JOURNAL OF ADVANCED ENGINEERING AND INNOVATION
Volume 4, Issue 1, June 2026 | pp.43-52
https://doi.org/10.54518/jaei.4.1.2026.1474

Data-Driven Predictive Maintenance Using Artificial Intelligence and Machine Learning

Muhammad Ilham Imami[1] ✉

Universitas Muhammadiyah Surabaya, Indonesia[1]

ABSTRACT

Digital transformation in engineering has accelerated the adoption of Data-Driven Predictive Maintenance (PdM) as a maintenance strategy that utilizes operational data to predict equipment failures more accurately. This study aims to examine recent advances in the application of Artificial Intelligence (AI) and Machine Learning (ML) for predictive maintenance through a Systematic Literature Review (SLR). The review analyzes peer-reviewed studies published over the last five years collected from major scientific databases. The analysis focuses on AI and ML algorithms, data sources, model evaluation techniques, implementation benefits, current challenges, and future research directions. The findings indicate that algorithms such as Random Forest, Support Vector Machine, Extreme Gradient Boosting, Convolutional Neural Network, Long Short-Term Memory, and Transformer have demonstrated strong performance in anomaly detection, fault diagnosis, and Remaining Useful Life prediction. While predictive maintenance significantly improves equipment reliability and maintenance efficiency, several challenges remain, including data quality, model interpretability, cybersecurity, and system integration. The study concludes that the integration of AI, ML, Digital Twin, Edge AI, and Explainable Artificial Intelligence offers promising opportunities to enhance prediction accuracy and support the development of adaptive, intelligent, and sustainable maintenance systems in the industry 5.0 era.

Keywords: *Artificial Intelligence, Data-Driven Predictive Maintenance, Industry 5.0, Machine Learning.*

CORRESPONDING AUTHOR:
Muhammad Ilham Imami
Universitas Muhammadiyah Surabaya, Indonesia

ARTICLE HISTORY

Received	: January 7, 2026
Final Revised	: February 24, 2026
Accepted	: April 18, 2026
Published	: June 30, 2026

1. | INTRODUCTION

The rapid advancement of digital transformation has fundamentally reshaped maintenance strategies across modern engineering and industrial systems. Traditional maintenance approaches, including reactive and preventive maintenance, are increasingly being replaced by intelligent maintenance strategies that utilize real-time operational data to improve equipment reliability, reduce maintenance costs, and minimize unexpected downtime. Among these strategies, Data-Driven Predictive Maintenance (PdM) has emerged as a core component of smart manufacturing because it enables maintenance decisions based on the actual health condition of industrial assets rather than predetermined maintenance schedules. This approach aligns closely with the objectives of Industry 4.0 and the emerging Industry 5.0 paradigm, where intelligent decision-making, automation, and data analytics are integrated to enhance operational efficiency and system sustainability (Islam et al., 2025).

The widespread adoption of Industrial Internet of Things (IIoT) technologies has enabled continuous acquisition of operational data from sensors installed in industrial equipment. These sensors generate large volumes of heterogeneous data, including vibration signals, temperature measurements, acoustic emissions, pressure readings, electrical current, and other condition-monitoring parameters. The availability of such high-dimensional datasets has created significant opportunities for applying Artificial Intelligence (AI) and Machine Learning (ML) techniques to identify hidden degradation patterns, detect anomalies, estimate Remaining Useful Life (RUL), and predict equipment failures before catastrophic breakdowns occur. Consequently, predictive maintenance has evolved from simple condition monitoring toward intelligent prognostic systems capable of supporting autonomous maintenance decision-making (Achouch et al., 2022).

Artificial Intelligence has become one of the most influential technologies for predictive maintenance because of its capability to analyze complex and nonlinear relationships among multiple operational variables. Machine Learning algorithms including Decision Trees, Random Forest, Support Vector Machines, Gradient Boosting, Artificial Neural Networks, and Deep Learning architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have demonstrated superior performance in fault diagnosis, anomaly detection, health-state assessment, and predictive analytics. Compared with conventional statistical approaches, these algorithms provide greater adaptability to dynamic industrial environments while continuously improving prediction accuracy through data-driven learning processes. Recent advances in explainable artificial intelligence (XAI), hybrid learning models, and digital intelligence have further strengthened the reliability and transparency of AI-enabled predictive maintenance systems (Rahman et al., 2025).

