STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, April 2026, pp 2743–2758.
Published online in International Academic Press (www.IAPress.org)

Predictive Maintenance in Industrial Systems Using Machine Learning : A Review

Oussama Benmansour [1,*], Ibtissam Medarhri [1], Mohamed Hosni [2]

[1]*MMCS Research Team, LMAID, ENSMR, Rabat, Morocco*
[2]*IEST Research Team, AIDTM Laboratory, ENSAM, University Moulay Ismail of Meknes*

Abstract Predictive maintenance (PdM) has been an important strategy in modern industry, especially with the use of Machine Learning (ML) techniques to enhance equipment reliability and reduce unplanned downtime. In contrast to old and traditional maintenance strategies, that were mainly relying on reactive or scheduled interventions, PdM provides a real-time defect detection and also failure prediction through a complete environment of sensor data records. Many recent studies highlight the effectiveness of ML techniques for optimizing intervention tasks. In this study, we present a systematic mapping study (SMS) of ML classification techniques in industrial contexts. A total of 166 articles in industry and manufacturing published between the year 2000 and 2024 were identified from Scopus digital Library, after a selection process. The findings emphasize an important aspect which is that the fault diagnosis subject is frequently investigated, with Random Forest (RF) being the predominant ML classifier with 64 appearances, followed by Support Vector Machine (SVM) with 55 uses. Also, recent research highlights the increasing role of Deep Learning (DL) in PdM via the use of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) with 28 and 17 appearances respectively. We have also measured the performance of the ML and DL models across the studied papers, by calculating the average performance metric for each model, thus providing a broader explanation and a clearer view on the use of each model. Although many papers did not explicitly specify the datasets used, we found that 85.2% of the papers that have cited their dataset have used real world datasets, thus assuring practicability. As far as the metrics are concerned, Accuracy is the most dominant metric with 100 occurrences, followed by Precision with 61 uses, Recall 57 uses and F1-score 37 uses. The most used tools are Python with 107 occurrences, R with 40 and MATLAB with 20. These findings show that there is a need for publicly available datasets, as well as the development of alternative classification techniques to advance industrial AI PdM applications.

Keywords Predictive Maintenance; Industrial Applications; Machine Learning; Classification; Deep Learning

DOI: 10.19139/soic-2310-5070-3058

1. Introduction

Maintenance is defined as the set of activities required to keep plant machinery, equipment, and systems in optimum operating conditions [1]. In literature, different categories of maintenance management strategies can be identified. Carvalho et al in his work categorize maintenance strategies into three main categories: Run-to-Failure (R2F), Preventive Maintenance (PvM), and Predictive Maintenance (PdM) [2]. R2F, or corrective maintenance, involves repairing equipment only after failure. PvM schedules maintenance at fixed intervals based on time or usage. PdM is a data-driven approach that leverages real-time monitoring and predictive models, such as machine learning (ML), to forecast failures, allowing maintenance to be performed only when needed. This method optimizes asset availability, reduces costs, and aligns with Industry 4.0 practices [3]. Rotating machinery plays a central role in

*Correspondence to: Oussama Benmansour (Email: oussama.benmansour1@outlook.com). MMCS Research Team, LMAID, ENSMR, Rabat, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

industrial manufacturing but is constantly subjected to wear, misalignment, and operational stresses, which makes maintenance a critical concern. Predictive maintenance (PdM) has emerged as a data-driven solution within the Industry 4.0 framework, enabling early fault detection and reducing unexpected downtime [4]. It integrates sensor data readings, statistical analysis, and Artificial Intelligence models to assess equipment's health. However, a key limitation lies in transparency: many PdM implementations depend on black-box ML models that are difficult to interpret [4]. Among diagnostic approaches, vibration analysis remains widely applied, using signal behavior to detect faults in bearings, rotors, and belt drives [5]. Time-domain signals contain valuable information but are often noisy, which has led many researchers to shift toward frequency-domain representations such as Fast Fourier Transformation (FFT) provides clearer fault signatures [4]. The data is generally acquired relying on sensors such as : accelerometers, thermal probes, and electrical sensors. ML and Deep Learning (DL) methods such as Random Forest (RF), Support Vector Machines (SVM), DTs in general, and Convolutional Neural Networks (CNNs) have consistently achieved strong results in both fault classification and remaining useful life prediction [5, 6]. Yet, the limited interpretability of these models forms many challenges in industrial settings where decisions involve significant financial and operational consequences. In order to address this interpretability issue, Explainable Artificial Intelligence (XAI) approaches such as SHAP and LIME are increasingly applied to show the features driving model outputs, thus enhancing trust and supporting decision-making [1, 3]. Recent studies also emphasize the role of feature selection in improving model performance and reducing complexity. Techniques such as Principal Component Analysis (PCA), Minimum Redundancy Maximum Relevance (mRMR), and Neighborhood Component Analysis (NCA) help isolate the most relevant variables, boosting prediction accuracy and at the same time enabling smarter sensor selection strategies. In some cases, combining feature selection with classifiers like RF can achieve precision and recall scores close to 98%, showcasing the impact of dimensionality reduction [7]. To address data privacy and decentralization needs, Federated Learning (FL) has been explored as an alternative to centralized ML, particularly in distributed industrial systems such as maritime fleets [8]. FL enables multiple nodes to collaboratively train models without exchanging raw data, by doing this, they preserve confidentiality while still supporting real-time maintenance predictions [8]. PdM applications are also expanding beyond traditional industrial domains. For example, in cultural heritage conservation, IoT-based systems combined with ML methods like Auto Encoders and k-Nearest Neighbors are widely applied to assess historic structures and predict degradation, facilitating more targeted and timely preservation strategies [9]. In our study, we focus on publications released between 2000 and 2024 that specifically investigate ML models for PdM across a range of industrial domains. The papers that we have selected are analyzed through six mapping questions that capture the essential dimensions of current research practices: (1) publication trends, to trace the evolution and dissemination of PdM research; (2) the ML models employed; (3) the datasets used, whether drawn from industrial sensors or simulations; (4) the evaluation metrics applied, such as precision, recall, and F1-score; (5) the tools and frameworks reported for model development and deployment; (6) the validation strategies, including cross-validation, train–test splits, and real-world testing.

The two main contributions of this paper are:

1. The identification of the papers that investigate classification techniques in predictive maintenance published between the years 2000 and 2024.
2. The analysis of the identified papers according to six criteria: (1) publication trends; (2) ML numerical models used; (3) the datasets used; (4) the metrics used to evaluate the models; (5) the tools used (6) the methods of validation used.

The rest of this paper is organized as follows: Section 2 describes the research methodology followed in our study. Section 3 presents the findings and results. Section 4 opens up a discussion and analysis on study implications. The conclusion and directions for future work are discussed in the last section.

2. Research Methodology

This paper is undertaken as a systematic mapping study (SMS) following the guidelines proposed by Petersen et al. [10, 11, 12, 13]. The mapping process was structured into five distinct phases which are described in upcoming sections.

