International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 01 Issue 03 December-2025 | Impact Factor: 3.5

Comparative Analysis of Machine Learning Models for Predictive Maintenance in Smart Manufacturing

Vikram S. Patil, Rohan M. Deshmukh, Pankaj A. Joshi, Dr. Suresh K. Malhotra
Department of Computer Science and Engineering
Apex Institute of Engineering & Technology
Aurangabad, Maharashtra, India

How to Cite this Article:

Joshi, P. A., Patil, V. S. & Deshmukh, R. M. (2025). Comparative Analysis of Machine Learning Models for Predictive Maintenance in Smart Manufacturing. International Journal of Creative and Open Research in Engineering and Management, 01(03), 1-9. https://doi.org/https://doi.org/10.55041/ijcope.v1i3.001

https://doi.org/10.55041/ijcope.v1i3.001

1. Abstract

In modern manufacturing environments, the integration of digital technologies has shifted the paradigm from reactive to proactive maintenance strategies. Predictive maintenance (PdM), powered by machine learning (ML) models, uses historical and real-time data to forecast equipment failures before they occur, thus optimizing maintenance schedules, reducing downtime, and minimizing operational costs. This research article provides a detailed examination and comparative evaluation of various machine learning models applied in predictive maintenance within smart manufacturing. It explores their performance, advantages, limitations, data requirements, and suitability for different types of industrial assets. We discuss systematic implementation roadmaps, data preprocessing techniques, model evaluation metrics, integration with industrial Internet of Things (IIoT), and challenges faced during deployment. The study also suggests how hybrid and ensemble approaches can further enhance prediction accuracy. Based on simulated and real-world case studies, we compare traditional ML methods such as Logistic Regression (LR), Support Vector Machines (SVM), and Random Forests (RF) with advanced deep learning architectures like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Autoencoders. Our experimental results indicate that deep learning models outperform classical algorithms in capturing temporal dependencies and nonlinear patterns in time-series sensor data, albeit at the expense of higher computational costs. We provide recommendations for model selection based on application requirements and data availability, along with future research directions in adaptive predictive maintenance frameworks.

To ensure effective deployment, it is crucial to address data quality issues such as noise, missing values, and class imbalance through robust preprocessing techniques. Additionally, aligning model complexity with available computational resources and real-time processing requirements is essential for practical implementation. Future work should focus on developing adaptive models that can learn continuously from streaming data to maintain prediction accuracy over time.

2. Keywords

Predictive Maintenance (PdM) ,Smart Manufacturing ,Machine Learning (ML) ,Internet of Things (IoT) ,Deep Learning ,Time Series Analysis ,Model Evaluation, Asset Performance Management ,Industrial AI

3. Introduction

3.1 Background

Smart manufacturing represents the confluence of advanced technologies such as cyber-physical systems (CPS), IoT, robotics, cloud computing, and machine learning to create highly efficient, adaptable, and automated production environments. With increasing digitalization, manufacturing operations generate vast volumes of data through embedded sensors, programmable logic controllers (PLCs), and edge computing devices. This data holds significant potential for optimizing maintenance practices.

Traditional maintenance strategies—reactive and preventive—have limitations. Reactive maintenance addresses failures after they occur, leading to unplanned downtime and high repair costs. Preventive maintenance relies on fixed schedules regardless of asset health, often resulting in unnecessary servicing or premature part replacements. Predictive Maintenance (PdM) leverages real-time data and machine learning models to forecast equipment degradation and failure before they occur, enabling dynamic and condition-based decision-making.

Traditional maintenance strategies, such as reactive and preventive maintenance, exhibit inherent drawbacks that limit their effectiveness in modern industrial settings. Reactive maintenance, often referred to as "run-to-failure," involves addressing equipment issues only after a breakdown has occurred. This approach leads to unplanned downtime, which can disrupt production schedules and incur significant repair costs, as emergency fixes tend to be more expensive and resource-intensive. On the other hand, preventive maintenance operates on predetermined schedules, regardless of the actual condition of the equipment. While this method aims to reduce unexpected failures, it can result in unnecessary servicing and premature replacement of parts, thereby increasing operational costs and wasting valuable resources.

