prone to failure and may need more frequent maintenance. Considering these factors is crucial for planning PdM and ensuring equipment reliability [7, 27]. Proper maintenance is essential for equipment reliability, operational efficiency, and overall performance [28]. Several problems with water pump injection systems in the OGI may require maintenance and repair. Some common issues are listed below [7, 27, 28].

(i).  *Wear and tear on mechanical components* Water pump injection systems contain various mechanical components such as bearings, seals, and impellers that may wear out over time and require replacement.

(ii).  *Corrosion* Water pump injection systems may be exposed to corrosive substances, such as saltwater, that can cause damage to the system over time.

(iii).  *Clogging* Water pump injection systems can become clogged with debris or scale, which can reduce the efficiency of the system and require cleaning or maintenance.

(iv).  *Electrical issues* Water pump injection systems may experience electrical problems such as short circuits or power surges, which can cause the system to fail.

(v).  *Leaks* Water pump injection systems may develop leaks in the piping, seals, or other components, which can cause the system to lose efficiency and require repair.

It is important to regularly maintain and service water pump injection systems to ensure that they operate at peak performance and to prevent costly repairs or downtime.

### 5.3 Critical discussion of previous studies

Numerous research studies have explored DL techniques in the realm of PdM. For instance, Janssens et al. [29] investigated the application of CNNs for monitoring the health of machines, leveraging infrared thermal images. Although the study demonstrates the potential of CNNs in detecting anomalies in rotating machinery, it lacks specific information about the types of machinery used, limiting the generalizability of the findings. An alternative strategy for PdM revolves around the utilization of artificial neural networks (ANNs) to predict failures. Sampaio et al. [30] proposed an ANN-based model to predict motor failure time, with an evaluation conducted on an AK-FN059 with a 12 cm cooling fan. However, the study lacks a comprehensive analysis of the ANN's performance and its comparison with other PdM methods [26–28]. While the authors mention the possibility of integrating fault diagnosis and prediction systems using ANNs for maintenance planning, no concrete results or insights are provided in this regard. Pertaining to data pre-processing and analysis, Bekar et al. [13] introduced an intelligent approach for PdM with a specific emphasis on machine motors. The study incorporated K-means clustering and principal component analysis (PCA) for data pre-processing. However, the research lacks detailed information about the specific motor types and data sources used, limiting the applicability of the findings.

Incorporating data from a multitude of sources stands as a pivotal element within the framework of PdM. Cheng et al. [11] focused on PdM in MEP systems for OGI facilities, combining ANNs and SVMs, but their study lacked specific MEP component details and thorough system assessment. In the realm of PdM, algorithm selection is crucial for accurate predictions. Falamarzi et al. [14] used ANN and SVR for tram track gauge deviation prediction but lacked comprehensive model performance analysis and comparative assessments. Susto et al. [31]. proposed a PdM system for epitaxy processes with linear regularization and Ridge regression but lacked equipment context and performance evaluation, using LiR, RF, and BN without comprehensive assessments. Praveenkumar et al. [32] diagnosed automobile gearbox faults using SVM, but both studies lacked detailed comparative assessments. Abu-Samah et al. [33] introduced proactive maintenance with Bayesian Networks (BN), monitored pump machine condition with MGGP but did not include comparisons with alternative techniques. Prytz et al. [27] used Random Forest (RF) for vehicle compressor PdM but lacked in-depth assessments regarding the incorporation of vehicle data for improved predictions. Biswal and Sabareesh [34] developed a test rig to examine condition monitoring aspects in WIPs by integrating ANNs into their methodology. Another relevant work by Susto and Beghi [35] presented a PdM system for epitaxy processes, utilizing SVM and k-Nearest Neighbors (k-NN) for predictive purposes. However, the study did not offer detailed discussions on the performance of the proposed system or comparisons with other PdM techniques. Durbhaka and Selvaraj [36] conducted an in-depth investigation into the realm of PdM specifically for wind turbine diagnostics. Their study revolved around the analysis of vibration signals and involved the application of k-NN, SVM, and k-means algorithms. Mathew et al. [37] introduced a regression kernel for prognostics with the application of SVM. Regrettably, the study did not provide specific information regarding the domain in which this approach was applied, nor did it offer comprehensive performance assessments. Kulkarni et al. [17] devised a PdM strategy tailored to pump systems, incorporating Random Forest (RF), with a particular emphasis on root-cause analysis for defects. In the realm of PdM, various machine learning algorithms like SVM, ANN, RF, and others have demonstrated their effectiveness [7, 38]. SVM is known for exceptional accuracy, making it a preferred choice for classification and regression tasks [30, 33]. Kumar et al. [39] developed a big data-driven framework for gas turbine CBM prediction with FURIA. Lasisi and Attoh-Okine [40], used LDA, SVM, and RF for track quality indexing. Amruthnath and Gupta explored early fault detection with unsupervised ML methods, suggesting broader research with more parameters and emphasizing

multiple ML methods [7]. Huuhtanen and Jung [28], focused on photovoltaic panel PdM with DL and CNN. Kolokas et al. [41] explored upstream industrial equipment fault prediction using DT, RF, NB-G, NB-B, and ANNs, highlighting the importance of equipment data. Luo et al. [41] proposed early fault detection in machine tools with innovative ML algorithms, particularly DL, centered around pump machines. Abdalla et al. [42] utilized XGBoosting for Electrical Submersible Pumps (ESPs) PdM but didn't provide specific performance discussions or comparisons with other PdM techniques. Table 1 offers a comprehensive overview of previous PdM studies, aiding in comparisons and identification of key findings in the field.

Implementing PdM programs in the oil and gas sector in developing countries faces challenges like limited skilled personnel, access to specialized sensors and software, infrastructure constraints, financial limitations, and cultural obstacles [27, 36]. In the Gulf Cooperation Council (GCC) countries, which heavily rely on oil and gas production, challenges include overdependence on oil, environmental impacts, social issues, and infrastructure limitations [48]. Efforts are being made to address these challenges through initiatives like reducing carbon emissions, diversifying the economy, and improving working conditions [48]. Maintenance challenges in the GCC countries include aging infrastructure, harsh climates, skill shortages, limited access to modern technologies, and financial constraints. Initiatives and policies are being implemented to improve sustainability, transparency, and diversification in the oil and gas sectors, along with investments in training and infrastructure development.

## 6 Applications of artificial intelligence in predictive maintenance

AI-based model has gained significant attention and demonstrated remarkable potential in the field of PdM. This section provides an overview of AI-based model for PdM, highlights relevant case studies and research papers that focus on AI in PdM, and emphasizes the importance of AI in the context of PdM for WIPs [6]. AI-based model for PdM leverages the power of AI techniques, such as ML, DL, and data analytics, to enable more accurate and efficient prediction of equipment failures [14]. These models utilize historical and real-time data from sensors, maintenance records, and other relevant sources to identify patterns, detect anomalies, and forecast the RUL of equipment. ML algorithms, including regression models, decision trees, and SVM, can be trained on historical data to predict the likelihood of equipment failures based on various parameters and environmental conditions [6–10]. These algorithms learn from the data patterns and adjust their

**Table 1** Summary of previous studies in PdM using AI-based models

| References | Algorithms used | Equipment | Future work |
|---|---|---|---|
| [29] | DL models like CNNs | Rotating machinery | The presented approach exhibits promising potential in various domains such as offshore manufacturing lines for online condition monitoring. Future research should focus on investigating the underlying physical phenomena associated with highlighted image regions and enhancing the interpretability of CNN-based models |
| [30] | ANN | AK-FN059 with 12 cm cooling fan | An evaluation of the feasibility of integrating fault diagnosis and prediction systems using ANNs is required, with the ultimate goal of developing a unified system for maintenance planning in the industry |
| [13] | K-means clustering and principal component analysis (PCA) | Machine motor | Additional efforts entail examining a multitude of measurements sourced from diverse data channels, particularly sensors, in order to establish connections between these measurements and systems related to production monitoring or maintenance |
| [11] | ANN and SVM | Building facilities | Future research will explore an ontology-based approach to develop a standardized data integration solution for different sensor types and application systems in PdM. This will facilitate improved data modeling and enhance the overall effectiveness of PdM systems |
| [14] | ANN and SVM | Rail-Tram track, 250 km of double tracks | Future research directions include exploring hybrid models such as combining ANN and SVM or integrating Genetic Algorithms with SVM. Additionally, expanding the analysis to different datasets will validate the effectiveness of the models in various tram systems |
| [31] | Linear regularization and ridge regression | – | Feature selection methods can be employed to identify the most informative features, reducing dimensionality, and improving computational efficiency. Future work should focus on applying these methods to enhance prediction accuracy |
| [43] | SVM | Rail network | A novel approach utilizing ML is presented with the aim of enhancing the velocity of rail networks through the application of PdM |
| [32] | SVM | Automobile gearbox | In the context of fault diagnosis for automobile gearboxes, ML techniques, particularly SVM, are employed to achieve accurate and effective results |
| [33] | BN | – | A Bayesian approach is employed for failure prediction in order to enhance proactive maintenance strategies |
| [27] | RF | Air compressors in trucks and buses | The achievement of PdM for vehicle compressor repairs, utilizing maintenance records and logged vehicle data, is made possible through the application of RF models |
| [35] | SVM and k-nearest neighbors (k-NN) | Tungsten filament | A system for PdM is devised to cater to epitaxy processes, employing filtering and prediction techniques that rely on SVM and k-Nearest Neighbors (k-NN) algorithms |