2.1. Formulation of Mapping Questions (MQs):

To develop a structured perspective on the application of ML techniques in industrial PdM, we defined six targeted mapping questions (MQs). These MQs are intended to reveal patterns concerning research dissemination, modeling approaches, dataset usage, evaluation metrics, software tools, and validation practices. They form the analytical backbone of our study and provide a systematic framework for classifying and interpreting the selected literature. The complete list of MQs considered in this work is presented in Table 1 along with their motivations.

2.2. Literature Search Strategy

The initial set of papers was carefully selected to guarantee both relevance and quality. Studies were excluded if they were not in written in English or if they focused on regression only or non-ML or DL approaches. This process followed by the inclusion and exclusion criteria, this ensures minimal bias and guarantees that only work on ML/DL classification in industrial PdM was considered. The purpose of our literature search was to systematically gather research papers of model-based ML methods for PdM in industrial settings. To do this, we adopted a three-step strategy:

2.2.1. Search String

A structured search query was developed by combining domain-relevant keywords such as "predictive maintenance," "machine learning,", "deep learning" and other relevant keywords. Boolean logic and database-specific syntax were employed to ensure inclusivity and precision. This rigorous search approach was designed to capture a comprehensive and representative sample of the already existing research papers that addresses the topic of PdM, laying the groundwork for the subsequent stages of selecting, key-wording, and data extraction. The final search string used was :
("Artificial intelligence" OR "Machine Learning" OR "Deep Learning") AND ("predictive maintenance") AND (classification).

2.2.2. Literature Resources

This step consists of identifying the candidate papers that can answer the MQs listed in Table 1. The search was carried out in Scopus library. This automatic search was performed using three restrictions [10] :

1. The automatic search covered only the title, abstract and keywords.
2. The starting date is from the year 2000
3. The research was launched only amongst English papers.

The automatic search was conducted in July, 25th 2025.

2.2.3. Search Process

The search process began with an automatic query that identified the initial set of candidate papers. A rigorous selection procedure was then applied to determine which of these were truly relevant (Section 2.3). To ensure completeness, we also used snowballing to capture any additional studies that the automatic search might have missed, including those indexed in other databases. Through this combined approach, our study was able to cover the widest possible range of research addressing ML techniques in industrial contexts.

2.3. Study Selection

In this stage, the relevant papers are selected from the pool of candidate papers that were gathered by means of the automatic search. Relevant papers are those that will be used to address our MQs. The candidate papers were, therefore, examined carefully. Inclusion and exclusion criteria were employed to classify the papers as retain or discard. The inclusion/exclusion criteria are listed below.

(a) Inclusion criteria:

1. Papers that address ML used in PdM Industrial applications (mandatory).
2. Papers performing a comparison with other classification techniques.
3. For duplicate papers the most recent at the time of the search was selected.

(b) Exclusion criteria:

1. Papers written in other languages than English.
2. Papers not discussing Classification techniques.

Note that at least one of the inclusion criteria, besides the mandatory criterion, had to be satisfied to include a paper. A paper was, however, excluded if it satisfied one of the exclusion criteria. The title, abstract and keywords were examined, and also the full text if necessary.

2.4. Data Extraction

Finally, data was extracted according to six predefined criteria following the form of Table 1, enabling both qualitative and quantitative synthesis. Extraction fields included publication metadata, ML models applied, dataset descriptions, performance metrics, tools used, and validation techniques. The extracted data was collected, sorted and visualized. By adhering to this structured SMS methodology in our study, we ensure the reproducibility, traceability, and objectivity of our work, offering a high-confidence overview of ML classification techniques in PdM.

Table 1. Data extraction form

ID	Mapping Questions
MQ1	In which years were the selected papers published?
MQ2	What machine learning models are used in the selected studies?
MQ3	What datasets are used to train and evaluate the models?
MQ4	What metrics are used to evaluate model performance?
MQ5	What tools or platforms are used to develop and evaluate the models?
MQ6	What validation methods are used to assess the proposed models?

2.5. Data Synthesis

The goal of data synthesis is to summarize the extracted data related to each MQ. A synthesis step was used to discuss the results obtained as regards each MQ. Some charts generated by means of MS Excel were used to facilitate the tabulating and visualization of the results.

3. RESULTS

3.1. Overview of the study selection process

A total of 658 articles were initially retrieved from Scopus. Following a systematic screening based on predefined inclusion and exclusion criteria, as shown in Figure 1, 166 articles were identified as relevant to our study. These were used to address the six mapping questions.

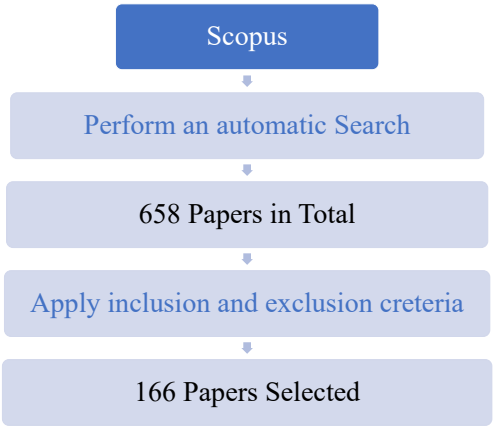

Figure 1. Search Process

3.2. *MQ1: In which years were the selected papers published?*

The publication trends in Figure 2 reveal a clear rise in interest toward ML-based PdM after 2017. A sharp increase appears in 2018, with activity reaching its peak in 2020 at 35 publications. This surge coincides with the spread of IoT-enabled sensing and the wider adoption of Industry 4.0 practices [14]. In the following years, 2021 and 2022, the output stabilized at 27 and 30 papers annually respectively. A noticeable dip in 2023, with only 18 publications, may reflect post-pandemic adjustments that shifted resources away from research-intensive initiatives. The number of research papers soon started increasing again in 2024 reaching a total of 30 papers.

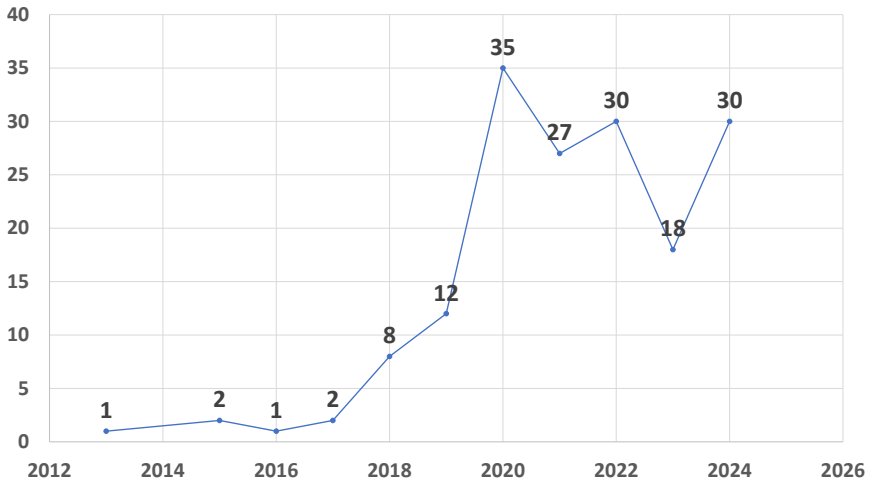

Figure 2. Number of publications per Year

3.3. *MQ2: What are the most frequently applied classification techniques in the literature of PdM in Industry?*

Figure 3 indicates that RF is the most widely applied model, appearing in 64 of the 166 reviewed studies. Its popularity stems from its robustness against noisy inputs and its ability to rank feature importance, a valuable property when dealing with high-dimensional industrial datasets. SVM is reported in 55 studies, particularly in earlier phases of research where data availability was more limited. DTs are used in 47 papers, often paired with boosting techniques to enhance predictive accuracy. DL methods, especially CNNs and LSTMs architectures,

appear in 45 studies and are steadily gaining ground, largely because they can process raw vibration and time-series signals directly, reducing the need for extensive preprocessing. XGBoost was also investigated in selected studies, with 17 studies highlighting their efficiency and strong performance.

