

Enhancing Predictive Maintenance of Industrial Assets Through Machine Diagnostic Parameter Grouping

Sławomir LUŚCIŃSKI¹, Mariusz BEDNAREK^{2,3}, Marek JABŁOŃSKI⁴

¹ *Kielce University of Technology, Department of Production Engineering, Poland*

² *WSB Merito University in Poznań, Poland*

³ *Universidad Autonoma de Chile, Temuco, Chile*

⁴ *WSB Merito University in Poznań, Management and Quality Institute, Poland*

Received: 07 December 2024

Accepted: 27 February 2025

Abstract

This article examines the advancement of predictive maintenance (PdM) for industrial assets through an innovative methodology that categorises diagnostic parameters into coherent groups. Predictive maintenance constitutes a vital component in mitigating unforeseen downtime and improving operational efficiency within manufacturing settings. The authors recommend a centralised framework for PdM, effectively addressing the complexities arising from data saturation by numerous sensor nodes. The proposed methodology refines the predictive maintenance process by systematically organising diagnostic parameters based on their significance and interconnections, thereby enhancing its effectiveness and efficiency. The study utilises the KNIME software platform for comprehensive data analysis and validation of the proposed approach, demonstrating its practicality with datasets obtained from SCADA/MES systems. The results confirm the robustness and accessibility of the methodology, highlighting its potential applicability across various industrial sectors. Future research directions include the integration of advanced machine learning techniques and the exploration of the methodology's relevance in diverse industries.

Keywords

Predictive Maintenance (PdM); Diagnostic Parameters; Hierarchical Clustering; SCADA/MES Systems; KNIME Software.

Introduction

Objective and scope of the study

Maintenance refers to the activities and processes undertaken to preserve or restore the functionality of equipment, machinery, or systems. It is essential in various industries, particularly manufacturing, to minimise unplanned downtime, ensure product quality, and maintain operational efficiency (Lee et al., 2020). According to the EN 13306:2017 standard (EN 13306, 2024), “maintenance is a combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore

it to, a state in which it can perform the required function”. Predictive maintenance is a proactive maintenance strategy that utilises real-time data and monitoring techniques to assess the condition of equipment and predict when maintenance should be performed. This approach aims to prevent unexpected equipment failures and minimise downtime by addressing potential issues before they escalate into serious problems (Mobley, 2002).

Industry practice is evolving towards decentralised models in predictive maintenance (PdM) due to the increasing number of sensor nodes in industrial environments that generate vast amounts of data, which may result in a need for more data processing efficiency (Nunes et al., 2023). However, the authors propose a centralised approach to adopting a predictive maintenance paradigm for the production system instead of individual machines. A diagnostic parameter grouping method was proposed to avoid struggling with data overload due to the application of a centralised approach. In this article, a novel

Corresponding author: Sławomir Luściński – Kielce University of Technology, Department of Production Engineering, Al. Tysiąclecia Państwa Polskiego 7, 25-314, Kielce, Poland, e-mail: luscinski@tu.kielce.pl

© 2025 The Author(s). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

methodology has been developed and validated to group diagnostic parameters and streamline the predictive maintenance process design. This innovative approach reduces the number of critical parameters monitored for predictive analysis. By systematically categorising these parameters according to their relative importance and interrelations, the proposed method facilitates the development of a more efficient and effective predictive maintenance strategy, consequently augmenting the overall value of the research and fostering optimism regarding the prospective advancements in predictive maintenance. A centralised approach driven by industrial asset diagnostic parameter redundancy may provide more holistic and meaningful information about the production system than decentralised edge computing of diagnostic data.

The paper's structure is organised to foster a comprehensive understanding of enhancing predictive maintenance within condition-based systems via a diagnostic parameter grouping methodology. The introductory section (Section 1) is followed by a literature review (Section 2), which elucidates the foundational principles of CBM, the role of predictive analytics, and the challenges associated with traditional parameter monitoring techniques. Fig. 1 presents the segmentation of the research design methodology employed in organising the research design section (Section 3).

Subsequent sections of the article present the outcomes derived from a computational experiment, a crucial step in validating the effectiveness and applica-

bility of the chosen method and thereby instilling confidence in its reliability. Section 4 discusses the obtained results, and Section 5 summarises this paper's contributions and future research prospects.

Literature review

Predictive maintenance

Maintenance refers to the activities and processes undertaken to preserve or restore the functionality of equipment, machinery, or systems. It is essential in various industries, particularly manufacturing, to minimise unplanned downtime, ensure product quality, and maintain operational efficiency. Fig. 2 presents the evolution of maintenance strategies.

Condition-based maintenance (CBM) has emerged as a fundamental strategy within contemporary industrial environments. It uses real-time data analytics to anticipate equipment malfunctions and refine maintenance schedules. The principal aim of CBM is to transition from reactive maintenance methodologies to a proactive framework, thereby augmenting operational efficiency and minimising expenses linked to unforeseen downtimes. A predictive maintenance (PdM) initiative represents a systematic framework for the upkeep of equipment and machinery by surveilling their real-time operational parameters to forecast the optimal timing for maintenance interventions. Predic-