**Table 1** (continued)

| References | Algorithms used | Equipment | Future work |
|---|---|---|---|
| [36] | k-Nearest neighbors (k-NN), SVM, and k-means clustering | Bearing | Vibration signal analysis using collaborative recommendation approaches is employed for PdM in wind turbine diagnostics, combining k-NN, SVM, and k-means clustering techniques |
| [37] | SVM | – | Regression kernel for prognostics using SVM is investigated for improving PdM capabilities |
| [44] | RF | – | The generation of new data sets for use in PdM is proposed to enable comparisons and evaluations of different PdM approaches |
| [45] | PCA, Hierarchical clustering, k-means, Fuzzy C-means, and model-based clustering | Exhaust fan | Further expansion of this research can involve the integration of supplementary physics-based parameters and exploration of their potential combinations. Additionally, it is crucial to assess the detection accuracy by employing larger sample sizes and incorporating multiple fault states into the evaluation process |
| [46] | Decision trees (DT), RF, Naive Bayes (NB-G, NB-B), and ANN | Industrial equipment | Improved performance can be achieved by incorporating equipment data, such as status bits, utility pressure, alert bits, liquid flow rates, etc |
| [40] | CNNs | – | A Cognitive Acoustic Analytics Service for the IoT is created, employing models based on CNNs. This service utilizes advanced DL techniques to analyze acoustic data in an intelligent and cognitive manner, enabling valuable insights and decision-making in IoT applications |
| [7] | Linear discriminant analysis (LDA), SVM, and RF | Sample mile track | A ML approach utilizing principal component analysis (PCA) and a track quality index is employed for PdM in a sample mile track |
| [47] | PCA, hierarchical clustering, k-means, Fuzzy C-means, and model-based clustering | Exhaust fan | This research has the potential to extend its scope by incorporating additional physics-based parameters and exploring their combinations. Furthermore, it would be intriguing to assess the detection accuracy using larger sample sizes and multiple fault states |
| [42] | Decision Trees (DT), RF, Naive Bayes (NB-G, NB-B), and ANN | Industrial equipment | By integrating equipment data such as status bits, utility pressure, alert bits, liquid flow rates, and more, it is possible to enhance performance and achieve improved results |

predictions as new information becomes available. DL techniques, such as CNNs and recurrent neural networks (RNNs), are capable of handling complex and unstructured data. They excel in feature extraction and can capture intricate relationships within the data, enabling more accurate fault diagnosis and prediction of failures [23, 25–27]. AI-based model for PdM offers several advantages, including early detection of equipment degradation, reduced downtime, optimized maintenance schedules, and improved cost-efficiency. By accurately predicting failures and identifying maintenance needs in advance, these models empower maintenance teams to take proactive measures and prevent costly breakdowns [13–18]. For WIPs in the OGI, specific case studies and research papers have explored the application of AI in PdM [3, 10]. These studies have emphasized the use of AI algorithms, such as ML and DL, to analyze sensor data, monitor pump performance, and detect potential faults or deviations. These case studies and research papers showcase the practical implementation of AI-based model in WIPs ' PdM. They provide valuable insights into the benefits, challenges, and lessons learned from using AI techniques in this specific context [6–8].

### 6.1 Barriers to using ML for predictive maintenance

The organization of this literature review is structured around the ML techniques and categories utilized, the equipment and devices employed for data acquisition, the attributes of the utilized data (including size and type), and various other factors. This review shows that PdM plays a vital role in enhancing efficiency within environments where machines undergo gradual deterioration and wear over time. The proliferation of inexpensive, interconnected sensors within the IoT enables a growing abundance of data for ML algorithms to facilitate PdM [13]. The study provides an extensive examination of ML techniques employed in recent years for PdM of industrial components [36]. The review explores potential market prospects for PdM and the challenges associated with implementing ML algorithms for PdM in the context of Industry 4.0, as delineated in Table 2. A survey by Prytz et al. [27] found that only 11% of companies surveyed had successfully implemented ML-based PdM**.**

### 6.2 Importance of AI in water injection pumps' predictive maintenance

By employing AI-based model for PdM, operators can proactively identify potential issues, monitor pump performance, and take preventive measures to avoid costly downtime [6]. AI algorithms can analyze sensor data in real-time, detect abnormal pump behavior, and provide early warnings for impending failures [13]. This enables maintenance teams to schedule maintenance activities, order spare parts, and allocate resources in a timely manner, minimizing disruptions and maximizing pump availability [24]. By leveraging AI in WIPs ' PdM, operators can optimize maintenance strategies, reduce unnecessary inspections or repairs, and enhance the overall reliability and productivity of their operations. In conclusion, the application of AI in PdM offers significant advantages for WIPs in the OGI [45]. AI-based model enables accurate failure prediction, optimized maintenance schedules,

**Table 2** Barriers to using ML for PdM

| References | Hurdles | Reflections |
|---|---|---|
| [49] | Determining the necessary data to gather | The launch of connected machines can be difficult for several reasons: |
| | | (1) Unclear evidence of data that provide value: It can be difficult to know which data is valuable when launching connected machines, as it is not always clear how the data is used or what insights they would provide. This can make it challenging to justify the time and resources required to collect and analyze the data |
| | | (2) Unclear business goal and planning: Without clear business goals, it is difficult to determine which data are necessary to collect and how they should be used to inform business decisions. Without a plan for how to use the data, it is also difficult to know how to collect and analyze the data |
| | | (3) Limited resources: Launching connected machines often requires significant time and resources. If one is unsure about which data is valuable or how to use them, it can be challenging to justify the use of these resources |
| | | (4) Complexity: Connected machines can be complex systems, and it can be challenging to effectively collect and analyze the data they generate |
| [11] | Collecting necessary datasets | Due to large dataset requirements, data preparation and computational expenses, establishing ML solutions can be time-consuming and resource intensive |
| [44] | Optimized data science | Selecting an appropriate method of analyzing data involves choosing the right approach to extract meaningful insights, such as statistical techniques, data visualization, or ML algorithms |

and improved cost-efficiency, ensuring the reliable and efficient operation of water injection systems [35].

### 6.3 AI models for predictive maintenance of water injection pumps

In recent years, a range of AI models have been employed for PdM of WIPs in the OGI. These models utilize advanced techniques such as ML, DL, and hybrid approaches to enable accurate prediction of pump failures, estimation of RUL, and optimization of maintenance strategies [36]. This section provides an overview of the AI models commonly used in water injection pump maintenance, highlighting their key characteristics, functionalities, and advantages. ML-based models, including regression models, decision trees, SVM, RF, and ANN, have demonstrated success in predicting pump failures based on historical sensor data, operational parameters, and maintenance records [10–14]. These models employ various algorithms to learn from the data patterns and identify potential failure modes. They can estimate the probability of pump failures, classify fault types, and optimize maintenance schedules to minimize downtime [27, 33]. DL-based models, such as CNNs and RNNs, have shown remarkable performance in analyzing sensor data and capturing complex patterns in water injection pump operation. CNNs excel in image-based fault detection by extracting spatial features from pump component images or spectrograms [45]. RNNs, on the other hand, are effective in analyzing sequential sensor data and capturing temporal dependencies for accurate failure prediction [8, 43]. These DL models enable more accurate fault diagnosis and prediction, leading to improved maintenance decision-making. Hybrid models, which integrate ML algorithms with DL architectures, have been developed to leverage the strengths of both approaches. These models combine feature extraction capabilities of CNNs with sequence modeling capabilities of RNNs, allowing for comprehensive analysis of both spatial and temporal aspects of water injection pump operation [33–36].