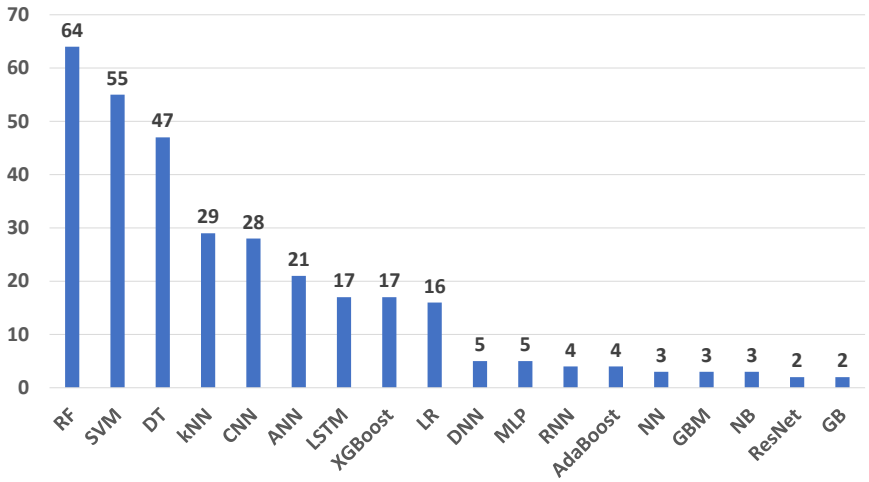

Figure 3. ML classification techniques used

3.4. MQ3 : What is the distribution of the Datasets used in this study ?

As illustrated in Table 2, 85.2% of the reviewed papers that have cited their datasets, rely on real-world industrial datasets. This strong dominance points to a field moving beyond theoretical exploration toward practical deployment [15] . While real datasets bring inevitable challenges such as noise, class imbalance, and sensor drift, they also provide the contextual richness that synthetic sources cannot match. The remaining 14.8% of studies that have mentioned their datasets, make use of simulated or synthetic datasets, which are useful for controlled benchmarking but less reflective of operational realities. Public repositories such as NASA's C-MAPSS and the PHM Society Challenge datasets continue to play a central role in ensuring reproducibility across studies. In our study, we emphasize that although these benchmarks are valuable, expanding access to proprietary industrial data will be important and beneficial for advancing PdM methods that remain both scientifically rigorous and industrially relevant.

Table 2. Distribution of Datasets

Type of Dataset	Occurrence
Real world	34
Synthetic / Simulated	8
CWRU	4
C- MAPSS	3
IMS	2
MaFaulDa	2
Krups EA8108 PYZOFLEW	1

3.5. MQ4 : What are the metrics used by the papers to calculate the precision of the used models ?

Figure 4 summarizes the distribution of evaluation metrics, with a total of 286 mentions across the reviewed papers. Four metrics dominate this landscape. Accuracy leads with 100 occurrences. Precision (61) and Recall (57) follow closely, and their near parity suggests that they are often reported in tandem. F1-score, with 37 mentions, is also

a common choice, though less frequent than precision and recall. Beyond this core group, the frequency drops sharply. The confusion matrix appears in only 8 studies, while AUC is reported 7 times. Error-oriented metrics are relatively rare: MSE and RMSE each appear 4 times, and MAE only 3. Other low-frequency measures include TPR (2). When grouped, the dominance of classification metrics becomes clear: accuracy, precision, recall, and F1 together account for nearly 90% of all mentions. Taken together, this points to a research community that continues to rely heavily on a narrow set of straightforward classification metrics, with limited diversification toward alternatives. As noted by Guyon and Elisseeff [16], choosing the right metric is essential for fair model evaluation. The over reliance on accuracy can obscure model weaknesses, especially in imbalanced PdM datasets, broader adoption of metrics such as precision, recall, and F1 is necessary to capture the true performance of predictive models. An in depth analysis will be given further in our paper, based on the results of Table 3