Despite substantial progress, several technical challenges continue to hinder the widespread deployment of AI-based predictive maintenance systems. Many industrial datasets suffer from data imbalance, missing values, noisy sensor signals, limited failure samples, and heterogeneous data structures, which reduce the robustness and generalizability of prediction models. Furthermore, the lack of model interpretability, insufficient real-world validation, cybersecurity concerns, and integration difficulties with legacy industrial infrastructures remain significant barriers to industrial implementation. Existing studies also tend to concentrate on specific algorithms or application domains, making it difficult to obtain a comprehensive understanding of the overall evolution of data-driven predictive maintenance technologies. These limitations highlight the necessity for a systematic synthesis of recent research developments to identify current trends, methodological advances, existing research gaps, and future opportunities (Cheng et al., 2026).

Therefore, this study conducts a Systematic Literature Review (SLR) of research published over the last five years on data-driven predictive maintenance using Artificial Intelligence and Machine Learning. The review aims to analyze the evolution of predictive maintenance technologies, identify the most widely adopted machine learning techniques, evaluate their advantages and limitations, and examine emerging research directions such as explainable AI, hybrid intelligent systems, digital twins, and next-generation predictive analytics. The findings are expected to provide a comprehensive academic reference for researchers while offering practical insights for engineers and industrial practitioners seeking to develop more reliable, scalable, and intelligent predictive maintenance systems for future smart manufacturing environments.

2. | RESEARCH METHOD

This study employed a Systematic Literature Review (SLR) to comprehensively investigate recent advances in Data-Driven Predictive Maintenance (PdM) using Artificial Intelligence (AI) and Machine Learning (ML). The SLR methodology was selected because it provides a rigorous, transparent, and reproducible approach for identifying, evaluating, and synthesizing scientific evidence from previously published studies. Unlike conventional narrative reviews, an SLR follows a predefined review protocol that minimizes selection bias and enables a systematic assessment of existing knowledge. Through this approach, the study aims to summarize current technological developments, compare different AI and ML techniques, identify implementation challenges, and determine future research opportunities related to predictive maintenance in modern engineering systems (Tsallis et al., 2025).

The review process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020**)** framework, which consists of four sequential phases: identification, screening, eligibility assessment, and final inclusion. During the identification stage, potentially relevant publications were collected from internationally recognized scientific databases, including Google Scholar, Scopus, IEEE Xplore, ScienceDirect, SpringerLink, Web of Science, and MDPI. These databases were selected because they provide comprehensive coverage of peer-reviewed publications in engineering, computer science, manufacturing, industrial automation, and artificial intelligence. To maximize search accuracy, combinations of keywords were developed based on the research objectives. The search strings included terms such as "data-driven predictive maintenance," "predictive maintenance," "artificial intelligence," "machine learning," "deep learning," "fault diagnosis," "remaining useful life," "condition monitoring," "Industrial Internet of Things," "smart manufacturing," *and* "Industry 5.0," which were combined using Boolean operators (AND/OR) to broaden or refine the search results.

To ensure the relevance and quality of the reviewed literature, explicit inclusion and exclusion criteria were established before the screening process. Eligible studies were required to present clear research objectives, methodological descriptions, experimental or analytical validation, and measurable outcomes related to predictive maintenance performance. Publications such as conference proceedings, editorials, book chapters, theses, technical reports, duplicated records, and studies unrelated to predictive maintenance were excluded. Furthermore, articles lacking sufficient methodological information or focusing solely on preventive or reactive maintenance without incorporating data-driven intelligence were not considered in the review.Following the initial search, all retrieved publications underwent a systematic screening process. The structured screening procedure also improved consistency in article selection while reducing the possibility of subjective judgment during the review process (Ghamrawi et al., 2025).

After the eligible studies had been identified, qualitative content analysis was employed to extract and organize relevant information from each publication. The extracted information included publication year, research objectives, engineering application domain, industrial equipment, sensor technologies, operational data characteristics, Artificial Intelligence and Machine Learning algorithms, deep learning architectures, feature engineering methods, evaluation metrics, implementation outcomes, identified challenges, and future research recommendations. Special attention was given to the comparative performance of supervised learning, unsupervised learning, ensemble learning, and deep learning techniques in predictive maintenance applications. In addition, information regarding Remaining Useful Life (RUL) prediction, anomaly detection, fault diagnosis, predictive accuracy, computational efficiency, and model interpretability was systematically documented to facilitate cross-study comparison.