### 6.4 Data collection and processing

Effective data collection and processing are crucial components of AI-based model for PdM. This section discusses the data sources and collection methods commonly used in PdM, the preprocessing techniques employed for AI models, and the associated challenges and considerations in data collection and processing [34]. To effectively implement PdM programs, organizations rely on a range of data sources to gather the requisite information for predictive analytics and ML models. Among these sources, sensor data plays a pivotal role. Sensors are strategically placed on equipment to collect real-time data on parameters such as temperature, vibration, pressure, and flow. For instance, in the case of wind turbines, vibration sensors are employed to monitor the condition of bearings and gearboxes, facilitating the early detection of potential faults [13]. Another crucial data source is Internet of Things (IoT) devices. IoT devices are becoming increasingly prevalent for data collection in PdM [36]. IoT devices, when integrated into machinery and equipment, offer continuous data streams, enabling remote monitoring and centralized issue detection [27]. Historical maintenance logs provide valuable insights into past equipment failures and maintenance activities, aiding in predicting future maintenance needs. Environmental data, including temperature, humidity, and weather conditions, can impact equipment health; correlating this data with performance enhances PdM model accuracy [14]. PdM utilizes various data collection methods, such as continuous condition monitoring for real-time sensor data, particularly beneficial for early anomaly detection in critical assets like engines and turbines [13]. IoT technology simplifies remote monitoring by gathering data from diverse locations, facilitating wide geographical equipment failure management [7]. Data historians represent specialized systems that collect and store historical data for analysis. They are commonly employed in industries such as oil and gas, where significant data volumes are generated. Data historians provide a historical perspective on equipment performance, aiding in the development of effective PdM strategies [50]. ML and AI techniques play a crucial role in analyzing and making predictions based on the collected data. These algorithms are capable of identifying patterns, anomalies, and trends that may not be immediately apparent through manual analysis. Consequently, they enhance the accuracy of PdM, contributing to reduced downtime and cost savings [14].

### 6.5 Data preprocessing techniques for AI models

Data preprocessing plays a pivotal role in preparing data for AI models, encompassing crucial tasks such as data cleaning, transformation, and organization [38]. This step is indispensable for enhancing data quality, consistency, and suitability, ensuring that the data is primed for effective analysis. Several preprocessing techniques are commonly employed in the context of AI-based model for PdM [27]. Data cleaning involves identifying and correcting or removing errors, outliers, missing values, and noise from the data. This step ensures that the data used for analysis is accurate and reliable. Data normalization is often applied to scale the data to a consistent range, enabling fair comparison and preventing the dominance of certain features. Common normalization techniques include min–max scaling, z-score normalization,

or logarithmic scaling. Data integration, as performed by combining data from multiple sources or sensors, ensures a comprehensive view of equipment condition, enhancing predictive model accuracy [7]. Hybrid AI models, achieved through integrating various AI techniques, improve PdM system performance. Accurate evaluation and validation of AI models are crucial for assessing their reliability in PdM of WIPs.

## 6.6 Challenges and considerations in data collection and processing

Data collection and processing for PdM pose several challenges and require careful considerations [34, 44]. Some of the key challenges include:

## 7 Data Availability and Accessibility

(1) Obtaining access to high-quality and relevant data can be a challenge, especially in cases where data is dispersed across different systems or not readily available. Collaboration with data owners, data sharing agreements, and data integration techniques may be necessary to address these challenges.

(2) *Data Volume and Velocity* PdM generates large volumes of data, particularly in real-time monitoring scenarios. Handling and processing such high-velocity and high-volume data require efficient storage and computational infrastructure.

(3) *Data Quality and Consistency* Ensuring data quality and consistency is crucial for reliable predictions. Data cleaning techniques need to be applied to address errors, outliers, and missing values. Additionally, consistency in data collection across different sensors and systems is essential to maintain accuracy and reliability.

(4) *Data Privacy and Security* Protecting sensitive data and ensuring compliance with privacy regulations are critical considerations. Proper anonymization techniques and data governance practices must be implemented to safeguard the privacy and security of collected data.

(5) Domain expertise plays a crucial role in comprehending the collected data and contextualizing equipment behavior. Effectively interpreting this data necessitates a deep understanding of the specific domain. Collaborative efforts between data scientists and subject-matter experts are vital to guarantee precise analysis and interpretation of the data, leveraging their respective knowledge and skills.

## 8 Role of artificial intelligence on predictive maintenance in the oil and gas industry

It is increasingly applied in the OGI for PdM programs. There are several ways in which AI can enhance PdM in the OGI [13–19]:

(i). *Predictive modeling* AI algorithms can be utilized to create predictive models that analyze data on equipment performance and identify potential issues before they occur.

(ii). *Real-time monitoring* AI-powered sensors and monitoring systems can be used to collect data on equipment performance in real time, allowing for the early identification of potential issues.

(iii). *Machine learning* AI algorithms can learn from data over time, allowing them to improve their predictive capabilities and become more accurate at identifying maintenance needs.

(iv). *Automated maintenance scheduling* AI can be used to automate the scheduling of maintenance, ensuring that it is performed at the optimal time to minimize downtime and optimize equipment performance.

Overall, the use of AI in PdM can help oil and gas companies improve safety, efficiency, and cost-effectiveness. By identifying and addressing potential issues before they occur, AI can reduce downtime and improve overall production efficiency.

## 8.1 Performance of predictive maintenance in the oil and gas industry

Predictive Maintenance is a crucial technique in the OGI, anticipating equipment failures and proactively addressing maintenance needs [7–23, 25–35]. It offers benefits such as increased operational efficiency, improved safety, cost savings, and environmental advantages [17, 29]. The theories guiding PdM include "wearout failure," addressing gradual equipment degradation [13], and "infant mortality," mitigating failures in new equipment [6]. Reliability engineering emphasizes designing reliable systems and conducting regular inspections and tests [33]. Risk-based maintenance prioritizes tasks based on potential impacts, utilizing techniques like failure mode and effects analysis [19]. The goal of PdM is to minimize equipment failure risk, optimize operations, reduce costs, and improve production through PdM, CBM and ML techniques [28, 30].

PdM, utilizing historical data and ML tools, enables timely fault detection and enhanced equipment condition [3]. The primary goal of every maintenance strategy is to minimize failure rates, improve equipment conditions,
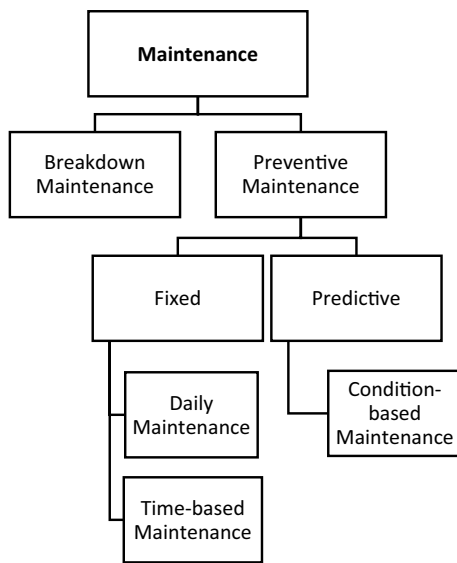
**Fig. 8** Maintenance types adopted from [28]

extend lifespan, and reduce costs. PdM is a highly promising approach that possesses significant potential for achieving these objectives [33]. However, it is crucial to carefully choose ML techniques, data types, and sizes that are practical for industrial applications. This review provides guidance for researchers and practitioners in selecting the most suitable ML techniques, data sizes, and types [13, 43]. Figure 8 illustrates the different maintenance types in two scenarios: when the system is well-planned and when it is not well-planned. ML, a subset of AI, is a powerful tool for creating predictive algorithms by training models on historical data to forecast future outcomes. ML demonstrates exceptional proficiency in managing high-dimensional multivariate data and uncovering

concealed relationships within intricate environments [51]. ML techniques have widespread applications across diverse manufacturing domains, encompassing maintenance, optimization, troubleshooting, and control [32]. This review provides an overview of recent advancements in ML techniques for PdM, categorizing them based on ML technique, category, equipment, data acquisition, description, size, and type. The review is based on a comprehensive search of the Scopus database, covering the current field, background, methodology, and a thorough review of ML techniques for PdM [7]. It also offers conclusions and future research guidelines. PdM utilizes AI and IT technology to maintain equipment safety through defect detection and RUL estimates, reducing costs and extending equipment lifespan.