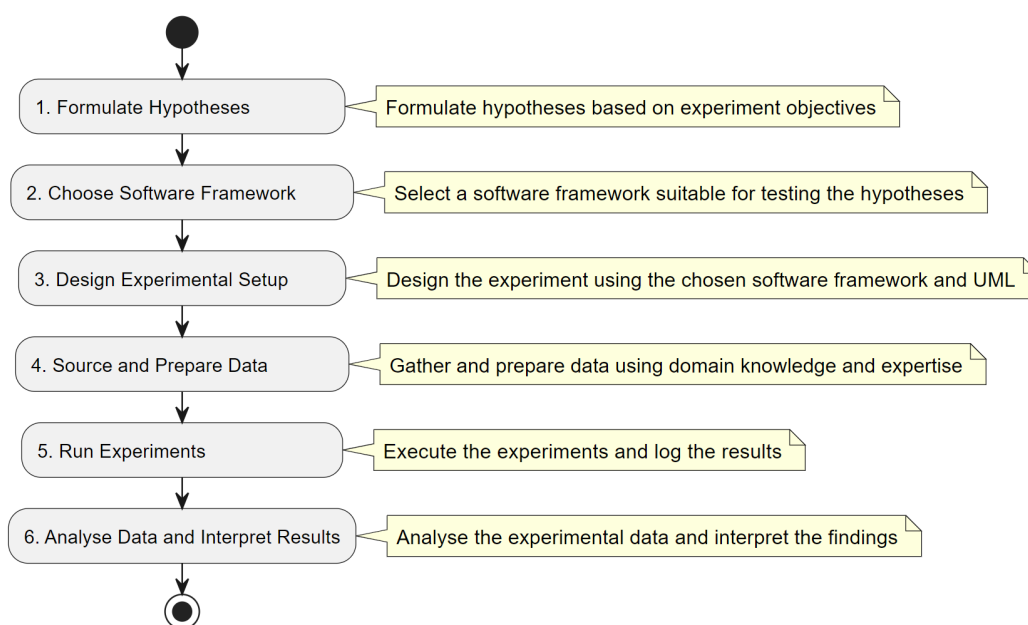


Fig. 1. Staged research design process. Source: own

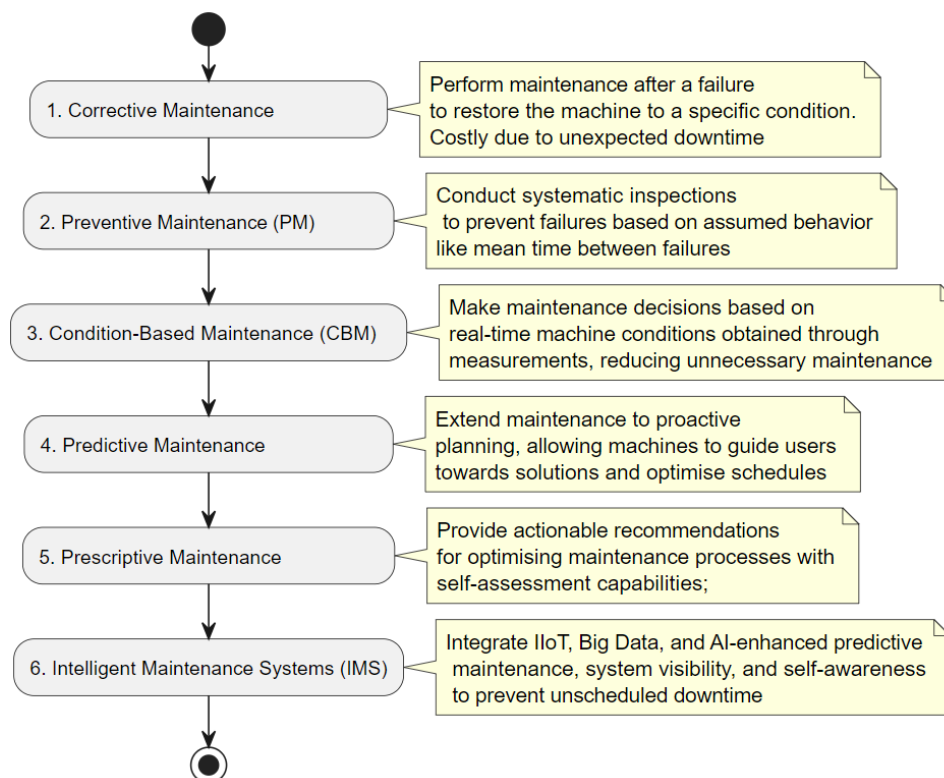


Fig. 2. The evolution of maintenance strategies. Source: based on [Berthold \(2008\)](#)

tive maintenance is a proactive maintenance strategy that utilises real-time data and monitoring techniques to assess the condition of equipment and predict when maintenance should be performed. This proactive approach aims to prevent unexpected equipment failures and minimise downtime by addressing potential issues before they escalate into serious problems ([Mobley, 2002](#)). Tab. 1 presents critical components of the predictive maintenance approach.

The authors highlight the increasing trend of using multiple sensors for condition monitoring and the necessity of data fusion techniques to enhance diagnostic and prognostic accuracy. Predictive analytics plays a vital role in CBM by harnessing both historical and real-time data to project possible equipment failures. Various statistical and machine-learning methodologies have scrutinised data patterns and discerned failure indicators ([Gallego Garcia & García, 2019](#)). Empirical studies have illustrated that sophisticated predictive models can enhance maintenance scheduling and resource allocation, ultimately improving operational performance. PdM has evolved significantly with advancements in Industry 4.0 and the Internet of Things (IoT). It combines sensor data and analytics to improve maintenance scheduling, moving beyond traditional corrective and preventive maintenance strategies.

Data sourcing for predictive maintenance

Anomaly detection is concerned with identifying data values that deviate from typical behaviour. Several factors may cause the anomaly; some of them are related to errors in the acquisition system, such as sensor malfunction, low battery, or errors during data transmission, while other anomalies may be caused by an industrial equipment malfunction or event, such as changes in the production line or a curative stop ([Nunes et al., 2023](#)).

