JOURNAL OF ADVANCED ENGINEERING AND INNOVATION
Volume 3, Issue 2, December 2025 | pp. 98-107
https://doi.org/10.54518/jaei.3.2.2025.1360

An Integrated Artificial Intelligence and Big Data Analytics Framework for Accurate Machine Failure Prediction

Widya Nurhalimah[1] ✉

Universitas Perwira Purbalingga, Indonesia[1]

ABSTRACT

Machine failure prediction has become a critical component of predictive maintenance in smart manufacturing environments. The growing complexity of industrial systems and the massive volume of data generated by Industrial Internet of Things (IIoT) sensors require intelligent approaches capable of processing large-scale data while providing accurate predictions. This study proposes an integrated conceptual framework that combines Artificial Intelligence (AI) and Big Data Analytics (BDA) to improve machine failure prediction accuracy. A Systematic Literature Review (SLR) was conducted by synthesizing peer-reviewed studies published over the last five years. The findings indicate that Big Data Analytics supports industrial data acquisition, storage, integration, preprocessing, and feature engineering, whereas Artificial Intelligence enables pattern learning, degradation identification, and machine failure prediction through intelligent algorithms. The integration of these technologies establishes a comprehensive predictive maintenance workflow that transforms industrial data into actionable maintenance decisions. The proposed framework is expected to improve prediction accuracy, reduce unplanned downtime, optimize maintenance scheduling, and enhance equipment reliability. Furthermore, it provides a conceptual foundation for future empirical research and practical implementation of intelligent predictive maintenance across various industrial sectors.

Keywords: *Artificial Intelligence, Big Data Analytics, Machine Failure Prediction, Predictive Maintenance, Smart Manufacturing.*

CORRESPONDING AUTHOR:
Widya Nurhalimah
Universitas Perwira Purbalingga, Indonesia

ARTICLE HISTORY

Received	: July 12, 2025
Final Revised	: Sepember 15, 2025
Accepted	: November 6, 2025
Published	: December 30, 2025

1. | INTRODUCTION

The rapid advancement of Industry 4.0 has transformed conventional manufacturing systems into intelligent production environments characterized by interconnected machines, Industrial Internet of Things (IIoT), cyber-physical systems, cloud computing, and real-time data exchange. In this digital ecosystem, industrial equipment continuously generates large volumes of operational data through embedded sensors, enabling organizations to monitor machine conditions more accurately than ever before. These technological developments have shifted maintenance strategies from reactive and preventive approaches toward predictive maintenance, which aims to forecast equipment failures before they occur, thereby minimizing unexpected downtime and improving operational efficiency (Shahin et al., 2023).

Machine failure remains one of the most significant challenges in industrial operations because unexpected equipment breakdowns can interrupt production processes, reduce product quality, increase maintenance costs, and decrease overall system reliability. Traditional maintenance strategies generally rely on scheduled inspections or corrective actions after failures occur. Although these approaches are widely implemented, they often fail to detect hidden degradation patterns and cannot accurately determine the optimal maintenance schedule. Consequently, industries increasingly require intelligent failure prediction systems capable of analyzing large-scale operational data and identifying early indications of machine degradation (Yan et al., 2023).

Artificial Intelligence (AI) has emerged as one of the most promising technologies for predictive maintenance because it enables computers to learn complex relationships from historical and real-time operational data. Various AI techniques, including Machine Learning, Deep Learning, Random Forest, Support Vector Machine, Extreme Gradient Boosting, and Artificial Neural Networks, have demonstrated remarkable capabilities in identifying fault patterns, estimating remaining useful life, and predicting machine failures with high accuracy. Compared with traditional statistical approaches, AI-based models are capable of handling nonlinear relationships and high-dimensional datasets while continuously improving predictive performance through iterative learning processes (Talaat et al., 2024).