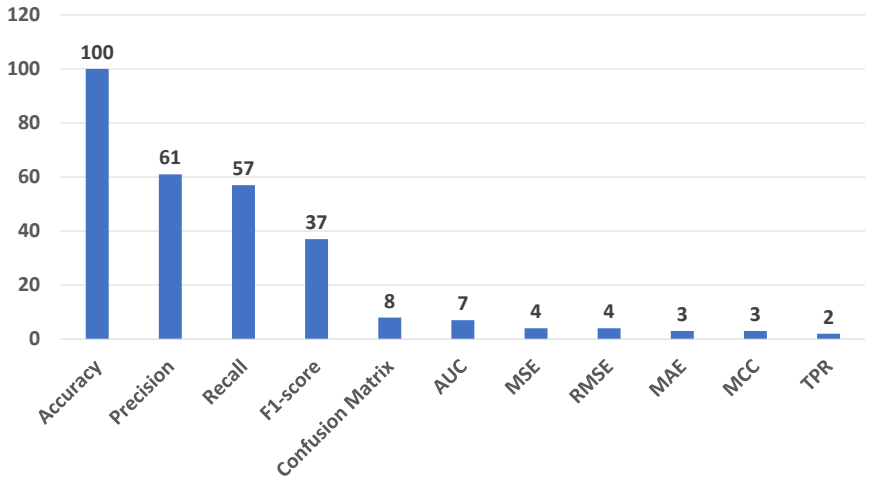

Figure 4. Evaluation Metrics

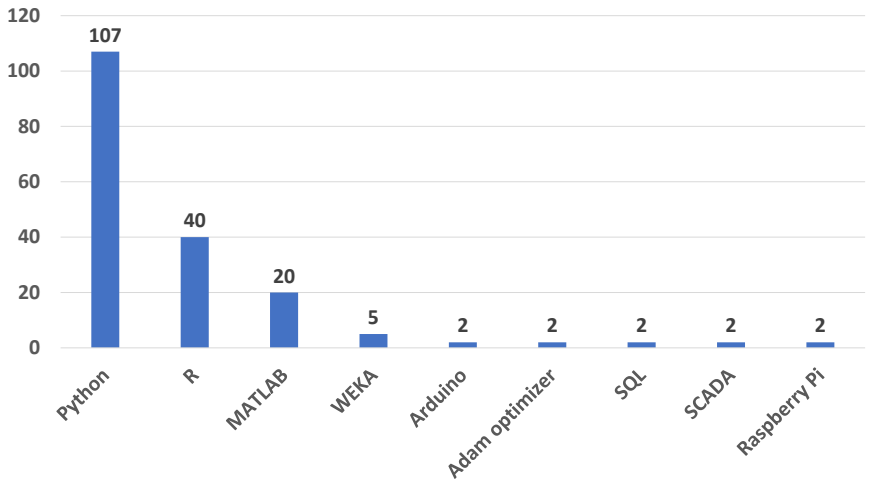

Figure 5. Tools

3.6. MQ5 : What tools are used in the papers ?

Figure 5 highlights a clear concentration on just a few tools, with Python leading by a wide margin at 107 mentions. It stands out as the default environment in most of the reviewed studies. R follows with 40 mentions, giving it a

strong presence in more statistically oriented work, while MATLAB appears 20 times, reflecting its traditional role in engineering and signal processing. Taken together, these three account for the vast majority of reported tools. After them, the numbers fall off quickly. WEKA is used in only 5 studies, and several other tools (Arduino, Adam optimizer, SQL, SCADA, and Raspberry Pi) are each used twice. These outliers suggest occasional experimentation with embedded hardware, optimizers, or industrial control systems rather than widespread adoption. Overall, more than 80% of the mentions belong to Python, R, and MATLAB, with Python alone making up over half. The rest form a long tail of rarely used options, showing that the field has converged on a few dominant platforms while other environments play only a marginal role.

3.7. MQ6 : What validation methods used ?

Figure 6 shows that validation in PdM still revolves around a handful of familiar methods. The 70/30 train–test split comes up the most, with 19 adaptations, and it holds its place largely because it's easy to apply and widely recognized. Close behind, 10-fold cross-validation appears in 16 studies, while MCCV is used 9 times. These methods are chosen when researchers want something more reliable than a single split. The 80/20 division shows up 6 times, a simpler variant but still common enough to suggest that fixed hold-out strategies are part of everyday practice. Beyond these, the numbers drop off quickly. Early stopping and incremental learning each appear only twice, pointing to very specific use cases, usually with DL models or continuous data streams. At the edge of the chart, FFT and SWRL are mentioned once each. Taken together, the figure paints a field that leans heavily on what is simple and familiar. While a few papers experiment with alternatives, they remain rare. In our study, this stood out as a sign that validation choices are often guided less by innovation and more by practicality. Researchers stay with what works, even if it doesn't always give the most nuanced picture of model performance [17].

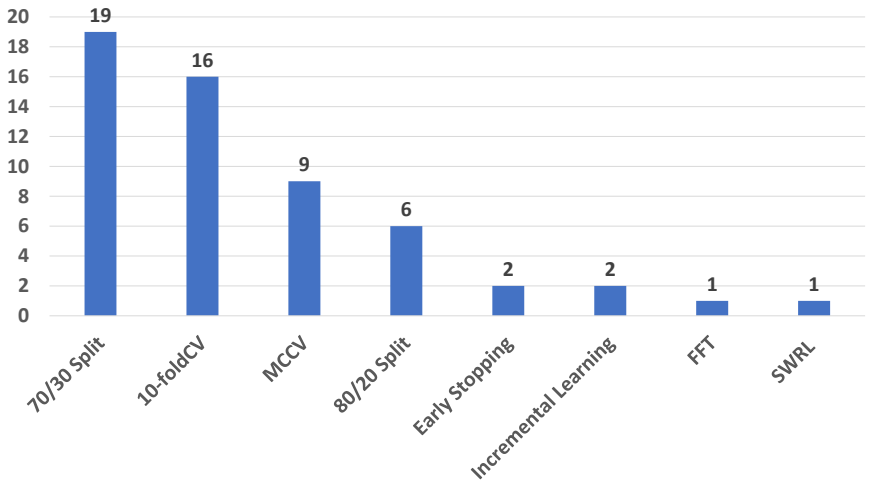

Figure 6. Validation Methods

4. Discussion, study Implications and interpretability

The performance analysis of the numerical models show relatively good results for most well used models [18]. from the results reported in table 3, we can say that despite the superior performance of CNN trained on the CWRU dataset, the classical ML models remain highly relevant for PdM; RF and GB, which are ensemble methods, show strong accuracy and a well balanced precision recall trade off. LR and SVM demonstrate competitive F1 scores and Recall values. These models offer low computational complexity as compared to Deep Learning models, easier hyperparameter tuning and improved transparency [19].

Table 3. Average Performance of ML Models

Models	**Accuracy(%)**	F1-Score(%)	Precision(%)	Recall(%)
Random Forest	93,74125	96,53666667	96,77666667	96,35
SVM	85,6125	89,73	86,175	96,82
Decision Tree	87,07142857	89,47333333	87,87333333	93
KNN	93,24	-	83,78	96,87
PdM-CNN (CWRU)	97,3	97,67	98,5	98
ANN	95,71	95,84	96,575	94,775
XGBoost	93,9875	86,67	92,205	100
Logistic Regression	89,175	94,16	94,08	94,24
DNN	98,03	-	-	-
MLP	99,1	-	-	-
AdaBoost	60,4	33,4	20,64	87,45
Naive Bayes	68	-	-	-
Gradient Boosting	93,75	90,86	90,86	-

4.1. Discussion and study implications

This paper identified 166 papers investigating PdM using ML in different industry fields, published between 2000 and 2024. The papers identified were analyzed according to 6 MQs. The main findings related to each MQ are discussed in this section.

MQ1: In which year were the selected papers published?

The publication trend between 2000 and 2024 shows three distinct phases. From 2000 up to 2017, activity was minimal which reflects a field still in its infancy, fragmented and mostly exploratory. In fact, only a single study appeared in both 2013 and 2016. A noticeable shift occurred in 2018 and 2019, when the number of papers climbed from 2 to 8 and then 12. This sudden rise suggests that interest in PdM began to expand quickly, supported by improved methods, more accessible datasets, and the first signs of community building, a trend also discussed by Carvalho et al. [2]. From 2020 onward, the domain entered what looks like a maturity phase. In 2020, output peaked at 35 papers. Zhang et al. [15] also identify this moment as a turning point, linking it to the industrial adoption of ML-based PdM solutions. Since then, the numbers have stayed relatively high: 27 publications in 2021, 30 in 2022, a dip to 18 in 2023, and then back up to 30 in 2024. The slowdown in 2023 may well reflect post-pandemic adjustments or simply a pause while researchers refined earlier work, but the rebound shows that the momentum has not been lost. In short, most of the research activity happened after 2018, with the peak in 2020 marking PdM's recognition as a serious research area. Today the field is firmly embedded across reliability, ML, aerospace, and industrial IoT, this should be interpreted as a sign that the field has entered a phase of methodological standardization, necessary for further industrial adoption. This consolidation stage echoes what has been described in broader reviews of PdM's maturity [20].