PHM systems have become necessary for effective equipment maintenance. PHM systems, utilizing advancements in AI and IT, assess equipment health systematically [6]. PdM, a PHM approach, collects data on equipment's physical health and performance (e.g., pressure, vibration, temperature) to detect faults early, assess health, and predict future states. This approach reduces maintenance costs and extends the equipment's RUL by avoiding potential failures [51]. ML, a subfield of AI, learns and improves without explicit instructions, finding applications in various fields, including manufacturing for tasks like maintenance, optimization, and control [51]. PdM implementation typically encompasses a diverse array of technologies, including smart sensors, networks, AI, Big Data, and cloud systems. Figure 9 visually represents these technologies, outlining the process and components of PdM. The technologies can be categorized into sensors, networks, integration, augmented intelligence, and augmented behavior. Sensors play a crucial role in collecting data pertaining to machine operations and environmental

**Fig. 9** PdM process and technologies to drive maintenance [49]
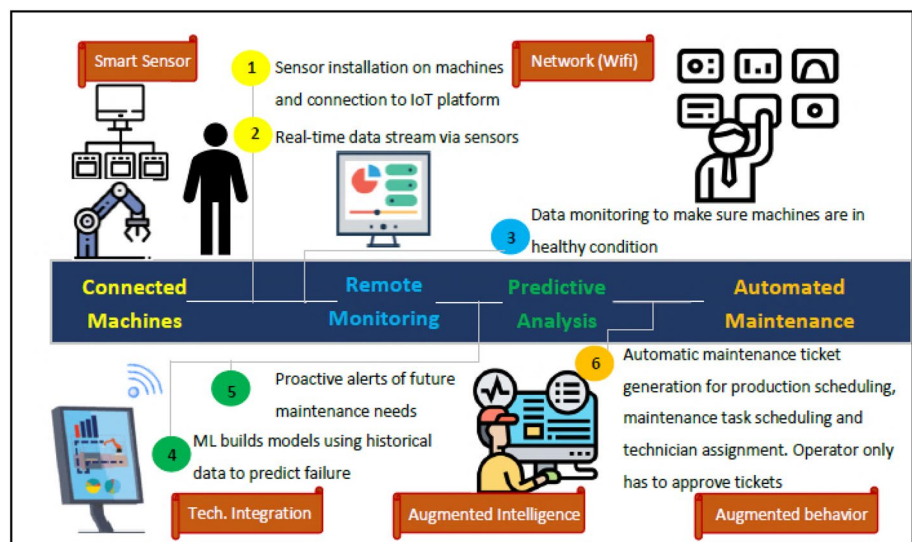
conditions, while the network facilitates data storage and transfer, utilizing technologies such as Bluetooth and WiFi. Effective data management and consolidation are made possible through the integration of various technologies, facilitated by IoT [13]. Augmented intelligence plays a pivotal role in data processing and analysis, enhancing the overall capabilities of the system [49]. Augmented behavior, enabled by applications and ticketing systems, offers virtualization, computing, and service platforms to support operators in their tasks. Furthermore, ML algorithms are categorized into supervised, unsupervised, and reinforcement learning methods, providing distinct approaches for data analysis and decision-making [49]. By combining and applying these algorithms, their classification power can be enhanced in both supervised and unsupervised learning scenarios.

The main categories of ML algorithms consist of supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, algorithms leverage labeled data to make predictions and identify patterns. Unsupervised learning algorithms, on the other hand, work with unlabeled data to discover inherent patterns and relationships. Lastly, reinforcement learning algorithms involve an agent interacting with an environment to learn a sequence of actions that maximize a reward or outcome [3–7]. Some algorithms can be applied to both



**Fig. 10** Classifications within ML techniques [52]

unsupervised and supervised learning, and different algorithms can be combined to improve classification accuracy [38, 45]. Unsupervised ML is a type of AI in which the algorithm uses existing data to identify patterns or clusters without guidance or feedback from an external expert. This is different from supervised learning, which involves providing the algorithm with labeled examples to learn from. Unsupervised learning encompasses a range of techniques, including clustering, self-organizing maps, and association rules. In the presented study, the reviewed articles are categorized into three main ML categories: classification, regression, and clustering, as illustrated in Fig. 10.

This categorization allows for a comprehensive understanding of the ML techniques employed across the analyzed articles. The data employed in These articles can be categorized into two primary types: firstly, real data acquired from genuine real-world sources, and secondly, simulated, or synthetic data generated with specific objectives in mind [13]. Reinforcement learning, on the other hand, represents a particular type of ML that entails providing the algorithm with information regarding the outcomes resulting from specific actions. Through a process of trial and error, the algorithm learns which actions yield the most favorable results, thus optimizing its decision-making capabilities. Some researchers consider reinforcement learning to be a special form of supervised learning, while others see it as distinct [27]. There are several different algorithms that can be used for supervising ML in PdM or manufacturing, each with its own strengths and limitations. Selecting the best algorithm for a specific problem can be difficult. To become proficient in applied ML, it is important to gain experience working with a variety of datasets [29]. Different approaches, data preparation, and modeling methods may be necessary for different problems. Datasets used in ML can be divided into two categories: real datasets collected from actual production processes and synthetic datasets created for ML.
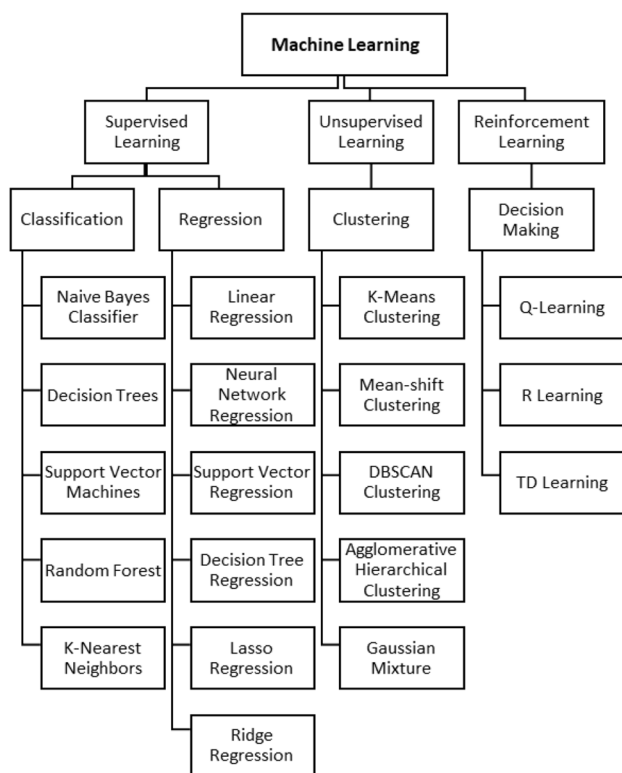
## 8.2 Applications of ML algorithms in predictive maintenance

ML algorithms play a pervasive role in the field of PdM, finding extensive utilization in diverse areas such as manufacturing systems, tools, and machines. This stems from their remarkable capability to analyze vast volumes of data and address a wide range of problems [19]. Within PdM, a multitude of ML algorithms are frequently employed, including but not limited to ANN, reinforcement learning, SVM, logistic regression, and decision trees [7]. Developing an ML algorithm involves historical data selection and pre-processing, model choice, training and validation, and ongoing performance maintenance [27]. ANNs, inspired by

biological neurons, excel in handling complex, nonlinear data and have been applied in soft sensing and predictive control systems [14]. They demonstrated superior performance in classifying tool state and predicting equipment failures compared to SVM and K-Nearest Neighbors (KNN) models [14, 50]. In rail track PdM, SVMs and ANNs are highly effective in predicting degradation and distinguishing healthy from faulty states [7, 37]. While physics-based models are adept at representing degraded conditions, ANN models perform strongly for equipment in optimal condition [27]. SVM, renowned for its exceptional accuracy in classification and regression, adapts and learns from data, excelling in pattern recognition, classification, and regression. It is successfully used in PdM for equipment status identification. In a study [14], both SVM and ANN models predicted gauge degradation for rail tracks, with SVM slightly outperforming ANN, particularly for curved segments. Xiang et al. [53] achieved over 80% accuracy in diagnosing and predicting machine conditions using SVM in a data-driven framework. Decision Trees, a versatile ML algorithm, create tree-like models for classification and regression tasks, processing large datasets, handling missing values, and working with categorical data. Their interpretability and visualizability make them effective for explaining predictions. RF, an ensemble ML method, integrates multiple decision trees to make predictions and excels in handling high-dimensional data and missing values [13]. In the realm of PdM, RF is highly favored for its capacity to manage large datasets, missing data, and nonlinear relationships. It predicts the RUL of machines by learning from present condition and past performance data, making it a reliable method for industrial equipment PdM [11–21]. SVM surpasses traditional ML algorithms and excels in tasks such as modeling machinery condition, detecting drifting behavior in continuous data streams, and diagnosing failures in rotating machines and aircraft maintenance systems [31]. RF combines predictions from multiple decision trees and is versatile for both classification and regression, known for its resistance to overfitting. Logistic Regression (LoR), a supervised algorithm,

estimates the probability of an event based on independent variables and is applied in PdM to gauge the likelihood of equipment failure, considering maintenance records, operating conditions, and environmental factors. LiR is particularly useful in the OGI due to its ability to handle large datasets, provide accurate predictions, and integrate seamlessly into existing maintenance systems [49]. In a study comparing LiR with RF and XGBoost, all three algorithms performed well in predicting machine downtime, with RF and XGBoost showing slightly better performance in decision thresholds [54, 55]. Note that Linear Regression (LiR) differs from Logistic Regression (LoR) in terms of their application domains. LiR is employed to predict continuous dependent variables, whereas LoR is specifically designed to handle binary dependent variables. Further details regarding this distinction can be found in Fig. 11.