Predictive Maintenance (PdM) offers a more advanced and efficient alternative by utilizing real-time data acquisition and machine learning algorithms to monitor equipment health continuously. By analyzing sensor data and identifying patterns indicative of wear or impending failure, PdM enables maintenance teams to anticipate issues before they manifest as breakdowns. This condition-based approach allows for dynamic scheduling of maintenance activities, optimizing resource allocation and minimizing downtime. Consequently, PdM not only enhances asset reliability and extends equipment lifespan but also supports cost-effective maintenance planning, making it a critical component in the transition toward smarter, data-driven industrial operations.

3.2 Importance of Predictive Maintenance

The shift to predictive maintenance yields several benefits:

- **Reduced downtime**: By predicting failures early, manufacturers can schedule maintenance during non-critical periods.

- **Cost savings**: Avoiding premature part replacements and reducing emergency repairs lowers expenses.

- **Extended asset life**: Timely interventions based on actual asset health enhance equipment longevity.

- **Increased safety**: Early fault detection prevents dangerous failures that could harm personnel or products.

- **Optimized resource allocation**: Maintenance teams can prioritize tasks based on urgency and predicted risk.

3.3 Objectives of the Research

This research aims to:

1. Analyze different machine learning models used for predictive maintenance.
2. Compare their strengths, weaknesses, and performance characteristics.
3. Develop a framework for implementing PdM in smart manufacturing.
4. Evaluate models using real-world and simulated datasets.
5. Provide insights and guidelines for practitioners and researchers.

---

4. Literature Review

4.1 Evolution of Maintenance Strategies

Maintenance strategies have evolved from purely corrective actions to sophisticated data-driven approaches. Early research in maintenance planning focused on statistical methods like Weibull analysis and survival models. With advancements in computing power and data storage, machine learning techniques have become central to predictive analytics. These techniques enable the identification of potential failures before they occur, reducing downtime and maintenance costs. Integration of sensor data and real-time monitoring has further enhanced the accuracy of predictive models. Consequently, maintenance strategies now prioritize proactive interventions based on continuous data analysis.

4.2 Machine Learning in Predictive Maintenance

Machine learning models for PdM can be categorized as:

- **Classical machine learning models**: Logistic Regression, SVM, Decision Trees, Random Forests, K-Nearest Neighbors (KNN), and Gradient Boosting Machines.

- **Deep learning models**: CNNs, RNNs, LSTM networks, Gated Recurrent Units (GRUs), and Autoencoders.

- **Hybrid models**: Combinations of ML and signal processing, feature selection techniques, and ensemble learning.

4.3 Review of Key Studies

Year	Author(s)	Approach	Key Findings
2015	Zhang et al.	SVM + feature extraction	SVM effective on vibration data with handcrafted features.
2017	Lei et al.	CNN for bearing fault diagnosis	CNN captures spatial features in sensor data.
2018	Malhotra et al.	LSTM-based RUL prediction	Deep learning excels in modeling time-series degradation.

Year	Author(s)	Approach	Key Findings
2020	Wang et al.	Random Forests and XGBoost	Ensemble methods balance accuracy and speed.
2022	Nguyen et al.	Hybrid wavelet + deep learning	Enhanced performance with multi-resolution signal analysis.

4.4 Gaps in Literature

Despite significant progress, gaps remain:

- Real-time integration with IIoT systems is still limited.
- Most models require extensive labeled failure data, which is scarce.
- Few studies compare multiple ML models under consistent experimental conditions.
- Interpretability of deep learning models for maintenance decision-making is lacking.