The amalgamation of data from Supervisory Control and Data Acquisition (SCADA) systems and Manufacturing Execution Systems (MES) has become increasingly common, furnishing a substantial reservoir of information for predictive maintenance ([Jardine et al., 2006](#)). Integrating Supervisory Control and Data Acquisition (SCADA) systems with Manufacturing Execution Systems (MES) is pivotal in enhancing the efficiency and reliability of highly automated production systems, particularly within the framework of Industry 4.0. Fusing data from SCADA and MES systems creates a substantial reservoir of information invaluable to predictive maintenance strategies. This integration allows for monitoring and predicting system performance, essential for condition-based

Table 1
Critical components of predictive maintenance

Component	Description
Condition monitoring	Employ various methodologies, including vibration analysis, thermographic imaging, tribological evaluation, and monitoring of process parameters, to evaluate the operational integrity of machinery. These techniques yield instantaneous data regarding the equipment's mechanical state and operational efficiency.
Data analysis	The systematic accumulation and examination of data from observational activities is essential for discerning patterns and irregularities that may signify impending failures. Such analytical endeavours facilitate the formulation of well-informed judgments regarding the optimal timing for executing maintenance activities.
Scheduling maintenance	Informed by condition monitoring and empirical data analysis, maintenance operations are orchestrated according to necessity rather than a predetermined schedule. This methodology enhances the efficiency of maintenance initiatives and mitigates expenditures.
Integration with other maintenance strategies	Although predictive maintenance represents a formidable approach, it is frequently joined with alternative maintenance methodologies, such as preventive and reactive maintenance, to formulate a holistic maintenance management framework.

Source: own elaboration based on (Mobley, 2002).

maintenance. These architectures enable the seamless integration of IT and OT fields, providing maintenance teams with valuable insights and customised alarms for proactive maintenance. Integrating SCADA systems with MES in Industry 4.0 enhances operational efficiency, data visibility, and decision-making in manufacturing. This fusion creates a substantial data pool for predictive maintenance, supporting effective condition-based maintenance strategies (Mallioris et al., 2024). The synergy between SCADA and MES systems enhances the performance of highly automated production systems, ensuring greater efficiency and dependability in industrial automation and control processes. SCADA systems play a crucial role in

real-time monitoring and data acquisition, which is essential for maintaining operational efficiency and minimising downtime in manufacturing environments (Chandna, 2024). For instance, SCADA systems facilitate real-time visualisation of manufacturing processes, identifying and addressing performance losses and quality control issues, thereby supporting continuous improvement and process enhancement (Kaundal et al., 2024). Conversely, MES is the central information hub in manufacturing, leveraging IoT-ready devices to generate and process real-time data.

Hierarchical Clustering

Conventional predictive maintenance techniques often rely on machine learning, constraint satisfaction, and Quality-of-Service (QoS) aggregation to forecast potential process issues. These methods have been shown to achieve a minimum accuracy of 70% when predictions are made with sufficient lead time, such as half of the process duration. However, they vary in their ability to predict violations and non-violations accurately, suggesting that combining different techniques can significantly enhance precision and recall (Metzger et al., 2015). Diminishing the number of critical parameters under surveillance enhances the efficacy of predictive maintenance systems for the systematic organisation of analogous data points predicated on distance metrics (Murtagh & Legendre, 2014). Hierarchical clustering could be applied here to group components with similar degradation patterns, allowing for more efficient maintenance scheduling and resource allocation. This approach could help reduce costs and spare part usage by identifying elements that can be maintained together, thereby improving the overall maintenance strategy.

Hierarchical clustering is a prominent data analysis methodology, particularly in predictive maintenance. Hierarchical clustering in predictive maintenance offers the advantage of capturing shared cluster patterns across multiple machines, enabling health assessment, degradation modelling, and machine comparison. This approach, exemplified by the Gaussian topic model (GTM) (Karlsson et al., 2021), allows for directly comparing cluster proportions between machines, leading to insights into degradation behaviours. Unlike traditional models such as the Gaussian mixture model (GMM), the GTM considers group-level information, making it more adept at identifying crucial cluster patterns. Therefore, hierarchical clustering facilitates a deeper understanding of degradation behaviour across machines, enhancing the efficiency and reliability of predictive maintenance systems (Karlsson et al., 2021). Another advantage of clustering in pre-

dictive maintenance is its ability to work with limited data. In scenarios where sensory data is insufficient to build a comprehensive physical model, clustering approaches such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) can be particularly beneficial. These methods allow for the segmentation and clustering of operational points using minimal features extracted from single sensor signals, making them general and adaptable to various contexts (Calzavara et al., 2021).

Hierarchical clustering in predictive maintenance offers the advantage of grouping similar data points into clusters based on proximity, allowing for the identification of patterns and trends in equipment performance. This method aids in predicting the right time for maintenance by tracing median values of clusters, as demonstrated in the development of predictive maintenance technology for wafer transfer robots (Kim et al., 2019).

Research Design

Hypothesis Formulation

For the preliminary validation of the innovative methodology that groups diagnostic parameters to streamline predictive maintenance processes in Condition-based Maintenance (CBM) systems, the hypotheses are as follows:

1. H1: "Grouping diagnostic parameters based on their importance and interdependencies, as proposed by the new methodology, can be effectively implemented with data reported from SCADA/MES systems".
2. H2: "The new method, designed with practicality in mind, should demonstrate domain neutrality and ease of following by practitioners with open-source software tools, ensuring its ease of implementation and use in your work".