Despite these advantages, the effectiveness of AI models strongly depends on the availability and quality of data. Modern manufacturing environments generate massive datasets characterized by high volume, velocity, variety, veracity, and value, which exceed the capability of conventional data processing techniques. Consequently, Big Data Analytics (BDA) has become an essential component of intelligent maintenance systems because it provides mechanisms for collecting, integrating, preprocessing, storing, and analyzing heterogeneous industrial data obtained from multiple sensing devices and production systems. Effective big data management enables AI algorithms to extract meaningful features, reduce noise, and improve prediction reliability (Aldoseri et al., 2023).

Recent studies have demonstrated substantial progress in applying AI to predictive maintenance; however, most investigations emphasize algorithm development while treating data management as a separate process. Similarly, studies focusing on Big Data Analytics frequently concentrate on data infrastructure without comprehensively integrating intelligent prediction models into a unified analytical framework. This separation limits the ability of maintenance systems to fully exploit the synergy between advanced analytics and intelligent decision-making, resulting in reduced prediction robustness and limited practical implementation (Bousdekis et al., 2021).

Furthermore, the increasing complexity of industrial systems requires predictive maintenance frameworks capable of processing real-time streaming data while maintaining high prediction accuracy and computational efficiency. The integration of AI and Big Data Analytics provides opportunities to improve feature extraction, optimize predictive model performance, support scalable decision-making, and enhance maintenance planning throughout the machine life cycle. Such integration is expected to contribute not only to failure prediction accuracy but also to operational reliability, maintenance efficiency, and production sustainability (Jasiulewicz-Kaczmarek et al., 2020).

Based on these considerations, this study proposes an integrated framework combining Artificial Intelligence and Big Data Analytics to improve machine failure prediction. Unlike previous studies that primarily examine these technologies independently, the proposed framework emphasizes their complementary roles throughout the predictive maintenance process, beginning with industrial data acquisition, preprocessing, feature engineering, predictive modeling, and intelligent maintenance decision support. The proposed framework is expected to provide a comprehensive conceptual foundation for future predictive maintenance systems while contributing to the development of intelligent, scalable, and data-driven industrial maintenance solutions.

2. | RESEARCH METHOD

This study employed a Systematic Literature Review (SLR) approach to develop an integrated framework of Artificial Intelligence (AI) and Big Data Analytics (BDA) for accurate machine failure prediction. The SLR method was selected because it provides a structured, transparent, and reproducible procedure for identifying, evaluating, synthesizing, and interpreting existing scientific evidence related to a particular research topic. Unlike conventional literature reviews, the SLR approach minimizes selection bias through predefined search strategies, inclusion criteria, exclusion criteria, and systematic analysis, thereby producing more reliable and comprehensive research findings. The review process adopted the general principles of systematic literature studies proposed by Kitchenham and further adapted for engineering and information technology research (Saltz & Krasteva 2022).

The literature search was conducted using Google Scholar as the primary academic database because it indexes a broad range of peer-reviewed journal articles, conference proceedings, and scholarly publications across engineering, computer science, manufacturing, and industrial informatics. To ensure the relevance and currency of the proposed framework, only published over the last five years were included in the review. The search process employed combinations of keywords such as Artificial Intelligence, Machine Learning, Deep Learning, Big Data Analytics, Predictive Maintenance, Machine Failure Prediction, Industrial Internet of Things, Smart Manufacturing, *and* Industry 4.0. Additional manual screening was performed by examining the reference lists of selected articles to identify other highly relevant studies (Haddaway et al., 2022).

The inclusion criteria comprised peer-reviewed journal articles written in English, published over the last five years, and discussing AI, Big Data Analytics, predictive maintenance, machine condition monitoring, or machine failure prediction within industrial contexts. Conversely, publications outside the specified publication period, non-peer-reviewed documents, duplicated studies, and articles lacking sufficient methodological information were excluded. Following the identification and screening stages, the selected studies were critically analyzed to extract information regarding research objectives, AI algorithms, big data techniques, industrial applications, predictive performance, and reported limitations. The synthesized findings were subsequently used to formulate an integrated conceptual framework illustrating the interaction between AI and Big Data Analytics throughout the machine failure prediction process, from industrial data acquisition and preprocessing to predictive modeling and intelligent maintenance decision support.