3. | RESULTS AND DISCUSSION

The analysis of the selected studies demonstrates a significant increase in research on data-driven predictive maintenance (PdM) published over the last five years, reflecting the accelerating adoption of Artificial Intelligence (AI), Machine Learning (ML), and Industrial Internet of Things (IIoT) technologies within modern engineering systems. The rapid digital transformation associated with Industry 4.0 has shifted maintenance strategies from reactive and preventive approaches toward intelligent maintenance frameworks capable of predicting equipment degradation before failure occurs. Rather than relying on fixed maintenance schedules, contemporary predictive maintenance systems continuously analyze operational data collected from sensors to estimate equipment health conditions and support maintenance decision-making. Recent systematic reviews consistently indicate that predictive maintenance has become one of the most active research domains in industrial engineering because of its potential to reduce unplanned downtime, increase asset availability, improve operational safety, and optimize maintenance costs (Lee et al., 2020).

One of the most significant findings identified in the reviewed literature is the dominant role of machine learning algorithms in fault prediction and equipment health assessment. Traditional statistical models are increasingly being replaced by supervised learning algorithms because of their superior capability to model nonlinear relationships among multiple operational variables. Algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) remain among the most frequently implemented techniques due to their robustness, relatively low computational complexity, and strong classification performance. These algorithms are particularly effective for structured datasets obtained from industrial monitoring systems where labeled failure records are available. Several comparative studies further indicate that ensemble learning techniques generally outperform single classifiers by improving prediction accuracy while reducing overfitting in complex industrial datasets (Mienye & Sun, 2022).

In addition to conventional machine learning algorithms, deep learning has become increasingly important for predictive maintenance because industrial data are commonly generated as multivariate time-series with highly nonlinear characteristics. Deep learning architectures, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Autoencoders, and Transformer-based models, have demonstrated remarkable performance in extracting representative features directly from raw sensor signals without requiring extensive manual feature engineering. CNN models are widely applied for fault classification using vibration signals and acoustic emission data, whereas recurrent neural networks such as LSTM are particularly effective for Remaining Useful Life (RUL) estimation because they capture long-term temporal dependencies within sequential operational data. More recent studies also report the integration of hybrid CNN-LSTM architectures to simultaneously exploit spatial and temporal characteristics of sensor data, resulting in improved prediction accuracy across multiple engineering applications (Wang et al., 2023).

Another important observation concerns the increasing availability of industrial data generated through IIoT infrastructures. Modern industrial systems employ distributed sensor networks capable of continuously collecting measurements such as vibration, temperature, pressure, rotational speed, electrical current, lubricant quality, and acoustic signals. These heterogeneous datasets form the foundation of data-driven predictive maintenance because the quality of prediction models strongly depends on data completeness, diversity, and reliability. Consequently, considerable attention has been devoted to data preprocessing activities, including missing-value treatment, signal filtering, normalization, feature extraction, feature selection, and dimensionality reduction before model development. Feature engineering remains an essential component for conventional machine learning methods, whereas deep learning approaches increasingly perform automated feature extraction from high-dimensional datasets. The literature consistently emphasizes that improving data quality frequently contributes more to prediction performance than merely adopting increasingly sophisticated algorithms (Piryonesi & El-Diraby, 2020).

The reviewed studies also demonstrate growing interest in integrating Digital Twin technology into predictive maintenance frameworks. A Digital Twin provides a dynamic virtual representation of physical equipment by continuously synchronizing operational information collected from IoT devices. This virtual environment enables engineers to simulate degradation processes, evaluate maintenance scenarios, estimate Remaining Useful Life, and optimize maintenance scheduling without interrupting actual industrial operations. Compared with conventional predictive maintenance systems, Digital Twin-based approaches provide enhanced decision support by combining real-time monitoring, physical modeling, simulation, and artificial intelligence within a unified architecture. Recent systematic reviews indicate that Digital Twin technology is becoming a critical enabling platform for intelligent maintenance because it facilitates continuous model updating and improves prediction reliability under changing operational conditions. Nevertheless, computational complexity, interoperability among industrial platforms, and high implementation costs remain significant obstacles to broader industrial adoption (Urrea & Kern, 2025).