MQ2: What are the most frequently applied classification techniques in the literature of PdM in Industry?

The figure of the most applied classification techniques, figure 3, illustrates the distribution of numerical techniques reported in the PdM literature for industrial applications, showing that a few ML methods dominate while many others remain relatively marginal. RF is the most frequently applied model, appearing 64 times, underscoring its popularity as an ensemble method valued for robustness, its ability to deal with high-dimensional inputs, and its interpretability through feature importance. RF is widely regarded as a reliable baseline, especially in noisy industrial datasets [2]. SVM is the next most common, with 55 mentions, particularly suited to classification tasks where data is limited or features are not linearly separable, which makes it effective for complex sensor signals [15]. DT also appear regularly, with 47 studies. Although simpler, they remain attractive because they are easy to interpret, computationally efficient, and useful in contexts where transparency is essential [20]. A second group

of methods includes KNN (29) and CNN (28). KNN is often used with moderate datasets and similarity-based reasoning, while CNN has become central in handling raw vibration data, spectrograms, and image-like sensor signals, reflecting the shift towards DL models [21]. In the mid-range, ANN (21), LSTM (17), XGBoost (17), and LR (16) each appear. ANN provides nonlinear modeling capabilities, LSTM excels with time-series data, XGBoost is praised for accuracy and efficiency, and LR continues to serve as a simple baseline for binary failure prediction. At the lower end, DNN and MLP (5 each), RNN and AdaBoost (4 each), and NN (3) are present but much less common, partly due to higher computational demands or limited datasets. Rarely used techniques include GBM (3), NB (3), ResNet (2), and GB (2), which are either specialized or overshadowed by stronger alternatives like RF or XGBoost. Overall, the findings confirm that tree-based methods (RF, DT, XGBoost) and SVM dominate due to their interpretability, resilience, and strong performance. However, the rise of CNNs and LSTMs in 28 studies points to an ongoing paradigm shift toward deep learning. Scientifically, this shift represents a move from feature-engineering-driven approaches toward end-to-end architectures capable of automatically extracting fault signatures from raw sensor data. The scientific challenge here is balancing accuracy with interpretability. In our study, we view this duality as a defining feature of PdM research: balancing proven ensemble methods that deliver reliability with DL architectures that open the door to new advances in fault detection and prediction [22].

MQ3 : What is the distribution of the Datasets used in this study ?

The dataset distribution confirms that real-world data dominates PdM research, with 46 instances reported across the reviewed studies. This reliance on industrial case studies underlines the applied character of the field and the need to validate algorithms under realistic conditions where noise, changing operating modes, and diverse fault types are present [23]. These datasets span domains such as manufacturing plants, conveyor systems, woodworking machines, and thermal installations, offering varied contexts for testing. Synthetic or simulated datasets form the second category, used in 8 studies. While they allow researchers to generate controlled degradation patterns, benchmark algorithms under repeatable setups, and explore rare or costly failure modes, their adoption remains limited compared to real data, reflecting the continuing demand for practical validation. Among public benchmarks, the CWRU dataset appears 4 times, NASA's C-MAPSS 3 times, and IMS bearing data 2 times, making them three of the most established datasets in PdM [24]. CWRU is well known for bearing fault classification, C-MAPSS is the reference standard for RUL prediction, and IMS provides vibration data for fault detection. Together, they ensure reproducibility and comparability across studies. The MaFaulDa dataset, though used only twice, is gaining traction as it provides structured labels and diverse operating conditions that facilitate systematic evaluation [25]. Beyond these, several unique domain-specific datasets appear once each, such as conveyor motor data with 11 variables, long-term vibration data from 51 sensors, MTS/UTS plant logs, woodworking machine failure records, and real-world vibration with current data from CeraCon GmbH's thermal systems. A more unusual case is the Krups EA8108 PYZOFLEW vibration dataset, which contains labeled audio data, demonstrating creative use of non-traditional signals for predictive maintenance. Overall, the results suggest three tiers: first, real-world datasets (29 undefined and 17 specified), which emphasize industrial validation; second, simulated data (8) and well-known public benchmarks (CWRU, C-MAPSS, IMS, MaFaulDa, totaling 11) that support reproducibility [26, 27]; and third, a handful of rare, specialized datasets (6 total), reflecting the adaptability of PdM research to different industrial settings. Taken together, this distribution shows that while the datasets remain crucial for comparability, the strength of the field lies in its strong grounding in real-world data, ensuring that PdM methods stay aligned with practical deployment. Also, the lack of diversity in public datasets limits cross-domain generalization. Expanding dataset availability beyond turbofan engines toward multi-component systems (gearboxes, conveyor belts, wind turbines) is essential for the next stage of PdM research.

MQ4 : What are the metrics used by the papers to calculate the precision of the used models ?

The distribution of evaluation metrics across PdM studies shows a strong dependence on a small group of standard classification measures, with accuracy standing out as the most reported. It appears in 100 papers, reflecting its intuitive appeal as the proportion of correct predictions over all cases and its role as a first benchmark of performance in industrial applications [28]. Still, accuracy alone can give a distorted view, particularly in imbalanced datasets where failures are rare compared to normal operation. To mitigate this, precision (61) and

recall (57) are frequently applied. Precision highlights how many predicted failures were correct, which is essential for reducing unnecessary interventions, while recall shows how many actual failures were detected, lowering the risk of missed faults. Their frequent joint use indicates an effort to balance these two perspectives [29]. F1-score (57) also figures prominently, serving as a harmonic mean of precision and recall, and offering a more balanced single measure when both false positives and false negatives carry high costs in PdM [30]. The confusion matrix, with 37 mentions, provides more granular insight by mapping outcomes into true positives, false positives, true negatives, and false negatives, supporting further metric derivation. Beyond this core group, usage drops sharply. AUC (8) is employed when evaluating performance across thresholds, while regression-based metrics such as MSE (5), RMSE (4), and MAE (4) appear mainly in RUL prediction contexts where the magnitude of error matters more than categorical correctness [31]. Less frequently reported measures include MCC (3), valued for balance in imbalanced data, and ROC curves (3), which plot true positive against false positive rates. RUL itself appears 3 times as a metric specific to prognostics, while TPR is rarely reported separately (2).

From a larger analytical perspective, these frequencies highlight a maturity gap between reporting simplicity and problem complexity. Our synthesis of model performance data confirms that most studies reporting only accuracy tend to show inflated results, meaning that high accuracy may not reliably indicate robust fault detection [32, 33]. In contrast, the few studies reporting F1-score, precision, or MCC reveal more nuanced outcomes: for example, some models with near perfect accuracy like CNNs on MaFaulDa and CWRU datasets show lower precision or recall when evaluated on unbalanced datasets. This pattern demonstrates that PdM literature often favors metric convenience over diagnostic depth, thus the importance of relying on studies that use different metrics [2]. This trend suggests that the field is gradually recognizing that reliability in PdM depends less on nominal accuracy and more on consistent sensitivity and specificity across different operating conditions, and referring to all different metrics possible to have a broader view on the model performance [34].