3. | RESULTS AND DISCUSSION

The results of this study demonstrate that the integration of Artificial Intelligence (AI) and Big Data Analytics (BDA) provides a comprehensive conceptual framework for improving machine failure prediction in smart manufacturing environments. Based on the synthesis of literature published over the last five years, predictive maintenance has evolved from conventional condition monitoring into an intelligent decision-support system capable of continuously processing industrial data generated by sensors, Industrial Internet of Things (IIoT) devices, cyber-physical systems, and connected manufacturing equipment. Recent studies consistently indicate that the increasing availability of real-time operational data has significantly enhanced the feasibility of developing predictive models that identify machine degradation before catastrophic failures occur. However, these studies also emphasize that predictive performance depends not only on sophisticated AI algorithms but equally on effective data acquisition, preprocessing, integration, and management (Aldoseri et al., 2023).

The proposed framework begins with industrial data acquisition, where operational parameters are collected from multiple sensing devices installed on manufacturing equipment. These parameters generally include vibration, temperature, rotational speed, pressure, torque, electrical current, voltage, lubrication conditions, and environmental variables. The heterogeneous nature of these datasets creates challenges associated with data volume, velocity, variety, and veracity, making conventional database systems insufficient for supporting predictive maintenance applications. Consequently, Big Data Analytics functions as the foundation for managing large-scale industrial datasets through scalable storage, distributed processing, data cleaning, feature extraction, and data integration before machine learning models are implemented. Previous studies emphasize that improving data quality substantially enhances the effectiveness of predictive maintenance algorithms because noisy, incomplete, or inconsistent data frequently reduce prediction accuracy regardless of the complexity of the learning algorithm (Hector & Panjanathan, 2024).

Following data preparation, Artificial Intelligence becomes the principal analytical component responsible for transforming processed industrial data into predictive knowledge. Machine learning algorithms such as Random Forest, Support Vector Machine, Artificial Neural Network, Extreme Gradient Boosting, Gradient Boosting, Decision Tree, and Deep Learning architectures have demonstrated superior capability in identifying nonlinear relationships among multiple machine condition variables. Unlike traditional statistical approaches that often require predefined assumptions regarding system behavior, AI algorithms automatically learn hidden degradation patterns from historical operational records. This capability enables intelligent prediction of machine failures before observable symptoms become critical, thereby supporting proactive maintenance scheduling and reducing unexpected production interruptions. Several comparative studies further report that ensemble learning methods and hybrid deep learning architectures consistently outperform conventional classification approaches when sufficient high-quality training data are available (Younas et al., 2023).

Another important finding derived from the reviewed literature concerns the complementary relationship between AI and Big Data Analytics. Numerous previous investigations focused primarily on improving prediction algorithms without simultaneously addressing industrial data management challenges. Conversely, several Big Data studies concentrated on scalable storage and processing technologies while giving limited attention to intelligent predictive modeling. This separation reduces the overall effectiveness of predictive maintenance because machine learning algorithms cannot achieve optimal performance without reliable and well-structured datasets. The proposed integrated framework therefore positions Big Data Analytics as the enabling infrastructure that continuously supplies high-quality information to AI models. Rather than functioning independently, both technologies operate as mutually reinforcing components within the same predictive maintenance ecosystem (Shaheen & Németh, 2022).

The synthesized framework also highlights the importance of feature engineering during industrial data preprocessing. Sensor-generated information frequently contains redundant variables, missing observations, measurement noise, and inconsistent sampling intervals that may degrade model performance. Data normalization, outlier detection, missing-value treatment, feature selection, and dimensionality reduction therefore become essential processes before model training. Effective feature engineering not only improves computational efficiency but also increases model generalization by preserving informative characteristics while removing irrelevant information. The literature consistently demonstrates that predictive maintenance systems incorporating robust preprocessing pipelines achieve higher predictive stability than systems relying solely on sophisticated algorithms without adequate data preparation (Alabadi & Habbal, 2023).