Methodology Design

The selection of computational tools and frameworks is crucial for effectively executing predictive maintenance strategies. The Konstanz Information Miner (KNIME) has emerged as a prominent open-source platform for data analysis and corporate analytics (Berthold et al., 2008). The modular architecture of KNIME enables users to develop workflows that amalgamate diverse data processing and analytical tasks, which is especially beneficial for performing intricate analyses such as hierarchical clustering (Ibrahim et al., 2016). The capability to integrate external programming languages, including Python and R, significantly

augments its adaptability, rendering it an appropriate option for professionals aiming to incorporate advanced analytics into their maintenance operations. It aids users in scrutinising, refining, and presenting data in an accessible and effective manner. Through its graphical user interface (GUI), KNIME empowers users to visually conceptualise and execute data processes utilising a modular arrangement of interconnected nodes. These nodes are the foundational elements of KNIME workflows, utilised for a broad spectrum of data manipulation and modelling functions. KNIME provides a diverse array of pre-designed components and extensions in the form of nodes, which address specific demands in data analysis, including text mining, image processing, and machine learning algorithms. It enhances the platform's adaptability and capacity to accommodate diverse data analysis needs.

Experiment Design

The experiment design process was supported by using an open-source low-code solution for automated diagram creation (PlantUML Web Server, 2024). Fig. 3 shows the part of the computational procedure in PlantUML notation developed to diagram the KNIME workflow design (Fig. 4).

```
@startuml
    Computational Sequence
    !theme vibrant
    hide footbox
    actor User
    participant "Node 01\nExcel\nReader" as EXR
    participant "Node 02\nTranspose" as TRS
    participant "Node 03\nMissing\nValue" as MVL
    participant "Node 04\nNormalizer" as NRM
    participant "Node 05\nDistant\nMatrix\nCalculate" as DMC
    participant "Node 06\nHierarchical\nClustering" as HCL
    participant "Node 07\nHierarchical\nCluster\nView" as HCV
    participant "Node 08\nHierarchical\nCluster\nAssigner" as HCA

    == Step 1 ==
    activate EXR #Green
    User-->EXR: Supply data time series
    EXR->EXR: Loads data
    User-->EXR: Configures node
    EXR->EXR: Extract data
    EXR->TRS: Pass data
    deactivate EXR
```

Fig. 3. Screenshot of the sequence diagram code in PlantUML notation. Source: own

The computational experiment was structured using the KNIME software framework, which provided a designated environment for data examination and analytics. The KNIME workflow (Fig. 5) was developed with parametrised computational nodes to implement an algorithm for hierarchical clustering of critical parameters.