XGBoost offers potent PdM capabilities. However, its complexity, data requirements, computational resources, lack of interpretability, model maintenance, and regulatory considerations present critical challenges. Successful implementation demands careful planning, expert knowledge, and attention to data quality and compliance [54, 55]. Its exceptional performance has been evident in
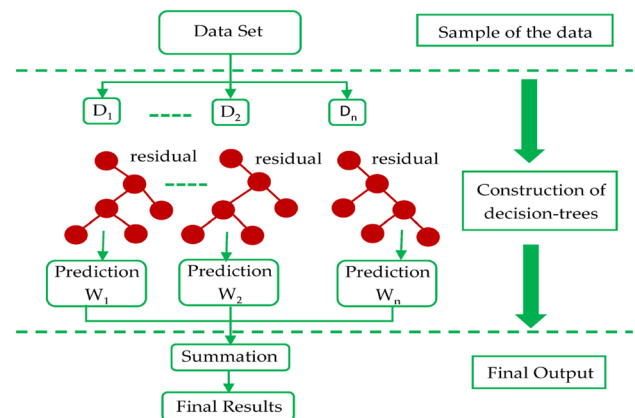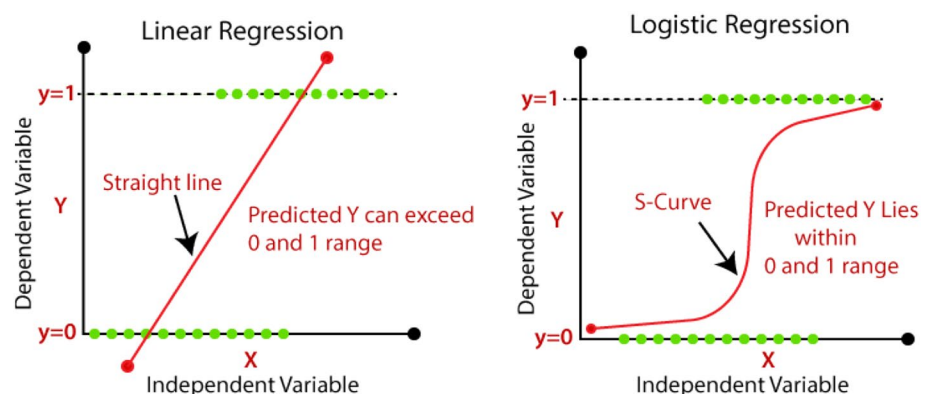


**Fig. 12** XGBoost algorithm tree [9, 54]



**Fig. 11** Logistic regression [9, 54]

numerous studies, including its ability to forecast machine downtime and predict maintenance requirements. XGBoost, alongside RF and Logistic Regression, exhibited outstanding results in receiver operating characteristic curves (ROC), with XGBoost and RF outperforming others in decision thresholds. The versatility and effectiveness of XGBoost render it an invaluable asset for enhancing maintenance processes and minimizing industry downtime [54, 55]. Figure 12 provides a visual representation of the tree structure within the XGBoost algorithm.

The Gradient Boosting Machine (GBM) is a powerful ensemble learning method used in PdM in the OGI. It can handle large datasets with high dimensionality and non-linear relationships, achieving high accuracy. GBM has been successfully applied in predicting equipment failures and estimating the RUL of components. In a study [27], GBM accurately predicted RUL with less than 5% mean absolute error. XGBoost, a specific implementation of GBM, is widely used and has achieved state-of-the-art results in various applications [13]. LiR is a widely employed statistical technique utilized for modeling the linear correlation between dependent and independent variables. It serves as a fundamental method for establishing the statistical relationship between these variables. In the OGI, it can be used to predict equipment failure or performance degradation based on factors such as operating conditions and maintenance history [19]. While LiR is simple and efficient, it may not capture non-linear relationships. In a study by Abbasi et al. [56], LiR, RF regression, and symbolic regression were used for modeling machinery condition and predicting drifting behavior (Fig. 13). The effectiveness of this approach was demonstrated through a real-world case study conducted on industrial radial fans [56]. The study validated the applicability and success of the approach in an industrial setting, specifically in the context of radial fans.

Symbolic regression is a ML technique used to find a mathematical equation that describes a given dataset. In the OGI, it can be used to predict the performance of wells based on input variables. For example, it can predict oil or gas production based on factors like well depth, rock composition, and pressure [36]. Symbolic regression, which utilizes mathematical symbols and functions to create a syntax tree, has shown promise in forecasting machinery conditions and identifying shifts in concepts [50]. Janssens et al. [29] leveraged Convolutional Neural Networks (CNNs) for monitoring machine health using infrared thermal images, outperforming conventional methods in fault detection and oil level prediction. In another study [27], CNNs excelled in PdM for electrical faults in pumps by accurately estimating power curves. An IoT-based cognitive acoustics analytics service exhibited exceptional performance when analyzing acoustic data in a separate research endeavor [43]. Another research effort [33] employed ML techniques to predict RUL of a machine process, achieving notable success. Moreover, in a separate study [13], PdM techniques were implemented in conjunction with a digital twin to actively prevent faults in machine tools, showcasing a proactive approach to maintenance.

## 9 Factors affecting predictive maintenance in the oil and gas industry

There are several factors that can affect the effectiveness of PdM for oil and gas operations. Some of these are described below [57].

### 9.1 Type of equipment

The type of equipment used in oil and gas operations can significantly impact the maintenance needs and the effectiveness of PdM. Different types of equipment may have different maintenance requirements and may be more or less susceptible to certain types of failure [58]. For example, some equipment may have complex systems that require more frequent maintenance, while other equipment may be simpler and require less maintenance [33–35]. The type of materials used in the construction of the equipment can also affect its maintenance needs. For example, certain materials may be more resistant to wear and tear, which could extend the lifespan of the equipment and reduce the need for maintenance [59].

### 9.2 Operating conditions

Operating conditions can significantly impact the maintenance needs and the effectiveness of PdM in oil and gas operations. The operating conditions of the equipment, such as temperature, pressure, and vibration, can affect its reliability and maintenance needs. For example,
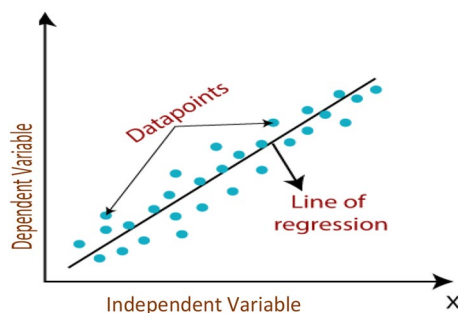


**Fig. 13** LiR in ML [56]

equipment that operates at high temperatures or pressures may require more frequent maintenance due to the increased wear and tear on the equipment [43]. Similarly, equipment that is subjected to high levels of vibration may require more frequent maintenance to prevent failure. Operating conditions play a significant role in maintenance, especially in challenging environments like offshore platforms or remote facilities [60]. These conditions affect maintenance frequency, types, and personnel safety. To ensure equipment reliability and safety in oil and gas operations, it's crucial to tailor PdM strategies to specific operating conditions. Doing so optimizes equipment effectiveness and longevity, enhancing overall reliability and safety [14].

### 9.3  Age of equipment

The age of the equipment used in OGI operations can significantly impact the maintenance needs and the effectiveness of PdM. Generally, older equipment is more prone to failure and may require more frequent maintenance. As equipment ages, it is subjected to wear and tear from regular use, which can lead to deterioration and an increased risk of failure [50]. In addition, older equipment may not have the same level of technological advancements as newer equipment, which can make it more prone to failure and more difficult to maintain. PdM is typically more effective for newer equipment, as there are greater amounts of data available about the equipment's performance and maintenance needs. For older equipment, it may be more challenging to predict when maintenance is needed, as there may be less data available, and the equipment may be more prone to unexpected failures [7]. Overall, the age of the equipment is an important factor to consider when planning PdM in OGI operations. It is generally more effective to prioritize maintenance for older equipment in order to prevent unexpected failures and extend the lifespan of the equipment.

### 9.4  Quality of maintenance

The quality of maintenance is an important factor that can impact the effectiveness of PdM in OGI operations. Properly maintaining equipment can help extend its lifespan and reduce the need for unexpected repairs. Poor quality maintenance can lead to equipment failure and unexpected downtime, which can be costly and disruptive to operations [25]. On the other hand, high-quality maintenance can help prevent equipment failure and improve the reliability and performance of the equipment. There are several factors that can affect the quality of maintenance in OGI operations, including the skill and training of the maintenance personnel, the availability of high-quality

spare parts and tools, and the use of best practices and procedures [61]. To ensure the quality of maintenance in OGI operations, it is important to invest in the training and development of maintenance personnel, use high-quality spare parts and tools, and follow established best practices and procedures [44]. Implementing these approaches can contribute to enhancing the dependability and efficiency of equipment, thereby leading to improved overall performance in OGI operations.