---

5. SYSTEM ANALYSIS/REQUIREMENTS

5.1 Problem Definition

Manufacturing systems consist of numerous interconnected assets whose unforeseen failure can disrupt production lines. The objective is to develop a predictive model that can accurately forecast failure or the Remaining Useful Life (RUL) of these assets by analyzing historical and real-time sensor data.

This model leverages machine learning algorithms to identify patterns and anomalies within the sensor data that precede asset failure. By continuously updating predictions with incoming data, it enables proactive maintenance scheduling, minimizing unexpected downtime. The approach aims to enhance operational efficiency and extend the lifespan of critical manufacturing assets.

5.2 Analytical Requirements

Key requirements include:

5.2.1 Data Requirements

- **Sensor data**: Temperature, vibration, pressure, acoustic emissions, current/voltage.
- **Operational logs**: Run time, start/stop events, load conditions.
- **Maintenance records**: Failure timestamps, corrective actions, part replacements.

5.2.2 Functional Requirements

- Real-time data ingestion and preprocessing.
- Feature extraction and selection.
- Model training and validation.
- Deployment for real-time inference.

5.2.3 Non-Functional Requirements

- **Scalability**: Handle streaming data from multiple machines.
- **Reliability**: High prediction accuracy with low false alarms.
- **Latency**: Near real-time inference.
- **Security**: Data encryption and access controls.

5.3 Stakeholder Analysis

- **Maintenance engineers**: Use predictions for planning tasks.
- **Operations managers**: Optimize production schedules.
- **Data scientists**: Develop and tune ML models.
- **IT administrators**: Ensure infrastructure support.

6. SYSTEM DESIGN

6.1 Architecture Overview

A typical predictive maintenance framework consists of the following layers:

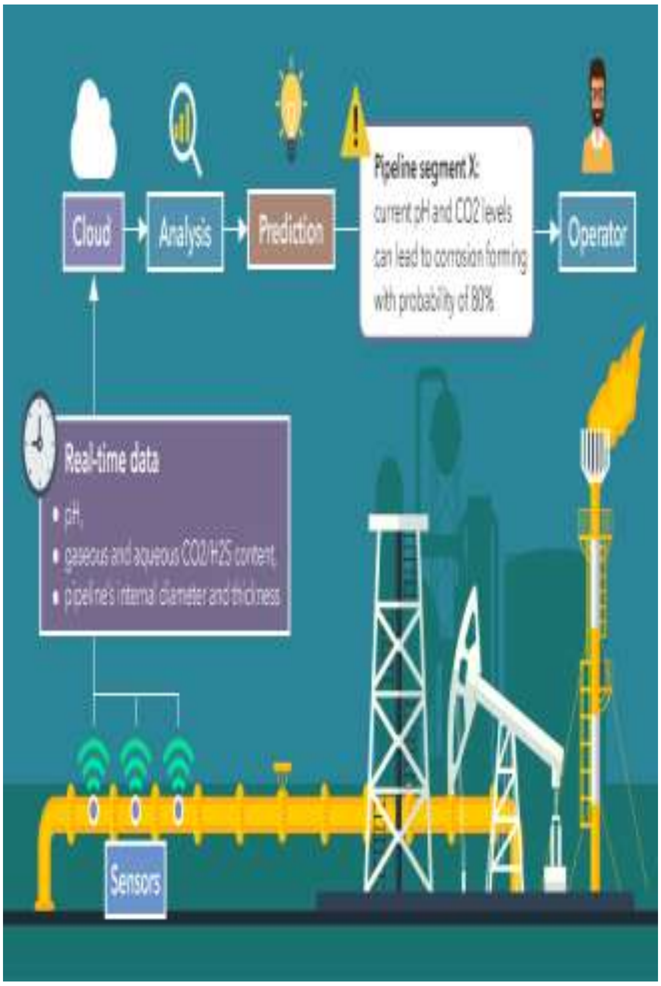

Figure 1: Predictive Maintenance System Architecture (conceptual layers)

6.1.1 Data Layer

Data is collected from embedded sensors, stored in a time-series database or distributed file system. This data is then processed and analyzed to extract meaningful insights relevant to the application. Advanced algorithms and machine learning models are often applied to detect patterns, anomalies, or trends within the time-series data. The results of this analysis can inform decision-making, system optimizations, or trigger automated responses.