Another significant observation concerns real-time predictive maintenance implementation. Modern industrial environments require continuous monitoring of production assets rather than periodic offline analysis. Consequently, AI models must process streaming data generated from IIoT-enabled production systems while maintaining acceptable computational latency. Big Data platforms support this requirement by enabling distributed processing architectures capable of handling high-frequency sensor streams from multiple machines simultaneously. When integrated with AI inference models, these architectures enable continuous health assessment, anomaly detection, failure prediction, and intelligent maintenance recommendations without interrupting manufacturing operations. Such capabilities directly contribute to reduced downtime, increased equipment availability, optimized spare-part management, and lower maintenance costs.

The review also reveals that model evaluation remains a critical component of predictive maintenance research. Performance assessment generally employs Accuracy, Precision, Recall, F1-score, Area Under the Curve (AUC), Receiver Operating Characteristic (ROC), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and confusion matrix analysis depending on whether classification or regression tasks are performed (Agrawal et al., 2022). Using multiple evaluation metrics provides a more comprehensive understanding of model performance because high classification accuracy alone may conceal weaknesses in detecting minority failure classes. Consequently, future predictive maintenance systems should evaluate algorithm robustness using multiple complementary performance indicators while considering computational efficiency and practical deployment constraints.

Although considerable progress has been achieved, several research gaps remain. First, many published studies continue to evaluate predictive algorithms using benchmark datasets instead of long-term industrial operational data, limiting their generalizability. Second, interoperability between heterogeneous industrial data sources remains a challenge because equipment from different manufacturers often produces incompatible data formats. Third, explainability and transparency of AI models remain insufficiently explored despite growing industrial demand for interpretable maintenance recommendations. Fourth, scalability, cybersecurity, privacy protection, and computational resource optimization continue to represent important barriers to large-scale industrial implementation. Addressing these limitations requires tighter integration between AI algorithms, Big Data infrastructures, cloud computing, edge computing, and IIoT communication technologies within unified predictive maintenance architectures (D'Agostino et al., 2024).

Based on the overall synthesis, the proposed integrated framework consists of six interconnected stages: industrial data acquisition, big data storage, data preprocessing, feature engineering, AI-based predictive modeling, and intelligent maintenance decision support. Information collected from sensors is first stored within scalable Big Data infrastructure before undergoing cleaning, transformation, and feature extraction. The processed datasets are subsequently analyzed using machine learning or deep learning algorithms to identify degradation patterns and estimate failure probability. Prediction results are then converted into maintenance recommendations that support maintenance scheduling, spare-part allocation, inspection prioritization, and operational planning. Compared with fragmented approaches that separately discuss AI or Big Data Analytics, the proposed framework provides a holistic perspective illustrating how both technologies interact throughout the entire predictive maintenance lifecycle (Nordal & El-Thalji, 2021).

Overall, the analysis indicates that integrating Artificial Intelligence and Big Data Analytics offers substantial potential for improving machine failure prediction accuracy while simultaneously increasing operational efficiency and maintenance effectiveness. The framework supports intelligent, scalable, and data-driven predictive maintenance capable of processing heterogeneous industrial data in real time. Furthermore, it establishes a conceptual foundation for future empirical studies that validate the proposed architecture using real industrial datasets, advanced deep learning algorithms, explainable artificial intelligence techniques, digital twins, and edge computing platforms to further enhance predictive reliability and industrial applicability.

4. | CONCLUSION

This study proposed an integrated framework that combines Artificial Intelligence (AI) and Big Data Analytics (BDA) to improve the accuracy of machine failure prediction in smart manufacturing environments. Based on the systematic synthesis of recent literature, the study demonstrates that AI and BDA should not be treated as independent technologies but rather as complementary components within a unified predictive maintenance ecosystem. Big Data Analytics provides the infrastructure required for collecting, storing, integrating, preprocessing, and managing large-scale industrial data, while Artificial Intelligence transforms these processed datasets into predictive knowledge through intelligent learning algorithms capable of identifying degradation patterns and estimating machine failures before breakdowns occur.