Overall, the evidence points to a heavy reliance on accuracy, precision, recall, and F1-score, often accompanied by confusion matrices, which together form the de facto evaluation standard. At the same time, a smaller set of works incorporate regression-style errors and threshold-independent metrics like AUC and MCC to address particular challenges. This shows that while PdM evaluation practices remain dominated by simple, interpretable measures, there is a gradual move toward more problem-specific and robust approaches [35].

MQ5 : What tools are used in the papers?

The analysis of tools reported in PdM studies shows a strong dominance of Python, which appears 107 times and has clearly established itself as the default environment for machine learning and data science. Its versatility, large ecosystem of libraries such as TensorFlow, PyTorch, Scikit-learn, Pandas, NumPy, and Keras, along with strong community support and open-source accessibility, explains why it is by far the most widely adopted platform across both research and industrial applications [36]. R is the second most popular environment, mentioned in 40 papers, and remains particularly relevant in statistical modeling, visualization, and exploratory data analysis, making it an appealing option for work on reliability, survival modeling, and statistical aspects of PdM [37, 38]. MATLAB follows with 20 occurrences, reflecting its historic importance in engineering and signal processing, areas that are tightly connected to vibration analysis and condition monitoring, and its extensive toolboxes have made it a useful option especially in earlier PdM studies before Python and R took over as standards. Tools outside of this top three are far less common. WEKA, cited 5 times, serves as a classic ML environment mainly for prototyping and teaching, but its limited scalability restricts broader adoption. A few specialized or hardware-oriented platforms also appear in the literature. Arduino and Raspberry Pi, each reported twice, indicate experimental setups where embedded systems are used for data acquisition and real-time monitoring [39]. SCADA and SQL, again with 2 mentions each, highlight integration with industrial control systems and database management [40]. Finally, the Adam optimizer is mentioned twice, not as a standalone platform but as a key component of neural network training, particularly in DL models such as CNNs and LSTMs. Altogether, more than 80 percent of tool mentions are concentrated in Python, R, and MATLAB, with Python alone accounting for over half of the total. This shows how the field has consolidated around a small set of dominant platforms while other tools remain on the margins, used mainly for specialized or experimental purposes. In our study this convergence is striking, reinforcing the

central role of open-source ecosystems led by Python while still acknowledging the complementary contributions of R and MATLAB in the PdM research landscape. From a scientific standpoint, the reliance on open-source Python libraries facilitates reproducibility, but it also raises the risk of over-homogenization. A stronger integration with cloud platforms and edge deployment environments would enhance the external validity of PdM research.

MQ6 : What validation methods used ?

The analysis of validation methods in PdM papers shows that most studies continue to depend on traditional train–test splits and cross-validation to assess model performance. The 70/30 split is the most common, reported 19 times, and remains popular largely because it is simple to apply and widely accepted as a baseline [41]. It offers a workable compromise between training size and test set availability. The second most frequent approach is 10-fold cross-validation, used in 16 studies, which provides a more robust picture by rotating training and testing across all partitions and thereby reducing the risk of variance tied to a single split [42]. MCCV appears in 9 papers and has the advantage of repeatedly generating random partitions, offering greater statistical confidence in the results, a feature particularly useful when datasets are small [43]. The 80/20 split is less common, with 6 mentions, but it shows that hold-out validation remains standard practice, especially in cases where researchers want to maximize training data. Beyond these, the numbers fall off. Early stopping is reported twice, mostly in DL studies where overfitting is a concern, as training is halted when performance plateaus [44]. Incremental learning, also cited twice, appears in scenarios where data arrives in streams and models need to adapt continuously without complete retraining [45]. More unusual cases include FFT and SWRL, each mentioned once, but these are more specialized procedures than mainstream validation methods.

While these frequencies might illustrate methodological preferences, our performance synthesis reveals what they imply: the dominance of internal validation such as the 70/30 split or k-fold on the same dataset is one reason why many studies report near perfect accuracies (more than 98%) for models such as RF, CNN, and XGBoost [46]. These schemes estimate consistency within a dataset but not necessarily across different machines, sensors, or operating conditions. As a result, models validated through single dataset cross validation tend to show inflated accuracy but reduced transferability, a pattern confirmed in our aggregated analysis where accuracy remains high but robustness indicators (F1, precision, MCC) vary more widely [47]. This highlights a structural issue in PdM research: most validation protocols are operationally narrow, providing little insight into how models would behave under real industrial drift or unseen equipment. Only a few studies applying early stopping or incremental learning attempt to address this limitation, pointing to a gradual but insufficient shift towards validation schemes aligned with deep and online learning contexts [48]. Overall, the field continues to favor simple hold-out splits and classical cross-validation because they are easy to use, consistent, and computationally efficient [41]. However, the gap between laboratory validation and industrial deployment remains evident. Future work should prioritize cross-dataset and temporal validation, federated or transfer learning-based assessments, and real-time incremental testing to better reflect actual production environments. This evolution will be critical to transforming PdM validation from statistical convenience to industrial reliability, thereby bridging the final link between predictive accuracy and real-world applicability.

4.2. Interpretability and Emerging Trends in PdM

Recent advances in PdM research indicate a shift from pure performance improvement toward models that are more interpretable, well distributed, and deployable. High accuracy deep learning models such as CNNs and LSTMs have proven to be effective but often operate as black boxes, limiting their adoption in safety-critical industrial contexts. To address this, recent work has emphasized Explainable Artificial Intelligence XAI as a means to increase trust and transparency. Techniques such as SHAP, LIME, and Grad-CAM are increasingly used to visualize how vibration, temperature, or acoustic signals influence predictions, enabling engineers to justify maintenance actions and improve confidence in automated systems [49]. In parallel, FL is emerging as a key strategy to overcome the well known problem of the data-sharing and privacy constraints across industrial sites. Instead of centralizing sensitive sensor data, FL enables multiple plants or devices to collaboratively train models while keeping data local, which enhances generalization across heterogeneous environments and aligns with recent communication-efficient FL

architectures for IIoT applications [50]. Complementing this trend, Edge AI brings predictive intelligence directly to industrial gateways and embedded systems, reducing latency, bandwidth usage, and cloud dependence—key for real-time monitoring in constrained environments. Recent studies integrating Edge AI with cloud computing frameworks demonstrate improved responsiveness and scalability of PdM implementations [51]. Similarly, new developments in industrial AI and edge computing further emphasize that embedding intelligence closer to machines strengthens operational continuity and autonomy [52]. Collectively, these developments indicate that the next generation of PdM systems will prioritize trustworthiness, data sovereignty, and responsiveness as much as accuracy. Integrating XAI, FL, and Edge AI represents the logical progression toward intelligent maintenance systems that are not only precise but also transparent, privacy-preserving, and industrially scalable.