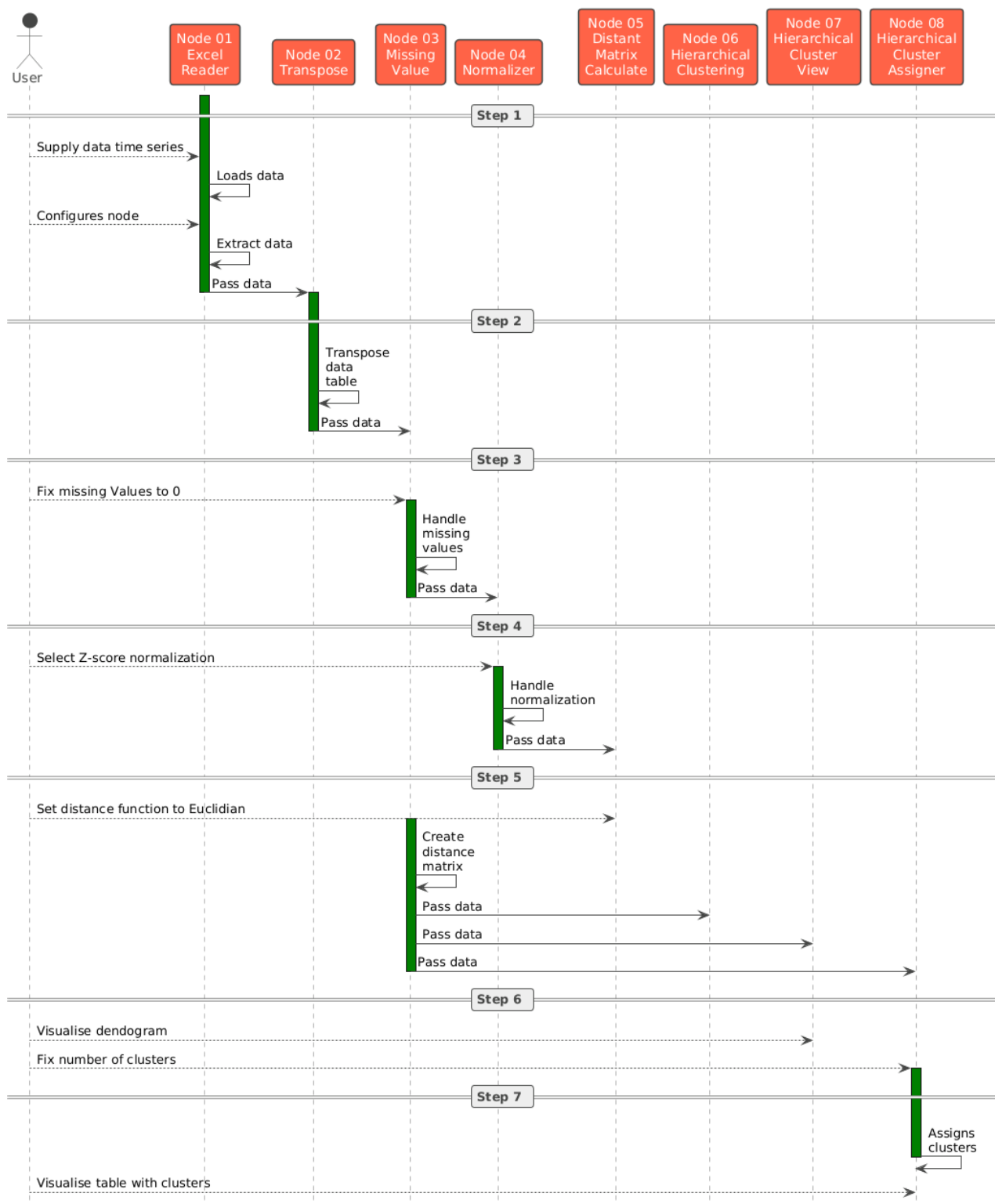


Fig. 4. Computation sequence diagram. Source: own

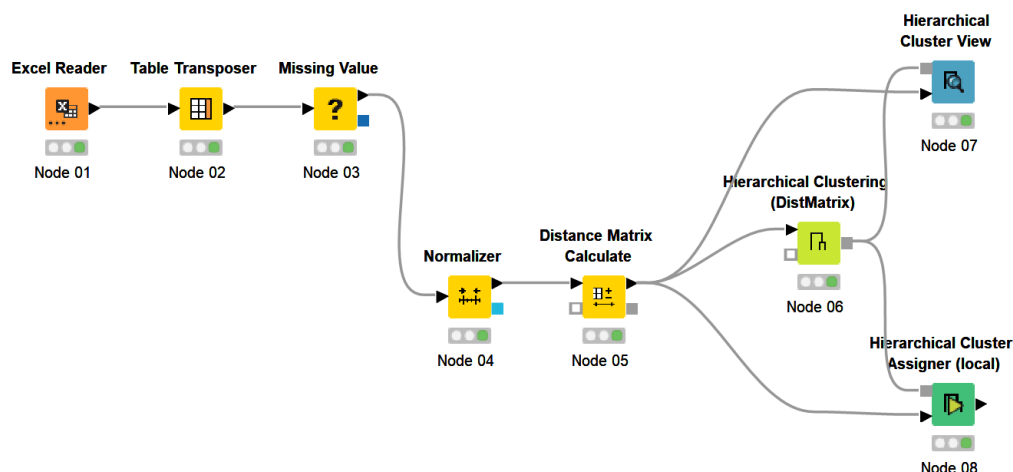


Fig. 5. Implemented KNIME workflow diagram. Source: KNIME screenshot of own elaboration

The workflow involved several nodes, such as *Excel Reader*, *Transpose*, *Missing Value*, *Normaliser*, *Distance Matrix Calculate*, *Hierarchical Clustering*, *Hierarchical Cluster View*, and *Hierarchical Cluster Assigner*. These nodes processed the data and implemented an algorithm for clustering critical parameters. The KNIME workflow was developed with parametrised computational nodes to standardise the data through Z-score normalisation, calculate Euclidean distance values, and hierarchically cluster the input data using a *Complete Linkage strategy* parameter. The results were visualised through the *Hierarchical Cluster View* node, which represented the hierarchical cluster tree, and the *Hierarchical Cluster Assigner* node, which allocated each data point to a specific cluster for further analysis.

Data Collection and Preparation

Anomaly detection identifies data points that significantly deviate from standard patterns. Such anomalies can arise from various factors; some are related to inaccuracies in data collection, like sensor failures, low battery levels, or errors during transmission. Others may result from malfunctions or incidents involving industrial equipment, such as changes in the production line or scheduled maintenance. To ensure the study's robustness, we obtained authentic raw data in CSV format from the industrial company's MES integrated with the SCADA system. (Tab. 2).

The obtained data series were transformed into an example malfunction data matrix in an MS Excel worksheet (Fig. 6). Each row represents an occurrence of

Table 2
Data set

Work order	Asset id	Error code	Error code description
10382	2177	TMQ 1117	FUNCTIONAL_TEST_REMOVE_ERROR
10364	11897	TMQ 150	COUNTER_TEST_WITH_TESTER_ERROR
1105	4915	TMQ 2010	ELECTRIC_TEST_CONVERTER_ERROR
10039	1080	TMQ 2038	ELECTRIC_TEST_WRONG_VOLTAGE_MEASURE_VCC
10597	4847	TMQ 2042	ELECTRIC_TEST_WRONG_VOLTAGE_MEASURE_UBAT
10285	2055	TMQ 2072	ELECTRIC_TEST_MOSFET_MOTOR_ERROR
1646	1625	TMQ 245	LED_VERIFICATION_ERROR
10294	1623	TMQ 66	INITIALIZE_FUNCTIONAL_TESTER_ERROR
10325	1622	TMQ 96	INIT_OPTO_CONNECTION_ERROR
10122	1621	TMQ 99	CHECK_SERIAL_NBR_ERROR

Source: supplied by AIUT Sp. z o.o. (AIUT.com).

	A	B	C	D	E	F	G	H	I	J	K
1	Asset ID	TMQ 1117	TMQ 150	TMQ 2010	TMQ 2038	TMQ 2042	TMQ 2072	TMQ 245	TMQ 66	TMQ 96	TMQ 99
2	A_10382	1									
3	A_10364		1								
4	A_1105			1							
5	A_10039				1						
6	A_10597					1					
7	A_10285						1				
8	A_1646							1			
9	A_10294							1			
10	A_10325									1	
11	A_10122										1

Fig. 6. Example malfunction data matrix. Source: screenshot from MS Excel worksheet elaborated with source data

a malfunction of a production asset, and the columns represent the Asset ID and Error codes taken from the SCADA system. The first column was filled with the asset IDs involved in malfunctions. Each malfunction was coded by putting the binary value “1” in the error code column, respectively. It is assumed that error codes represent critical parameters monitored online in SCADA, making them useful for predictive maintenance.