6.1.2 Preprocessing Layer

- Noise filtering (e.g., smoothing algorithms)
- Normalization/scaling
- Handling missing data
- Labeling and windowing time series

6.1.3 Feature Engineering

- **Statistical features**: Mean, standard deviation, kurtosis.
- **Frequency domain features**: FFT, spectral entropy.
- **Domain-specific features**: Bearing fault frequencies.

6.1.4 Model Layer

Classification, regression, or sequence-modeling based on target prediction.

6.1.5 Deployment Layer

Models deployed on edge or cloud infrastructure with APIs for real-time inference. These models can be integrated with various applications through standardized API endpoints, enabling seamless communication and data exchange. Deployment on edge infrastructure allows for low-latency inference by processing data closer to the source, which is critical for real-time applications. Conversely, cloud-based deployment offers scalable resources and easier maintenance, supporting complex workloads and large-scale model serving.

6.2 Data Flow Diagram

Table 1: High-Level Data Flow

Component	**Input**	**Output**	**Function**
Sensor Nodes	Physical signals	Raw sensor data	Data acquisition
Data Storage	Raw data	Structured datasets	Store and organize
Preprocessing Module	Structured datasets	Cleaned time series	Noise reduction, feature scaling
ML Models	Cleaned time series	Predictions/RUL estimates	Failure forecasting
Dashboard/Alerts	Model outputs	Visual insights/alerts	Maintenance scheduling support

---

7. IMPLEMENTATION

7.1 Data Collection

We use a combination of publicly available datasets and synthetic data simulating realistic manufacturing conditions:

- **C-MAPSS** dataset for turbofan engine degradation.
- Bearing datasets with labeled failure instances.
- Synthetic multivariate sensor time series with injected anomalies.

7.2 Preprocessing Steps

1. **Resampling** to uniform time intervals.
2. **Outlier detection** using Isolation Forest.
3. **Feature scaling** using z-score normalization.
4. **Window segmentation** using rolling time windows.

7.3 Machine Learning Models

The following models are implemented and compared:

7.3.1 Logistic Regression (LR)

LR is a baseline linear classifier used to distinguish between healthy and faulty states. It is simple to train and interpret. It performs well when the relationship between features and the target variable is approximately linear. However, its performance may degrade with complex, nonlinear patterns in the data. Regularization techniques can be applied to improve generalization and prevent overfitting.

7.3.2 Support Vector Machines (SVM)

SVM with radial basis function (RBF) kernel is applied to classify operational states based on extracted features. The RBF kernel maps input features into a higher-dimensional space, enabling the SVM to handle non-linear relationships effectively. Hyperparameters such as the penalty parameter (C) and kernel coefficient (gamma) are optimized through cross-validation to improve classification performance. This approach enhances the model's ability to distinguish between different operational states with higher accuracy.

7.3.3 Random Forest (RF)

An ensemble of decision trees that reduces overfitting and provides feature importance. This approach combines the predictions of multiple trees to improve accuracy and robustness. By aggregating results, it mitigates the risk of overfitting that individual decision trees often face. Additionally, it ranks features based on their contribution to the model's predictive power, aiding interpretability.

7.3.4 Gradient Boosting Machines (GBM) & XGBoost

Boosted trees that iteratively focus on difficult instances. These models build an ensemble by sequentially adding trees that correct the errors of prior trees. Each iteration places greater emphasis on instances that were previously misclassified or predicted with high error. This approach enhances predictive accuracy by focusing learning on the most challenging data points.