### 9.5  Environmental factors

Maintenance requirements and the efficacy of PdM in operations within OGI can be substantially influenced by environmental factors. These factors exert a noteworthy impact on the maintenance demands of equipment and the overall effectiveness of PdM practices within OGI operations [62]. It is imperative to thoroughly consider these environmental factors to ensure that PdM strategies are suitably tailored to the specific conditions, enabling optimal maintenance outcomes in the OGI. The environment in which the equipment is operating, such as offshore platforms or land-based facilities, can impact its maintenance needs. For example, equipment that is exposed to harsh weather conditions, such as extreme heat, cold, or saltwater, may require more frequent maintenance due to the increased wear and tear on the equipment [27]. Likewise, machinery operating in environments prone to corrosion, such as offshore platforms, may necessitate more frequent maintenance interventions to safeguard against corrosion and uphold the equipment's safety and reliability. Apart from directly influencing the equipment, the operational environment also exerts an impact on the maintenance procedures themselves [18]. It is crucial to consider these environmental factors comprehensively, as they significantly affect both the maintenance requirements and the execution of maintenance activities, contributing to the overall effectiveness of the maintenance process. For example, if the equipment is located in a hazardous or difficult-to-access environment, such as offshore platforms or remote land-based facilities, it may be more challenging to perform maintenance [7]. This can impact the frequency and types of maintenance that can be carried out, as well as the safety of personnel carrying out the maintenance. Overall, it is important to consider the environmental factors that may impact the maintenance needs of the equipment when planning PdM in OGI operations.

### 9.6  Type of oil or gas

The type of oil or gas being processed can impact the wear and tear on equipment and may require different types of maintenance in OGI operations. Different types of OGI may

have different physical and chemical properties, which can affect the equipment that is used to process them [63]. For example, some types of oil or gas may be more corrosive than others, which can impact the corrosion resistance of the equipment and the frequency of maintenance required [46]. Similarly, some types of oil or gas may be more viscous, which can affect other equipment used to process the oil or gas. In order to effectively plan PdM in OGI operations, it is important to consider the specific properties of the oil or gas being processed and the impact it may have on the equipment [50]. This can help to ensure the reliability and performance of the equipment and the overall efficiency of the OGI operations.

### 9.7 Safety considerations

PdM is a maintenance approach characterized by its reliance on data and analytics to determine when maintenance should be performed on equipment. Unlike traditional maintenance practices that adhere to fixed schedules or wait for equipment failure, PdM leverages data-driven insights to proactively identify maintenance needs [37]. By continuously monitoring and analyzing equipment performance, PdM aims to detect potential issues early on, allowing for timely and targeted maintenance interventions [62]. This data-centric approach enables organizations to optimize maintenance efforts, minimize downtime, and enhance equipment reliability and performance. In the OGI sector, it is important to prioritize safety in all maintenance activities due to the hazardous nature of the work environment [6, 7]. Safety is paramount in PdM implementation, ensuring personnel well-being and equipment integrity. Robust safety measures mitigate risks and foster a secure working environment. This includes ensuring that all personnel are properly trained and equipped to perform maintenance tasks safely, and that the work environment is safe and secure [27]. Additionally, the safety of the equipment itself must be considered, as faulty or malfunctioning equipment can pose a serious risk to workers and the surrounding environment. Overall, safety considerations are critical to the success of any maintenance program and must be given the highest priority in the OGI sector.

## 10 Challenges and future directions

This study highlights the importance of digitizing oil fields to improve production efficiency and manage risks. It identifies a research gap in the role of machine learning for equipment status understanding, especially in WIPs, and its use in predictive maintenance to reduce operational downtime [43]. Existing literature lacks a comprehensive overview, doesn't address real-time ML complexities or provide clear ML architecture guidelines. Moreover, there's a lack of end-to-end solutions and benchmarks for algorithm selection [33]. In the context of AI-based predictive maintenance for WIPs in the oil and gas industry, specific features like component wear, corrosion, and clogging represent unexplored areas for AI-based maintenance prediction [64]. Furthermore, issues like electrical problems (short circuits and power surges) within water injection systems, and the development of leaks in components (piping and seals) causing efficiency reduction and the need for repairs, have received limited attention in AI-based predictive maintenance literature [7, 27]. The lack of research dedicated to predicting maintenance needs based on these specific WIP features underscores the need for further exploration. This study uniquely focuses on establishing a connection between these features and AI-driven PdM solutions in the OGI. The study addresses these gaps by investigating the potential of AI algorithms (ANN, SVM, and RF) in predicting maintenance requirements related to wear and tear, corrosion, clogging, electrical problems, and leaks in water injection pump systems [45]. By exploring these algorithms in this specific application, the study aims to enhance our understanding of AI-based PdM models for WIPs in the OGI and fill the research gap.

ANN is a powerful ML technique inspired by the human brain [44]. It excels at learning complex patterns and relationships within data. ANN can handle non-linear relationships and adapt to varying input types, making it suitable for analyzing the diverse and dynamic data collected from WIPs. Additionally, ANN can handle large datasets and perform well in both regression and classification tasks [6]. SVM's ability to handle both linear and non-linear relationships makes it a suitable choice for analyzing the data from WIPs, where multiple factors may influence the maintenance needs. RF is an ensemble learning technique that integrates multiple decision trees to make predictions. Renowned for its remarkable accuracy, scalability, and adeptness in handling extensive datasets with numerous features, RF has established itself as a powerful tool in ML [65]. RF can capture complex interactions between variables and identify important features for PdM analysis. This makes it a valuable tool for identifying patterns and predicting the maintenance needs of WIPs. Implementing PdM using AI techniques, such as analyzing data from WIPs, can offer several benefits, particularly in terms of cost optimization [66]. By accurately predicting maintenance requirements, companies can adopt a proactive approach, scheduling maintenance activities at optimal times. This approach can minimize unplanned downtime, prevent critical failures, and reduce repair costs [13]. By incorporating AI-based PdM, companies can potentially reduce costs associated with reactive maintenance, which often

involves addressing unexpected breakdowns and emergency repairs. With AI-driven maintenance, the timing of maintenance activities can be optimized, minimizing costly downtime and extending the lifespan of equipment. By allocating resources more efficiently, companies can achieve cost savings in the long run. After maintenance, the age of the equipment can be extended due to the proactive approach taken with PdM. By addressing potential issues before they escalate, equipment reliability can be significantly improved. PdM allows for the detection of early signs of degradation or malfunction, enabling timely repairs or replacements [67]. By estimating the RUL of equipment, engineers can schedule maintenance activities more accurately, optimize operating efficiency, and avoid unexpected failures or production interruptions [68]. The growing use of AI in PdM raises ethical considerations that should be addressed. These include:

(1) *Privacy and Data Security* AI models rely on vast amounts of data, including sensitive operational and maintenance information. Ensuring data privacy, security, and compliance with relevant regulations are paramount to protect the integrity and confidentiality of the data.

(2) *Bias and Fairness* AI models trained on biased data can perpetuate unfair practices and discrimination. It is crucial to identify and ensure fairness and prevent negative societal impacts, it is essential to actively address and mitigate biases in data collection, preprocessing, and model training processes.

(3) *Transparency and Accountability* Establishing transparent and accountable AI systems is essential for gaining user trust. Documentation.

## 11 Conclusion

This systematic review highlights the benefits of AI-based PdM for WIPs in the OGI. It offers early fault detection, accurate RUL estimation, and optimized maintenance planning. Real-world case studies show improved accuracy, reduced downtime, and cost savings, enhancing equipment reliability and environmental performance. Yet, addressing data challenges and ethical implementation is essential for further progress. Embracing these recommendations can optimize operations and decision-making. In this study, we examined the recent advancements in AI and ML in the oil and gas industries, with a specific focus on their application for diagnosing and forecasting WIPs defects. We found that the use of AI for WIPs defect diagnosis has significantly increased in the past decade, especially in oil and gas. AI, driven by Industry 4.0, is transforming mechanical manufacturing and automation, enhancing

data processing, computational power, and storage. It improves work efficiency, quality control, safety, problem diagnosis, PdM, and supply chain intelligence. The synergy between AI and mechanical manufacturing is accelerating the fourth Industrial Revolution. Our comprehensive study identifies common faults, algorithms, and parameters used in previous research, offering valuable insights for future investigations. However, challenges remain, as no universal method exists, and solutions depend on data type and algorithm suitability. Notably, ANNs, SVM, and hybrid models are frequently used for oil and gas WIPs health evaluation, with hybrid models showing the most promise. The health of WIPs is influenced by various factors, yet reliability, precision, and processing time were not consistently considered. This research highlights the prevalent use of four types of datasets, emphasizing a need for more diverse data sources. Currently, the focus is primarily on algorithms, and there is a lack of comprehensive data characterization for WIPs condition assessment, leading to potential inaccuracies in dynamic operational environments.