Computational Experiment Execution

The experiment began by loading a malfunction data matrix in an MS Excel worksheet. Each row represented a malfunction occurrence of a production asset, and the columns represented the Asset ID and Error codes from the SCADA system. The error codes were considered critical parameters for predictive maintenance. Subsequently, Node 02, using the “Table Transposer,” transposes the entire input table by swapping rows and columns. Subsequently, Node 03, acting as “Missing Value,” handles any missing data in the table cells by replacing them with “0”.

Advancing further, Node 04 utilises the “Normalisation Processor” to standardise the data through Z-score normalisation (Gaussian). This linear transformation

ensures all values in each column follow a Gaussian distribution with a mean of 0.0 and a standard deviation of 1.0 (Fig. 7).

Node 05 runs the “Distance Matrix Calculate” function, which computes the Euclidean distance values for all pairs of rows in the input table. Afterwards, the result of this computation is added to the original input table as a new column containing the distance vector values. (Fig. 8).

Node 06 utilises the functionality of “Hierarchical Clustering (DistMatrix)” to hierarchically cluster the input data by employing a distance matrix as an input. The node implements a bottom-up or agglomerative approach, wherein the algorithm initiates with each data-point as an individual cluster and endeavours to merge the most similar ones into superclusters until eventually culminating in one extensive cluster encompassing all subclusters. A Complete Linkage strategy for this node is pre-defined, determining the distance between two clusters, c_1 and c_2 , as the maximum distance between any pair of points x and y , where x is in c_1 , and y is in c_2 . Consequently, a hierarchical cluster tree is generated, which can be input into either the “Hierarchical Cluster View” node or the “Hierarchical Cluster Assigner” node. The “Hierarchical Cluster View” node

Row ID	D Row0	D Row1	D Row2	D Row3	D Row4	D Row5	D Row6	D Row7	D Row8	D Row9
TMQ 1117	2.846	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316
TMQ 150	-0.316	2.846	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316
TMQ 2010	-0.316	-0.316	2.846	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316
TMQ 2038	-0.316	-0.316	-0.316	2.846	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316
TMQ 2042	-0.316	-0.316	-0.316	-0.316	2.846	-0.316	-0.316	-0.316	-0.316	-0.316
TMQ 2072	-0.316	-0.316	-0.316	-0.316	-0.316	2.846	-0.316	-0.316	-0.316	-0.316
TMQ 245	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	2.846	2.846	-0.316	-0.316
TMQ 66	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316
TMQ 96	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	2.846	-0.316
TMQ 99	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	-0.316	2.846

Fig. 7. Normalised data. Source: Own elaboration & KNIME screenshot

Row ID	D Row0	Distance
TMQ 1117	2.846	0 []
TMQ 150	-0.316	1 [4.472135954999579]
TMQ 2010	-0.316	2 [4.472135954999579, 4.472135954999579]
TMQ 2038	-0.316	3 [4.472135954999579, 4.472135954999579, 4.472135954999579]
TMQ 2042	-0.316	4 [4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579]
TMQ 2072	-0.316	5 [4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579]
TMQ 245	-0.316	6 [5.47722557505166, 5.47722557505166, 5.47722557505166, 5.47722557505166, 5.47722557505166, 5.47722557505166]
TMQ 66	-0.316	7 [3.162277660168379, 3.162277660168379, 3.162277660168379, 3.162277660168379, 3.162277660168379, 3.162277660168379, 3.162277660168379]
TMQ 96	-0.316	8 [4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579]
TMQ 99	-0.316	9 [4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579, 4.472135954999579]

Fig. 8. Clustered data view. Source: Own elaboration & KNIME screenshot

enables the visualisation of the hierarchical cluster tree, offering a lucid representation of the relationships and groupings present within the data (Fig. 9).

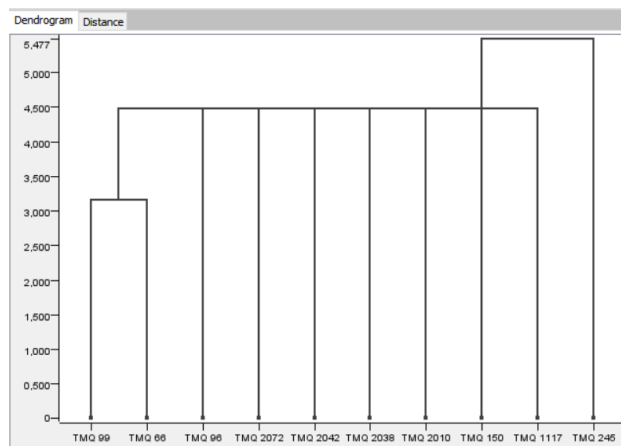


Fig. 9. Dendrogram view. Source: Own elaboration & KNIME screenshot

Conversely, the “Hierarchical Cluster Assigner” node assigns each data point to a specific cluster based on the hierarchical structure, thereby facilitating a more detailed analysis and interpretation of the clustered data (Fig. 10).