5. Conclusions and Further work

Our systematic mapping study (SMS) analyzed 166 papers published between 2000 and 2024 related to the use of ML classification techniques for predictive maintenance (PdM) in industrial contexts, sourced from the Scopus digital library. The papers were examined according to six mapping questions (MQs), providing a comprehensive overview of research trends, methodological choices, and emerging challenges. Overall, the study confirms that PdM has evolved into a mature and data-driven research area under the Industry 4.0 paradigm. Research activity has intensified significantly since 2017, reaching its peak in 2020, and stabilizing thereafter (MQ1). From a methodological standpoint (MQ2), tree-based models, especially Random Forest (64), SVM (55), and Decision Trees (47) remain the most frequently applied due to their robustness and interpretability. However, deep learning methods such as CNNs (28) and LSTMs (17) show increasing adoption, reflecting a paradigm shift toward end-to-end feature learning. Regarding datasets (MQ3), the analysis clarifies that approximately 85% of the articles that have announced the used datasets, employ real-world industrial data, but others rely on publicly available datasets such as C-MAPSS, CWRU and IMS rather than proprietary factory datasets. This indicates that while PdM research has achieved greater realism, it still depends heavily on standardized datasets. Expanding access to diverse and multimodal industrial data remains a key requirement for future maturity. In terms of performance evaluation (MQ4), accuracy remains the most frequently reported metric (100 mentions) but often provides a distorted view in imbalanced scenarios. Our synthesis revealed that studies reporting only accuracy tend to show inflated scores (often above 98%), whereas those using multi-metric evaluations (Precision, Recall, F1, AUC, MCC) provide more reliable insights into model robustness. This underlines the need for metric pluralism and cost-sensitive evaluation frameworks in future PdM research. Tool usage (MQ5) continues to be dominated by Python (107) due to its extensive libraries and open-source accessibility, with R (40) and MATLAB (20) supporting specialized applications in statistical modeling and signal processing. Validation practices (MQ6) remain a methodological bottleneck: most studies employ internal validation schemes such as train–test splits (70/30, 80/20) and 10-fold cross-validation, which assess internal consistency but not cross-domain generalization. Few works use advanced approaches such as Monte Carlo CV, early stopping, or incremental learning. Future studies should move toward cross-dataset and federated validation protocols to assess true reliability under distribution shift and data heterogeneity. Emerging trends such as Explainable AI (XAI), Federated Learning (FL), and Edge AI are reshaping the PdM landscape. XAI enhances transparency in high-performing models, FL allows collaborative learning across distributed industrial sites while preserving data privacy, and Edge AI enables real-time inference in resource-constrained environments. Integrating these technologies can lead to more trustworthy, decentralized, and interpretable PdM systems. In conclusion, this study provides an analytical synthesis that bridges descriptive mapping with actionable insights. The results emphasize three overarching recommendations for future research:

- Adopt multi-metric and cost-sensitive evaluation to ensure credible and balanced model assessment.
- Diversify datasets and validation schemes to improve generalization and representativeness.
- Leverage interpretability and emerging AI paradigms (XAI, FL, Edge AI) to enhance industrial trust, privacy, and scalability.

By addressing these directions, PdM research can progress from isolated algorithmic advances to operationally reliable, explainable, and industry-ready AI systems that fully embody the goals of the Industry 4.0 era. The ongoing work will focus on consolidating empirical evidence on the relative performance of classical ML methods

versus deep learning architectures in PdM, particularly under conditions of noisy, imbalanced, and heterogeneous industrial datasets. Building on the gaps revealed by this systematic mapping study, several directions are identified for future research:

- Expanding the use of heterogeneous ensembles and hybrid frameworks that combine interpretable models as such Random Forest, Decision Trees with deep architectures like CNNs and LSTMs to balance accuracy with explainability, in line with findings from MQ2 and the emerging trend of XAI.
- Conducting systematic comparisons of pre processing techniques including data balancing, feature selection, dimensionality reduction, and missing data to quantify their influence on model robustness and generalization across industrial conditions.
- Extending performance assessment beyond accuracy by integrating multi-metric evaluation schemes such as AUC, MCC, F1, and cost-sensitive measures, as emphasized in MQ4, to better reflect the asymmetric costs of false positives and missed detections in critical maintenance contexts..
- Developing richer and more diverse datasets beyond traditional benchmarks such as turbofan engines and bearings, incorporating other assets like gearboxes, conveyor belts, and wind turbines to ensure stronger external validity and broader generalization.
- Advancing validation practices through cross-dataset, temporal, and federated validation frameworks to more accurately assess real world reliability and transferability.
- Investigating the integration of Federated Learning, Edge AI, and IoT-based PdM solutions to enable decentralized, privacy-preserving, and low-latency analytics in distributed industrial systems.

From a practical standpoint, industries are encouraged to invest in collaborative data-sharing initiatives that promote the creation of anonymized, diverse public repositories. Stronger partnerships between industrial engineers and AI researchers are essential for developing explainable, robust, and deployable PdM systems—advances that can directly translate into reduced downtime, optimized maintenance schedules, and improved operational trust in AI-driven decision support.

REFERENCES

1. R. K. Mobley, *An Introduction to Predictive Maintenance*, vol. 37. Elsevier, 2002.
2. T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers and Industrial Engineering*, vol. 137, no. April, p. 106024, 2019.
3. W. Zhang, D. Yang, and H. Wang, "Data-driven methods for predictive maintenance of industrial equipment: A survey," *IEEE systems journal*, vol. 13, no. 3, pp. 2213–2227, 2019.
4. S. Gawde, S. Patil, S. Kumar, P. Kamat, and K. Kotecha, "An explainable predictive maintenance strategy for multi-fault diagnosis of rotating machines using multi-sensor data fusion," *Decision Analytics Journal*, vol. 10, p. 100425, 2024.
5. A. A. Shandookh, A. A. F. Ogaili, and L. A. Al-Haddad, "Failure analysis in predictive maintenance: Belt drive diagnostics with expert systems and Taguchi method for unconventional vibration features," *Heliyon*, vol. 10, no. 13, 2024.
6. S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and S. Alfarhood, "Explainable predictive maintenance of rotating machines using LIME, SHAP, PDP, ICE," *IEEE Access*, vol. 12, pp. 29345–29361, 2024.
7. F. E. Bezerra, G. C. de Oliveira Neto, G. M. Cervi, R. Francesconi Mazetto, A. M. de Faria, M. Vido, G. A. Lima, S. A. de Araújo, M. Sampaio, and M. Amorim, "Impacts of feature selection on predicting machine failures by machine learning algorithms," *Applied Sciences*, vol. 14, no. 8, p. 3337, 2024.
8. A. Angelopoulos, A. Giannopoulos, N. Nomikos, A. Kalafatelis, A. Hatziefremidis, and P. Trakadas, "Federated learning-aided prognostics in the shipping 4.0: Principles, workflow, and use cases," *IEEE Access*, vol. 12, pp. 6437–6454, 2024.
9. M. Casillo, F. Colace, R. Gaeta, A. Lorusso, D. Santaniello, and C. Valentino, "Revolutionizing cultural heritage preservation: an innovative IoT-based framework for protecting historical buildings," *Evolutionary Intelligence*, vol. 17, no. 5, pp. 3815–3831, 2024.
10. K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *12th international conference on evaluation and assessment in software engineering (EASE)*, BCS Learning & Development, 2008.
11. I. Medarhri, M. Hosni, M. Ettalhaoui, Z. Belhaj, and R. Zine, "Reviewing Machine Learning Techniques in Credit Card Fraud Detection," in *International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K - Proceedings*, vol. 1, pp. 179–187, 2024.
12. R. S. Wahono, "A systematic literature review of software defect prediction," *Journal of software engineering*, vol. 1, no. 1, pp. 1–16, 2015.
13. C. Marshall and P. Brereton, "Tools to support systematic literature reviews in software engineering: A mapping study," in *2013 ACM/IEEE international symposium on empirical software engineering and measurement*, pp. 296–299, IEEE, 2013.