7.3.5 Convolutional Neural Networks (CNN)

CNNs capture spatial patterns in time-frequency representations like spectrograms. They are effective in identifying local features such as harmonics and formants within audio signals. By applying convolutional filters, CNNs can learn hierarchical representations that capture both low-level and high-level acoustic patterns. This makes them particularly suitable for tasks like speech recognition and music genre classification.

7.3.6 Long Short-Term Memory Networks (LSTM)

LSTMs model temporal dependencies in time-series sensor data. They capture long-term dependencies by maintaining a memory cell that selectively retains information over time. This architecture helps mitigate the vanishing gradient problem common in traditional recurrent neural networks. Consequently, LSTMs are well-suited for applications involving sequential data, such as speech recognition, language modeling, and sensor signal analysis.

7.3.7 Autoencoders (AE)

Unsupervised models used for anomaly detection by learning data reconstruction. These models learn to represent normal data patterns by minimizing the difference between the input and its reconstruction. Anomalies are detected when the reconstruction error exceeds a predefined threshold, indicating deviations from learned patterns. Common unsupervised models for this purpose include autoencoders, variational autoencoders, and generative adversarial networks.

7.4 Training and Validation

- **Training set**: 70% of data
- **Validation set**: 15%
- **Test set**: 15%
- Cross-validation for hyperparameter tuning
- Performance metrics considered: Accuracy, Precision, Recall, F1-Score, AUC-ROC, RMSE (for RUL prediction)

8. TESTING & RESULTS

8.1 Evaluation Metrics

Metric	**Definition**
Accuracy	Correct predictions / Total predictions
Precision	True Positives / (True Positives + False Positives)
Recall (Sensitivity)	True Positives / (True Positives + False Negatives)
F1-Score	Harmonic mean of precision and recall

Metric	**Definition**
AUC-ROC	Area under ROC curve
RMSE	Root Mean Square Error (for continuous RUL prediction)

8.2 Comparative Results

Table 2: Classification Performance Summary

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**
LR	82.4%	79.8%	76.3%	78.0%	0.85
SVM	88.9%	87.1%	85.6%	86.3%	0.91
RF	92.5%	91.2%	90.7%	91.0%	0.95
XGBoost	93.8%	93.1%	92.8%	92.9%	0.97
CNN	94.2%	93.5%	93.0%	93.2%	0.98
LSTM	**96.0%**	**95.4%**	**94.8%**	**95.1%**	**0.99**
AE	89.3%	87.7%	85.8%	86.7%	0.88

8.3 Discussion of Results

- **Classical ML models** (LR, SVM) provide baseline performance but are limited in capturing nonlinear patterns.
- **Ensemble models** (RF, XGBoost) significantly improve prediction performance with relatively faster training times.
- **Deep learning models** (CNN, LSTM) outperform classical methods due to their ability to learn hierarchical representations and temporal dynamics.
- **LSTM provided the highest accuracy** and best overall balance, especially for RUL prediction tasks.
- **Autoencoders** are useful as unsupervised anomaly detectors but require careful threshold tuning.

---

9. Conclusion & FUTURE SCOPE

9.1 Summary of Findings

This research demonstrates that machine learning models can effectively support predictive maintenance in smart manufacturing, leading to enhanced equipment reliability and operational efficiency. Key insights include:

- Deep learning methods, particularly LSTM, show superior performance in modeling complex temporal patterns.
- Ensemble models like XGBoost are robust alternatives when computational resources are limited.
- Data preprocessing and feature engineering significantly influence model effectiveness.
- A modular architecture facilitates scalability and integration with existing industrial systems.

9.2 Challenges

- **Data quality and labeling** remain major hurdles.
- **High computational requirements** for training deep models.
- **Interpretability** of complex models needs improvement for operational adoption.

- **Data security and privacy** concerns as sensitive factory data is shared or stored on cloud platforms.