The proposed framework illustrates a complete data-driven workflow consisting of industrial data acquisition, big data management, data preprocessing, feature engineering, AI-based predictive modeling, and intelligent maintenance decision support. The integration of these stages is expected to improve prediction accuracy, reduce unplanned downtime, optimize maintenance scheduling, enhance equipment reliability, and support more efficient operational decision-making. Furthermore, the framework provides a comprehensive conceptual foundation that can be adapted to various industrial sectors regardless of specific equipment types or manufacturing environments.

Future research should focus on validating the proposed framework using real industrial datasets, investigating the integration of explainable artificial intelligence, edge computing, digital twins, and Industrial Internet of Things technologies, as well as evaluating the scalability and practical implementation of the framework in complex industrial systems. These efforts are expected to further strengthen the effectiveness, reliability, and industrial applicability of intelligent predictive maintenance systems.

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

Agrawal, H., Jain, P., & Joshi, A. M. (2022). Machine learning models for non-invasive glucose measurement: towards diabetes management in smart healthcare. *Health and technology*, *12*(5), 955-970.

Alabadi, M., & Habbal, A. (2023). Next-generation predictive maintenance: leveraging blockchain and dynamic deep learning in a domain-independent system. *PeerJ Computer Science*, *9*, e1712.

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. *Applied Sciences*, *13*(12), 7082.

Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2021). A review of data-driven decision-making methods for industry 4.0 maintenance applications. *Electronics*, *10*(7), 828.

D'Agostino, P., Violante, M., & Macario, G. (2024). A scalable fog computing solution for industrial predictive maintenance and customization. *Electronics*, *14*(1), 24.

Haddaway, N. R., Grainger, M. J., & Gray, C. T. (2022). Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. *Research synthesis methods*, *13*(4), 533-545.

Hector, I., & Panjanathan, R. (2024). Predictive maintenance in Industry 4.0: A survey of planning models and machine learning techniques. *PeerJ Computer Science*, *10*, e2016.

Jasiulewicz-Kaczmarek, M., Legutko, S., & Kluk, P. (2020). Maintenance 4.0 technologies–new opportunities for sustainability driven maintenance. *Management and production engineering review*, *11*.

Nordal, H., & El-Thalji, I. (2021). Modeling a predictive maintenance management architecture to meet industry 4.0 requirements: A case study. *Systems Engineering*, *24*(1), 34-50.

Saltz, J. S., & Krasteva, I. (2022). Current approaches for executing big data science projects—a systematic literature review. *PeerJ Computer Science*, *8*, e862.

Shaheen, B. W., & Németh, I. (2022). Integration of maintenance management system functions with industry 4.0 technologies and features—A review. *Processes*, *10*(11), 2173.

Shahin, M., Chen, F. F., Hosseinzadeh, A., & Zand, N. (2023). Using machine learning and deep learning algorithms for downtime minimization in manufacturing systems: An early failure detection diagnostic service. *The international journal of advanced manufacturing technology*, *128*(9), 3857-3883.

Talaat, F. M., Elnaggar, A. R., Shaban, W. M., Shehata, M., & Elhosseini, M. (2024). CardioRiskNet: a hybrid AI-based model for explainable risk prediction and prognosis in cardiovascular disease. *Bioengineering*, *11*(8), 822.

Yan, W., Wang, J., Lu, S., Zhou, M., & Peng, X. (2023). A review of real-time fault diagnosis methods for industrial smart manufacturing. *Processes*, *11*(2), 369.

Younas, F., Usman, M., & Yan, W. Q. (2023). A deep ensemble learning method for colorectal polyp classification with optimized network parameters: A deep ensemble learning method for colorectal polyp classification with optimized network parameters. *Applied Intelligence*, *53*(2), 2410-2433.