Model evaluation constitutes another fundamental aspect identified across the reviewed literature. Classification-based predictive maintenance models are commonly evaluated using Accuracy, Precision, Recall, F1-score, and Area Under the ROC Curve (AUC), whereas regression models for Remaining Useful Life prediction frequently employ Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). Recent studies increasingly recommend using multiple evaluation metrics simultaneously because reliance on overall accuracy alone may produce misleading conclusions, particularly when industrial datasets exhibit severe class imbalance where failure events occur much less frequently than normal operating conditions. Consequently, precision-recall analysis, confusion matrices, and cost-sensitive evaluation have become increasingly important for assessing predictive maintenance systems intended for real-world engineering applications. Despite the promising performance of AI-based predictive maintenance systems, the reviewed literature consistently highlights several challenges that continue to limit their large-scale industrial implementation. One of the most frequently reported issues is the quality and availability of industrial data. Machine learning algorithms require sufficient historical failure records to learn representative degradation patterns; however, actual industrial environments often generate highly imbalanced datasets in which normal operating conditions vastly outnumber failure events. This imbalance reduces the capability of classification models to accurately identify rare fault conditions and frequently results in biased predictions. In addition, industrial sensor data are commonly affected by missing values, measurement noise, inconsistent sampling frequencies, and communication failures, all of which negatively influence model robustness and prediction reliability. Consequently, recent studies emphasize that data preprocessing, anomaly filtering, feature engineering, and data augmentation remain essential steps before model development. Furthermore, the standardization of industrial data collection protocols has become an important research topic to improve interoperability among heterogeneous monitoring systems (Amjad et al., 2021).

Another major challenge concerns the interpretability of advanced machine learning models. Although deep learning architectures generally achieve higher predictive accuracy than conventional algorithms, their decision-making processes are often difficult for engineers to understand. This "black-box" characteristic limits industrial trust, particularly in safety-critical engineering systems where maintenance decisions may involve significant operational risks and financial consequences. As a result, Explainable Artificial Intelligence (XAI) has emerged as an important research direction during the 2021–2025 period. XAI techniques provide interpretable explanations for prediction outcomes by identifying influential variables, visualizing feature importance, and explaining the contribution of sensor measurements to maintenance recommendations. The integration of explainability with predictive maintenance not only improves user confidence but also facilitates regulatory compliance, model validation, and collaborative decision-making between intelligent systems and maintenance engineers. Current evidence suggests that future predictive maintenance platforms will increasingly combine high prediction accuracy with transparent and interpretable decision-support mechanisms (Ucar et al., 2024).

The literature also indicates that the rapid expansion of Industrial Internet of Things (IIoT) infrastructures has increased interest in distributed intelligence through Edge AI and Federated Learning. Conventional cloud-based predictive maintenance systems often experience communication latency, bandwidth limitations, and privacy concerns because large volumes of sensor data must be transmitted to centralized servers for analysis. Edge AI addresses these limitations by deploying machine learning models directly on edge devices located near industrial equipment, thereby enabling real-time anomaly detection and maintenance prediction with reduced latency. Similarly, Federated Learning allows multiple industrial facilities to collaboratively train machine learning models without sharing sensitive operational data, thereby improving privacy protection while maintaining model performance. These distributed intelligence paradigms are expected to become increasingly important as manufacturing systems continue to expand toward interconnected cyber-physical production environments. Their adoption also supports more resilient, scalable, and energy-efficient predictive maintenance architectures suitable for next-generation smart factories (Bitam et al., 2025).

Another emerging trend identified across the reviewed studies is the integration of predictive maintenance with Digital Twin technology and advanced simulation environments. Digital Twins continuously synchronize physical assets with virtual models, enabling maintenance engineers to monitor equipment conditions, evaluate degradation scenarios, and optimize maintenance strategies in real time. The combination of Digital Twins with AI algorithms enhances prediction accuracy by incorporating both historical operational data and physics-based simulation models (Sajadieh & Noh, 2025). In addition, recent advances in Foundation Models and Generative Artificial Intelligence have begun to influence predictive maintenance research by enabling automated knowledge extraction, maintenance report generation, intelligent fault diagnosis assistance, and adaptive decision support. Although these technologies remain at an early stage of industrial implementation, they demonstrate considerable potential for improving predictive maintenance through multimodal data integration, contextual reasoning, and autonomous maintenance planning. Future engineering systems are therefore expected to incorporate hybrid intelligence that combines data-driven learning, physical system knowledge, and human expertise to achieve higher reliability and operational efficiency.