One significant limitation of the study is related to its scope and coverage, as it primarily focuses on AI-based PdM for WIPs in OGI, potentially restricting its broader relevance to the wider field of AI in industrial maintenance. The study also lacks an in-depth exploration of the challenges related to data reliability and availability in the OGI, which are crucial for the success of AI applications. Moreover, the absence of empirical evidence from real-world applications and the limited discussion on potential drawbacks or limitations of specific AI techniques for PdM are notable limitations. Additionally, the study does not explicitly address the practical implementation challenges and barriers associated with AI-based PdM in the OGI, which is vital for industry practitioners and decision-makers. Future research should aim to overcome these limitations by broadening the scope, addressing data-related challenges, incorporating real-world case studies, and providing a more balanced assessment of AI algorithm limitations and practical implementation challenges, thereby enhancing the study's value and applicability to industry professionals in the OGI and beyond.

**Data availability**  Not applicable.

**Code availability**  Not applicable.

## Declarations

**Conflict of interest**  On behalf of all authors, the corresponding author states that there is no conflict of interest regarding the publication of this article.

**Ethical approval**  Not applicable.

**Consent for publication**  All authors gave their consent to publish this article.

## References

1. Ghasemian S, Faridzad A, Abbaszadeh P, Taklif A, Ghasemi A, Hafezi R (2020) An overview of global energy scenarios by 2040: identifying the driving forces using cross-impact analysis method. Int J Environ Sci Technol. https://doi.org/10.1007/s13762-020-02738-5

2. Dudin MN, Frolova EE, Protopopova OV, Mamedov O, Odintsov SV (2019) Study of innovative technologies in the energy industry: nontraditional and renewable energy sources. Entrep Sustain Issues 6:1704–1713. https://doi.org/10.9770/jesi.2019.6.4(11)

3. Hallaji SM, Fang Y, Winfrey BK (2022) Predictive maintenance of pumps in civil infrastructure: state-of-the-art, challenges and future directions. Autom Constr 134:104049. https://doi.org/10.1016/j.autcon.2021.104049

4. Ouadah A, Zemmouchi-Ghomari L, Salhi N (2022) Selecting an appropriate supervised machine learning algorithm for predictive maintenance. Int J Adv Manuf Technol 119:4277–4301. https://doi.org/10.1007/s00170-021-08551-9

5. Al-Janabi YT (2020) An overview of corrosion in oil and gas industry: upstream, midstream, and downstream sectors. In: Saji VS, Umoren SA (eds) Corrosion inhibitors in the oil and gas industry. John Wiley & Sons, New Jersey, pp 1–39. https://doi.org/10.1002/9783527822140.ch1

6. Canizo M, Onieva E, Conde A, Charramendieta S, Trujillo S (2017) Real-time predictive maintenance for wind turbines using Big Data frameworks. In: 2017 IEEE international conference on prognostics and health management (ICPHM). IEEE, Texas. p 70–77

7. Amruthnath N, Gupta T (2018) A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In: 2018 5th international conference on industrial engineering and applications (ICIEA). IEEE, Singapore, p 355–361

8. Cheng H, Yang D, Lu C, Qin Q, Cadasse D (2022) Intelligent oil production stratified water injection technology. Wirel Commun Mob Comput 2022:1–7. https://doi.org/10.1155/2022/3954446

9. Garg A, Vijayaraghavan V, Tai K, Singru PM, Jain V, Krishnakumar N (2015) Model development based on evolutionary framework for condition monitoring of a lathe machine. Measurement 73:95–110. https://doi.org/10.1016/j.measurement.2015.04.025

10. Khan A, Goodell JW, Hassan MK, Paltrinieri A (2022) A bibliometric review of finance bibliometric papers. Finance Res Lett 47:102520. https://doi.org/10.1016/j.frl.2021.102520

11. Cheng JCP, Chen W, Chen K, Wang Q (2020) Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. Autom Constr 112:103087. https://doi.org/10.1016/j.autcon.2020.103087

12. Gallab M, Bouloiz H, Garbolino E, Tkiouat M, ElKilani MA, Bureau N (2017) Risk analysis of maintenance activities in a LPG supply chain with a multi-agent approach. J Loss Prev Process Ind 47:41–56. https://doi.org/10.1016/j.jlp.2017.02.026

13. Bekar ET, Nyqvist P, Skoogh A (2020) An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. Adv Mech Eng 12:168781402091920. https://doi.org/10.1177/1687814020919207

14. Falamarzi A, Moridpour S, Nazem M, Cheraghi S (2019) Prediction of tram track gauge deviation using artificial neural network and support vector regression. Aust J Civ Eng 17:63–71. https://doi.org/10.1080/14488353.2019.1616357

15. Nguyen TV, Barbosa YM, da Silva JAM, Junior SDO (2019) A novel methodology for the design and optimisation of oil and gas offshore platforms. Energy 185:158–175. https://doi.org/10.1016/j.energy.2019.06.164

16. Munim ZH, Dushenko M, Jimenez VJ, Shakil MH, Imset M (2020) Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Marit Policy Manag 47:577–597. https://doi.org/10.1080/03088839.2020.1788731

17. Kulkarni K, Devi U, Sirighee A, Hazra J, Rao P (2018) Predictive maintenance for supermarket refrigeration systems using only case temperature data. In: 2018 annual American control conference (ACC). IEEE, Wisconsin. p 4640–4645

18. Uhlmann E, Pontes RP, Geisert C, Hohwieler E (2018) Cluster identification of sensor data for predictive maintenance in a selective laser melting machine tool. Procedia Manuf 24:60–65. https://doi.org/10.1016/j.promfg.2018.06.009

19. Susto GA, McLoone S, Pagano D, Schirru A, Pampuri S, Beghi A (2013) Prediction of integral type failures in semiconductor manufacturing through classification methods. In: 2013 IEEE 18th conference on emerging technologies & factory automation (ETFA). IEEE, Italy, p 1–4

20. Zhao J, Dong Y, Fu J, Zhao L, Zhang Y (2020) Design and experiment of energy-saving water injection pump. J Pet Explor Prod Technol 10:2127–2135. https://doi.org/10.1007/s13202-020-00860-1

21. Waeyenbergh G, Pintelon L (2002) A framework for maintenance concept development. Int J Prod Econ 77:299–313. https://doi.org/10.1016/s0925-5273(01)00156-6

22. Tchakoua P, Wamkeue R, Ouhrouche M, Slaoui-Hasnaoui F, Tameghe T, Ekemb G (2014) Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges. Energies 7:2595–2630. https://doi.org/10.3390/en7042595

23. Aliyu R, Mokhtar AA, Hussin H (2022) Prognostic health management of pumps using artificial intelligence in the oil and gas sector: a review. Appl Sci 12:11691. https://doi.org/10.3390/app122211691

24. Panda AK, Rapur JS, Tiwari R (2018) Prediction of flow blockages and impending cavitation in centrifugal pumps using support vector machine (SVM) algorithms based on vibration measurements. Measurement 130:44–56. https://doi.org/10.1016/j.measurement.2018.07.092

25. Tiwari R, Bordoloi DJ, Dewangan A (2021) Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm. Measurement 173:108676. https://doi.org/10.1016/j.measurement.2020.108676

26. Grundfos Research and Technology (2009) Mechanical shaft seals for pumps

27. Prytz R, Nowaczyk S, Rögnvaldsson T, Byttner S (2015) Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data. Eng Appl Artif Intell 41:139–150. https://doi.org/10.1016/j.engappai.2015.02.009

28. Huuhtanen T, Jung A (2018) Predictive maintenance of photovoltaic panels via deep learning. In: 2018 IEEE data science workshop (DSW). IEEE, Switzerland. p 66–70

29. Janssens O, Van de Walle R, Loccufier M, Van Hoecke S (2017) Deep learning for infrared thermal image based machine health monitoring. IEEE/ASME Trans Mechatron 23:151–159. https://doi.org/10.1109/tmech.2017.2722479

30. Sampaio GS, Filho ARDAV, Silva LSD, Silva LAD (2019) Prediction of motor failure time using an artificial neural network. Sensors 19:4342. https://doi.org/10.3390/s19194342

31. Susto GA, Beghi A, De Luca C (2012) A Predictive maintenance system for epitaxy processes based on filtering and prediction techniques. IEEE Trans Semicond Manuf 25:638–649. https://doi.org/10.1109/tsm.2012.2209131

32. Praveenkumar T, Saimurugan M, Krishnakumar P, Ramachandran KI (2014) Fault diagnosis of automobile gearbox based on machine learning techniques. Procedia Eng 97:2092–2098. https://doi.org/10.1016/j.proeng.2014.12.452

33. Abu-Samah A, Shahzad MK, Zamai E, Said AB (2015) Failure prediction methodology for improved proactive maintenance using Bayesian approach. IFAC-PapersOnLine 48:844–851. https://doi.org/10.1016/j.ifacol.2015.09.632