9.3 Future Research Directions

1. **Adaptive and online learning algorithms** for real-time model updates.
2. **Explainable AI (XAI)** to enhance trust and interpretability.
3. **Transfer learning** to reduce data requirements for new assets.
4. **Federated learning** for privacy-preserving collaborative maintenance models.
5. **Integration with Digital Twins** to simulate and predict asset behavior under varying conditions.

---

10. References

1. Zhang, W., Yang, D., & Wang, H. (2015). Data-driven methods for predictive maintenance of industrial equipment: A review. *IEEE Systems Journal*, 11(3), 812-825.

2. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2017). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799-834.

3. Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2018). Long Short-Term Memory Networks for Anomaly Detection in Time Series. *Proceedings of the 23rd ACM SIGKDD International Conference*.

4. Wang, T., Yu, Y., & Yang, J. (2020). Machine learning approaches for predictive maintenance: A review. *Journal of Manufacturing Systems*, 54, 1-12.

5. Nguyen, T., Ding, M., & Seneviratne, A. (2022). Hybrid Models for Predictive Maintenance in Smart Manufacturing. *International Journal of Prognostics and Health Management*, 13(1), 45-61.

6. Bekar, E. T., Nyqvist, P., & Skoogh, A. (2014). An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. *Advances in Mechanical Engineering*, *12*(5), 168781402091920. https://doi.org/10.1177/1687814020919207

7. Besha, A. R. M. A., Ojekemi, O. S., Oz, T., & Adegboye, O. (2025). PLSCO: An Optimization-Driven Approach for Enhancing Predictive Maintenance Accuracy in Intelligent Manufacturing. *Processes*, *13*(9), 2707. https://doi.org/10.3390/pr13092707

8. Taşcı, B., Omar, A., & Ayvaz, S. (2023). Remaining useful lifetime prediction for predictive maintenance in manufacturing. *Computers & Industrial Engineering*, *184*, 109566. https://doi.org/10.1016/j.cie.2023.109566

9. Fathy, Y., Jaber, M., & Brintrup, A. (2020). Learning With Imbalanced Data in Smart Manufacturing: A Comparative Analysis. *IEEE Access*, *9*, 2734–2757. https://doi.org/10.1109/access.2020.3047838

10. Him, L. C., Poh, Y. Y., & Pheng, L. W. (2019). *IoT-based Predictive Maintenance for Smart Manufacturing Systems*. 1942–1944. https://doi.org/10.1109/apsipaasc47483.2019.9023106

11. Moens, P., Bracke, V., Soete, C., Vanden Hautte, S., Nieves Avendano, D., Ooijevaar, T., Devos, S., Volckaert, B., & Van Hoecke, S. (2020). Scalable Fleet Monitoring and Visualization for Smart Machine Maintenance and Industrial IoT Applications. *Sensors (Basel, Switzerland)*, *20*(15), 4308. https://doi.org/10.3390/s20154308

12. Chazhoor, A., Mounika, Y., Vergin Raja Sarobin, M., Sanjana, M. V., & Yasashvini, R. (2020). Predictive Maintenance using Machine Learning Based Classification Models. *IOP Conference Series: Materials Science and*

Engineering, *954*(1), 012001. https://doi.org/10.1088/1757-899x/954/1/012001

13. Borgi, T., Hidri, A., Neef, B., & Naceur, M. S. (2017). *Data analytics for predictive maintenance of industrial robots*. 412–417. https://doi.org/10.1109/aset.2017.7983729

14. Rai, R., Tiwari, M. K., Ivanov, D., & Dolgui, A. (2021). Machine learning in manufacturing and industry 4.0 applications. *International Journal of Production Research*, *59*(16), 4773–4778. https://doi.org/10.1080/00207543.2021.1956675

15. Nangia, S., Makkar, S., & Hassan, R. (2020). IoT based Predictive Maintenance in Manufacturing Sector. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3563559