4. | CONCLUSION

Data-driven predictive maintenance has become a fundamental approach for improving the reliability, efficiency, and sustainability of modern engineering systems. Based on the systematic review of studies published over the last five years, this research demonstrates that the integration of Artificial Intelligence and Machine Learning has significantly enhanced the capability of predictive maintenance systems to detect equipment anomalies, estimate Remaining Useful Life, and support intelligent maintenance decision-making. Conventional machine learning algorithms remain effective for structured industrial datasets, while deep learning models provide superior performance for analyzing complex and high-dimensional sensor data. Furthermore, the adoption of Industrial Internet of Things technologies has enabled continuous data acquisition, allowing predictive maintenance to evolve toward real-time and autonomous maintenance strategies.

Despite these advancements, several challenges remain, including data quality issues, class imbalance, limited model interpretability, cybersecurity concerns, and integration with legacy industrial systems. Addressing these challenges is essential to ensure reliable and scalable implementation in real-world engineering environments. Emerging technologies such as Explainable Artificial Intelligence, Digital Twins, Edge AI, Federated Learning, and Foundation Models offer promising opportunities to overcome current limitations and further improve predictive maintenance capabilities.

Future research should focus on developing transparent, adaptive, and trustworthy intelligent maintenance frameworks that effectively combine advanced data analytics, domain knowledge, and real-time decision support. Such developments are expected to play a crucial role in accelerating the realization of resilient, efficient, and sustainable smart engineering systems in the era of Industry 4.0 and Industry 5.0.

Acknowledgment

We gratefully acknowledge the contributions of individuals who supported the completion of this article.

Funding Information

This research did not receive any funding.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

Ethical Approval and Originality Statement

Ethical approval was obtained for this study. The manuscript represents original work and has not been previously published, nor is it under consideration by another journal.

Data Disclosure Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On predictive maintenance in industry 4.0: Overview, models, and challenges. *Applied sciences*, *12*(16), 8081.

Amjad, A., Azam, F., Anwar, M. W., & Butt, W. H. (2021). A systematic review on the data interoperability of application layer protocols in industrial IoT. *Ieee Access*, *9*, 96528-96545.

Bitam, T., Yahiaoui, A., Boubiche, D. E., Martínez-Peláez, R., Toral-Cruz, H., & Velarde-Alvarado, P. (2025). Artificial intelligence of things for next-generation predictive maintenance. *Sensors (Basel, Switzerland)*, *25*(24), 7636.

Cheng, H., Husen, J. H., Lu, Y., Racharak, T., Yoshioka, N., Ubayashi, N., & Washizaki, H. (2026). Generative ai for requirements engineering: A systematic literature review. *Software: Practice and Experience*, *56*(2), 141-170.

Ghamrawi, N., Shal, T., Ghamrawi, N. A., Abu-Tineh, A., Alshaboul, Y., & Alazaizeh, M. A. (2025). A step-by-step approach to systematic reviews in educational research. *European Journal of Educational Research*, *14*(2), 549-566.

Islam, M. T., Sepanloo, K., Woo, S., Woo, S. H., & Son, Y. J. (2025). A review of the industry 4.0 to 5.0 transition: exploring the intersection, challenges, and opportunities of technology and human–machine collaboration. *Machines*, *13*(4), 267.

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*(11), 110805.

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *Ieee Access*, *10*, 99129-99149.

Piryonesi, S. M., & El-Diraby, T. E. (2020). Role of data analytics in infrastructure asset management: Overcoming data size and quality problems. *Journal of Transportation Engineering, Part B: Pavements*, *146*(2), 04020022.

Rahman, M. A., Shahrior, M. F., Iqbal, K., & Abushaiba, A. A. (2025). Enabling intelligent industrial automation: A Review of machine learning applications with digital twin and edge AI Integration. *Automation*, *6*(3), 37.

Sajadieh, S. M. M., & Noh, S. D. (2025). From simulation to autonomy: Reviews of the integration of artificial intelligence and digital twins. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *12*(5), 1597-1628.

Ucar, A., Karakose, M., & Kırımça, N. (2024). Artificial intelligence for predictive maintenance applications: key components, trustworthiness, and future trends. *Applied Sciences*, *14*(2), 898.

Urrea, C., & Kern, J. (2025). Recent advances and challenges in industrial robotics: A systematic review of technological trends and emerging applications. *Processes*, *13*(3), 832.

Wang, X., Bai, Y., & Liu, X. (2023). Prediction of railroad track geometry change using a hybrid CNN-LSTM spatial-temporal model. *Advanced Engineering Informatics*, *58*, 102235.