34. Biswal S, Sabareesh GR (2015) Design and development of a wind turbine test rig for condition monitoring studies. In: 2015 international conference on industrial instrumentation and control (ICIC). IEEE, India. p 891–896

35. Susto GA, Beghi A (2016) Dealing with time-series data in predictive maintenance problems. In: 2016 IEEE 21st international conference on emerging technologies and factory automation (ETFA). IEEE, Germany. p 1–4

36. Durbhaka GK, Selvaraj B (2016) Predictive maintenance for wind turbine diagnostics using vibration signal analysis based on collaborative recommendation approach. In: 2016 international conference on advances in computing, communications and informatics (ICACCI). IEEE, India. p 1839–1842

37. Mathew V, Toby T, Singh V, Rao BM, Kumar MG (2017) Prediction of Remaining Useful Lifetime (RUL) of turbofan engine using machine learning. In: 2017 IEEE international conference on circuits and systems (ICCS). IEEE, India. p 306–311

38. Pan Z, Ge Y, Zhou YC, Huang JC, Zheng YL, Zhang N, Liang XX, Gao P, Zhang GQ, Wang Q, Shi S-B (2017) Cognitive acoustic analytics service for internet of things. In: 2017 IEEE international conference on cognitive computing (ICCC). IEEE, USA. p 96–103

39. Kumar A, Shankar R, Thakur LS (2018) A big data driven sustainable manufacturing framework for condition-based maintenance prediction. J Comput Sci 27:428–439. https://doi.org/10.1016/j.jocs.2017.06.006

40. Lasisi A, Attoh-Okine N (2018) Principal components analysis and track quality index: a machine learning approach. Transp Res C Emerg Technol 91:230–248. https://doi.org/10.1016/j.trc.2018.04.001

41. Kolokas N, Vafeiadis T, Ioannidis D, Tzovaras D (2018) Forecasting faults of industrial equipment using machine learning classifiers. In: 2018 innovations in intelligent systems and applications (INISTA). IEEE, Greece. p 1–6

42. Abdalla R, Samara H, Perozo N, Carvajal CP, Jaeger P (2022) Machine learning approach for predictive maintenance of the electrical submersible pumps (ESPs). ACS Omega 7:17641–17651. https://doi.org/10.1021/acsomega.1c05881

43. Li H, Parikh D, He Q, Qian B, Li Z, Fang D, Hampapur A (2014) Improving rail network velocity: a machine learning approach to predictive maintenance. Transp Res C Emerg Technol 45:17–26. https://doi.org/10.1016/j.trc.2014.04.013

44. Paolanti M, Romeo L, Felicetti A, Mancini A, Frontoni E, Loncarski J (2018) Machine learning approach for predictive maintenance in industry 4.0. In: 2018 14th IEEE/ASME international conference on mechatronic and embedded systems and applications (MESA). IEEE, Finland. p 1–6

45. Su C-J, Huang S-F (2018) Real-time big data analytics for hard disk drive predictive maintenance. Comput Electr Eng 71:93–101. https://doi.org/10.1016/j.compeleceng.2018.07.025

46. Eke S, Aka-Ngnui T, Clerc G, Fofana I (2017) Characterization of the operating periods of a power transformer by clustering the dissolved gas data. In: 2017 IEEE 11th international symposium on diagnostics for electrical machines, power electronics and drives (SDEMPED). IEEE, Greece. p 298–303

47. Butte S, Prashanth AR, Patil S (2018) Machine learning based predictive maintenance strategy: a super learning approach with deep neural networks. In: 2018 IEEE workshop on microelectronics and electron devices (WMED). IEEE, USA. p 1–5

48. Ruqaishi M, Bashir HA (2015) Causes of delay in construction projects in the oil and gas industry in the Gulf cooperation council countries: a case study. J Manag Eng 31:05014017. https://doi.org/10.1061/(asce)me.1943-5479.0000248

49. Çınar ZM, Nuhu AA, Zeeshan Q, Korhan O, Asmael M, Safaei B (2020) Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability 12:8211. https://doi.org/10.3390/su12198211

50. Amihai I, Chioua M, Gitzel R, Kotriwala AM, Pareschi D, Sosale G, Subbiah S (2018) Modeling machine health using gated recurrent units with entity embeddings and K-means clustering. In: 2018 IEEE 16th international conference on industrial informatics (INDIN). IEEE, Portugal. p 212–217

51. Santos dos T, Ferreira FJTE, Pires JM, Damasio C (2017) Stator winding short-circuit fault diagnosis in induction motors using random forest. In: 2017 IEEE international electric machines and drives conference (IEMDC). IEEE, USA. p 1–8

52. Duc TL, Leiva RG, Casari P, Östberg P-O (2019) Machine learning methods for reliable resource provisioning in edge-cloud computing. ACM Comput Surv 52:1–39. https://doi.org/10.1145/334114

53. Xiang S, Huang D, Li X (2018) A generalized predictive framework for data driven prognostics and diagnostics using machine logs. In: TENCON 2018–2018 IEEE region 10 conference. IEEE, South Korea. p 0695–0700

54. Rueckel J, Kunz WG, Hoppe BF, Patzig M, Notohamiprodjo M, Meinel FG, Cyran CC, Ingrisch M, Ricke J, Sabel BO (2020) Artificial intelligence algorithm detecting lung infection in supine chest radiographs of critically ill patients with a diagnostic accuracy similar to board-certified radiologists. Crit Care Med 48:e574–e583. https://doi.org/10.1097/ccm.0000000000004397

55. Osman AIA, Najah Ahmed A, Chow MF, Feng Huang Y, El-Shafie A (2021) Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Eng J 12:1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

56. Abbasi T, Lim KH, Rosli NS, Ismail I, Ibrahim R (2018) Development of predictive maintenance interface using multiple linear

regression. In: 2018 international conference on intelligent and advanced system (ICIAS). IEEE, Malaysia. p 1–5

57. Aivaliotis P, Arkouli Z, Georgoulias K, Makris S (2021) Degradation curves integration in physics-based models: towards the predictive maintenance of industrial robots. Robot Comput Integr Manuf 71:102177. https://doi.org/10.1016/j.rcim.2021.102177

58. de Jonge B, Scarf PA (2020) A review on maintenance optimization. Eur J Oper Res 285:805–824. https://doi.org/10.1016/j.ejor.2019.09.047

59. Carvalho TP, Soares FAAMN, Vita R, Francisco RDP, Basto JP, Alcalá SGS (2019) A systematic literature review of machine learning methods applied to predictive maintenance. Comput Ind Eng 137:106024. https://doi.org/10.1016/j.cie.2019.106024

60. Malla C, Panigrahi I (2019) Review of condition monitoring of rolling element bearing using vibration analysis and other techniques. J Vib Eng amp Technol 7:407–414. https://doi.org/10.1007/s42417-019-00119-y

61. Fausing Olesen J, Shaker HR (2020) Predictive maintenance for pump systems and thermal power plants: state-of-the-art review, trends and challenges. Sensors 20:2425. https://doi.org/10.3390/s20082425

62. Rathore A, Patidar NP (2019) Reliability assessment using probabilistic modelling of pumped storage hydro plant with PV-Wind based standalone microgrid. Int J Electr Power Energy Syst 106:17–32. https://doi.org/10.1016/j.ijepes.2018.09.030

63. Syed FI, Alshamsi M, Dahaghi AK, Neghabhan S (2022) Artificial lift system optimization using machine learning applications. Petroleum 8:219–226. https://doi.org/10.1016/j.petlm.2020.08.003

64. Guo J, Wan J-L, Yang Y, Dai L, Tang A, Huang B, Zhang F, Li H (2023) A deep feature learning method for remaining useful life prediction of drilling pumps. Energy 282:128442. https://doi.org/10.1016/j.energy.2023.128442

65. Berbiche N, El Alami J (2023) Enhancing anomaly-based intrusion detection systems: a hybrid approach integrating feature selection and bayesian hyperparameter optimization. Ing Syst Inf 28:1177–1195. https://doi.org/10.18280/isi.280506

66. Lee WJ, Wu H, Yun H, Kim H, Jun MBG, Sutherland JW (2019) Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data. Procedia CIRP 80:506–511. https://doi.org/10.1016/j.procir.2018.12.019

67. Manchadi O, Ben-Bouazza F-E, Jioudi B (2023) Predictive maintenance in healthcare system: a survey. IEEE Access 11:61313–61330. https://doi.org/10.1109/access.2023.3287490

68. Han X, Wang Z, Xie M, He Y, Li Y, Wang W (2021) Remaining useful life prediction and predictive maintenance strategies for multi-state manufacturing systems considering functional dependence. Reliab Eng Syst Saf 210:107560. https://doi.org/10.1016/j.ress.2021.107560