14. A. Rejeb, K. Rejeb, E. Süle, A. Hassoun, and J. G. Keogh, "Knowledge flows in industry 4.0 research: a longitudinal and dynamic analysis," *Journal of Data, Information and Management*, pp. 1–23, 2025.
15. W. Zhang, D. Yang, H. Wang, and W. Wang, "Data-driven predictive maintenance for Industry 4.0: A survey," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3125–3135, 2019.
16. I. Iguyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
17. D. Berrar and Others, "Cross-validation.," 2019.
18. J. C. Obi, "A comparative study of several classification metrics and their performances on data," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308–314, 2023.
19. J. a Ilemobayo, O. Durodola, O. Alade, O. J Awotunde, A. T Olanrewaju, O. Falana, A. Ogungbire, A. Osinuga, D. Ogunbiyi, A. Ifeanyi, *et al.*, "Hyperparameter tuning in machine learning: A comprehensive review," *Journal of Engineering Research and Reports*, vol. 26, no. 6, pp. 388–395, 2024.
20. A. Bousdekis, B. Magoutas, D. Apostolou, and G. Mentzas, "Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance," *Journal of Intelligent Manufacturing*, vol. 29, no. 6, pp. 1303–1316, 2018.
21. C. Ding, G. Wang, X. Zhang, Q. Liu, and X. Liu, "A hybrid CNN-LSTM model for predicting PM2. 5 in Beijing based on spatiotemporal correlation," *Environmental and Ecological Statistics*, vol. 28, no. 3, pp. 503–522, 2021.
22. S. Ayvaz and K. Alpay, "Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time," *Expert Systems with Applications*, vol. 173, p. 114598, 2021.
23. Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical systems and signal processing*, vol. 138, p. 106587, 2020.
24. A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 international conference on prognostics and health management*, pp. 1–9, IEEE, 2008.
25. W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mechanical systems and signal processing*, vol. 64, pp. 100–131, 2015.
26. K. Medjaher, D. A. Tobon-Mejia, and N. Zerhouni, "Remaining useful life estimation of critical components with application to bearings," *IEEE Transactions on Reliability*, vol. 61, no. 2, pp. 292–302, 2012.
27. M. A. Marins, F. M. L. Ribeiro, S. L. Netto, and E. A. B. Da Silva, "Improved similarity-based modeling for the classification of rotating-machine failures," *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 1913–1930, 2018.
28. A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
29. A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical systems and signal processing*, vol. 23, no. 3, pp. 724–739, 2009.
30. Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
31. K. Goebel, M. J. Daigle, A. Saxena, I. Roychoudhury, S. Sankararaman, and J. R. Celaya, *Prognostics: The science of making predictions*. 2017.
32. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *The journal of machine learning research*, vol. 15, no. 1, pp. 3133–3181, 2014.
33. N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent data analysis*, vol. 6, no. 5, pp. 429–449, 2002.
34. E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo, "Combining evaluation metrics via the unanimous improvement ratio and its application to clustering tasks," *Journal of Artificial Intelligence Research*, vol. 42, pp. 689–718, 2011.
35. J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manufacturing letters*, vol. 3, pp. 18–23, 2015.
36. A. C. Müller and S. Guido, *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.", 2016.
37. G. Sharma and J. Martin, "MATLAB®: a language for parallel computing," *International Journal of Parallel Programming*, vol. 37, no. 1, pp. 3–36, 2009.
38. R. Ihaka and R. Gentleman, "R: a language for data analysis and graphics," *Journal of computational and graphical statistics*, vol. 5, no. 3, pp. 299–314, 1996.
39. C. Shalini and I. V. M. Prakash, "IoT based industrial sensor monitoring and alerting system using Raspberry Pi," in *IOP Conference Series: Materials Science and Engineering*, vol. 981, p. 42010, IOP Publishing, 2020.
40. M. Nuruzzaman and S. Rana, "IoT-Enabled Condition Monitoring in Power Distribution Systems: A Review of Scada-Based Automation, Real-Time Data Analytics, and Cyber-Physical Security Challenges," *Journal of Sustainable Development and Policy*, vol. 1, no. 01, pp. 25–43, 2025.
41. R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, pp. 1137–1145, Montreal, Canada, 1995.
42. S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," 2010.
43. Q.-S. Xu, Y.-Z. Liang, and Y.-P. Du, "Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 2, pp. 112–120, 2004.
44. L. Prechelt, "Early stopping-but when?," in *Neural Networks: Tricks of the trade*, pp. 55–69, Springer, 2002.
45. J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM computing surveys (CSUR)*, vol. 46, no. 4, pp. 1–37, 2014.
46. H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data mining and knowledge discovery*, vol. 33, no. 4, pp. 917–963, 2019.
47. P. Goodarzi, A. Schütze, and T. Schneider, "Comparing automl and deep learning methods for condition monitoring using realistic validation scenarios," *arXiv preprint arXiv:2308.14632*, 2023.

48. P. Jul-Rasmussen, M. Stevnsborg, X. Liang, and J. K. Huusom, "Implementation of real-time incremental learning for ensemble hybrid model prediction in pilot scale bubble column aeration," *Digital Chemical Engineering*, vol. 14, p. 100212, 2025.
49. A. Ucar, M. Karakose, and N. Kça, "Artificial intelligence for predictive maintenance applications: key components, trustworthiness, and future trends," *Applied Sciences*, vol. 14, no. 2, p. 898, 2024.
50. Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6348–6358, 2020.
51. A. Kota, "PREDICTIVE MAINTENANCE: INTEGRATING EDGE AI WITH CLOUD COMPUTING FOR INDUSTRIAL IOT," *Technology (IJRCAIT)*, vol. 7, no. 2, 2024.
52. A. Bala, R. Z. J. A. Rashid, I. Ismail, D. Oliva, N. Muhammad, S. M. Sait, K. A. Al-Utaibi, T. I. Amosa, and K. A. Memon, "Artificial intelligence and edge computing for machine maintenance-review," *Artificial Intelligence Review*, vol. 57, no. 5, p. 119, 2024.