Row ID	Cluster...
TMQ 1117	2
TMQ 150	1
TMQ 2010	1
TMQ 2038	1
TMQ 2042	1
TMQ 2072	1
TMQ 245	0
TMQ 66	1
TMQ 96	1
TMQ 99	1

Fig. 10. Euclidian distance vector values. Source: Own elaboration & KNIME screenshot

Data Analysis and Interpretation

Two hypotheses were formulated within the framework of the computational experiment aimed at validating the new methodology for categorising diagnostic parameters in condition-based maintenance systems, and it was found that:

1. H1 was confirmed. The computational experiment using the KNIME software framework demonstrated that grouping diagnostic parameters based on their importance and interdependencies can be effectively implemented with data reported from SCADA/MES systems. The methodology designed in the experiment streamlined the predictive maintenance processes in Condition-based Maintenance (CBM) systems, showcasing the feasibility and effectiveness of the proposed approach.
2. H2 was confirmed. The computational experiment conducted using the KNIME software framework successfully demonstrated that the new method showcased domain neutrality and was readily accessible to practitioners by utilising open-source software tools.

The experiment emphasised the importance of identifying the optimal number of clusters when using hierarchical clustering. Techniques such as the elbow method, silhouette score, or gap statistics can be utilised to determine the best number of clusters, each offering different insights into the data's structure and aiding decision-making. It could be improved with a new method, possibly by combining multiple evaluation metrics to increase the accuracy and robustness of selecting the number of clusters, thereby further refining the clustering process.

Moreover, the experiment revealed that incorporating domain knowledge and subject matter expertise into data preparation, cleaning, and the clustering process can further enhance the accuracy and relevance of the cluster assignments, leading to more insightful insights and actionable outcomes in predictive maintenance strategies.

Discussion

The results from the computational experiment utilising the KNIME software framework emphasise the potential inherent in the proposed methodology for grouping diagnostic parameters, which augment predictive maintenance strategies within CBM systems.

The successful validation of two hypotheses – H1 and H2 – demonstrates this innovative strategy’s methodological robustness and practical applicability in actual industrial environments. Reducing the number of monitored parameters enables quicker decision-making, allowing prompt interventions that prevent unexpected equipment failures. Furthermore, the results suggest that the proposed methodology can be seamlessly integrated with data generated from SCADA/MES systems. This integration is paramount, as it allows organisations to capitalise on pre-existing data frameworks while improving the precision of predictive analytics. The capability to effectively utilise historical and real-time data strategically positions the proposed approach as a significant asset for industries aiming to evolve from reactive to proactive maintenance paradigms.

The validation of H2 highlights the impartiality and accessibility inherent in the proposed methodological framework. By employing open-source software platforms such as KNIME, the methodology ensures that practitioners from various industries can implement it without needing expensive proprietary software solutions. This democratisation of technological resources is particularly relevant in today’s rapidly evolving industrial landscape, where small and medium enterprises (SMEs) often face challenges in adopting advanced predictive maintenance strategies due to financial constraints. The modular design of KNIME allows users to customise workflows to suit their specific requirements, making it applicable to diverse operational scenarios. This flexibility enhances the practicality of the methodology and encourages broader acceptance across multiple sectors, ultimately leading to improved operational efficiency and reduced maintenance costs.

The findings of this study pave the way for future innovations in predictive maintenance, contributing to the ongoing evolution of industrial practices in the era of Industry 4.0. However, the study has several fundamental limitations that warrant consideration:

1. The sample size used in the study was fairly limited, which could restrict the broader applicability of the results. Additionally, relying on self-reported data may introduce systematic bias, as participants might have varying levels of understanding of the principles of predictive maintenance. Future re-

search should aim to address these issues by using larger, more diverse samples and employing objective performance measures to verify the findings.

2. The effectiveness of the proposed framework heavily depends on the quality and thoroughness of the data extracted from SCADA and MES systems. When data is unbalanced, erratic, or inaccurate, the predictive maintenance approach’s effectiveness may be compromised, leading to unreliable predictions and maintenance schedules. Although the methodology has been tested within a specific industrial setting, its applicability across different industries or machinery types may be limited. Variations in equipment classifications, operational contexts, and maintenance practices could impact the transferability of findings and the success of the proposed parameter categorisation strategy. Despite using the KNIME platform for hierarchical clustering, managing large datasets can pose challenges due to high computational demands. In real-world applications, the speed and efficiency of the technique might be hindered by vast data volumes, potentially causing delays in real-time decision-making.
3. The applied methodology assumes that the relationships between diagnostic parameters are linear, which may not always be the case in complex industrial systems. Non-linear interdependencies among parameters could lead to suboptimal classifications, reducing accuracy in predictive maintenance outcomes.
4. Although the study demonstrates the methodology’s effectiveness through computational simulations, empirical validation in various operational settings is essential to establish its practical relevance. The absence of comprehensive field trials may reduce confidence in the methodology’s performance across diverse industrial environment conditions.
5. The categorisation of diagnostic parameters based on historical data may cause overfitting, where the model performs well with the training data but struggles to generalise to new, unseen data. This limitation highlights the importance of continuous monitoring and refining the predictive maintenance strategy as new data becomes available.
6. While the study highlights the methodology’s accessibility via open-source platforms, effective implementation still requires data analysis and machine learning skills. Practitioners lacking sufficient technical knowledge might struggle to apply the proposed method. In conclusion, although the study significantly advances predictive maintenance strategies, these limitations emphasise the

need for further research and validation to ensure the robustness and applicability of the proposed methodology across various industrial settings.

Conclusions

The research aimed to validate the innovative methodology's effectiveness and applicability in enhancing predictive maintenance processes in CBM systems by formulating and testing these hypotheses. The results served to confirm the following:

- The adaptability and accessibility of the methodology across different domains and its compatibility with commonly used software tools,
- The practicality and feasibility of the proposed methodology in real-world industrial settings.

A systematic methodology has been developed and validated to categorise diagnostic parameters that characterise the operational status of machinery and equipment, facilitating predictive analysis. This innovative strategy reduces the number of critical parameters requiring predictive evaluation and monitoring. By systematically arranging these parameters based on their significance and interrelations, the methodology provides a foundation for a more efficient and effective predictive maintenance approach, thereby enhancing the overall value of the research. Additionally, integrating advanced machine learning algorithms, such as deep learning or reinforcement learning, could offer a more sophisticated and nuanced way to perform cluster analysis, potentially revealing hidden patterns and relationships within the data that traditional methods might miss.

Acknowledgements

The authors would like to thank AIUT Sp. z o.o. for their invaluable support in providing access to the SCADA system data used in this research. We appreciate their collaboration, which has significantly contributed to the study.

References

- Berthold, M.R., Cebon, N., Dill, F., Gabriel, T.R., Köter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2008). KNIME: The Konstanz Information Miner. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications* (pp. 319–326). Springer. DOI: [10.1007/978-3-540-78246-9_38](https://doi.org/10.1007/978-3-540-78246-9_38)
- Calzavara, G., Oliosi, E., & Ferrari, G. (2021). A Time-aware Data Clustering Approach to Predictive Maintenance of a Pharmaceutical Industrial Plant. *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 454–458. DOI: [10.1109/ICAIIIC51459.2021.9415206](https://doi.org/10.1109/ICAIIIC51459.2021.9415206)
- Chandna, M. (2024). Monitoring and Prediction of Supervision System for Industrial and Manufacturing Sectors Using Cloud Computing and Blockchain Technology. *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, 1–6. DOI: [10.1109/IC3SE62002.2024.10592888](https://doi.org/10.1109/IC3SE62002.2024.10592888)
- EN 13306:2017 – *Maintenance – Maintenance terminology*. (2024). Retrieved 23 October 2024, from <https://standards.iteh.ai/catalog/standards/cen/5af77559-ca38-483a-9310-823e8c517ee7/en-13306-2017>
- Gallego Garcia, S., & García, M. (2019). Industry 4.0 implications in production and maintenance management: An overview. *Procedia Manufacturing*, 41, 415–422. DOI: [10.1016/j.promfg.2019.09.027](https://doi.org/10.1016/j.promfg.2019.09.027)
- Ibrahim, A.A.E.-H., Hashad, A.I., & Shawky, N.E.M. (2016). Performance Analysis of Various Open Source Tools on Four Breast Cancer Datasets using Ensemble Classifiers Techniques. *International Journal of Engineering Research and Technology*, 5(3).
- Jardine, A.K.S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510. DOI: [10.1016/j.ymssp.2005.09.012](https://doi.org/10.1016/j.ymssp.2005.09.012)
- Karlsson, A., Bekar, E.T., & Skoogh, A. (2021). Multi-Machine Gaussian Topic Modeling for Predictive Maintenance. *IEEE Access*, 9, 100063–100080. DOI: [10.1109/ACCESS.2021.3096387](https://doi.org/10.1109/ACCESS.2021.3096387)
- Kaundal, R., Soni, S.K., & Rajguru, S. (2024). SCADA-Enhanced Real-Time OEE Visualization Driving Industry 4.0 Advancements. *2024 International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES)*, 1–6. DOI: [10.1109/ICSSES62373.2024.10561431](https://doi.org/10.1109/ICSSES62373.2024.10561431)
- Kim, H.-G., Yoon, H.-S., Yoo, J.-H., Yoon, H.-I., & Han, S.-S. (2019). Development of Predictive Maintenance Technology for Wafer Transfer Robot using Clustering Algorithm. *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, 1–4. DOI: [10.23919/ELINFOCOM.2019.8706485](https://doi.org/10.23919/ELINFOCOM.2019.8706485)
- Lee, J., Ni, J., Singh, J., Jiang, B., Azamfar, M., & Feng, J. (2020). Intelligent Maintenance Systems and Predictive Manufacturing. *Journal of Manufacturing Science and Engineering*, 142(110805). DOI: [10.1115/1.4047856](https://doi.org/10.1115/1.4047856)

- Mallioris, P., Aivazidou, E., & Bechtsis, D. (2024). Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. *CIRP Journal of Manufacturing Science and Technology*, 50, 80–103. DOI: [10.1016/j.cirpj.2024.02.003](https://doi.org/10.1016/j.cirpj.2024.02.003)
- Metzger, A., Leitner, P., Ivanović, D., Schmieders, E., Franklin, R., Carro, M., Dustdar, S., & Pohl, K. (2015). Comparing and Combining Predictive Business Process Monitoring Techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(2), 276–290. IEEE Transactions on Systems, Man, and Cybernetics: Systems. DOI: [10.1109/TSMC.2014.2347265](https://doi.org/10.1109/TSMC.2014.2347265)
- Mobley, R. K. (2002). *An Introduction to Predictive Maintenance* (2nd ed.). Butterworth-Heinemann.
- Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, 31(3), 274–295. DOI: [10.1007/s00357-014-9161-z](https://doi.org/10.1007/s00357-014-9161-z)
- Nunes, P., Santos, J., & Rocha, E. (2023). Challenges in predictive maintenance – A review. *CIRP Journal of Manufacturing Science and Technology*, 40, 53–67. DOI: [10.1016/j.cirpj.2022.11.004](https://doi.org/10.1016/j.cirpj.2022.11.004)
- PlantUML Web Server (2024). Retrieved 23 October 2024, from <https://www.plantuml.com/plantuml/uml/SyffKj2rKt3CoKnELR1Io4ZDoSa700003>