International Journal of Advanced Multidisciplinary Application | IJAMA
Volume 3 Issue 3 March 2026
ISSN No: 3048-9350

Machine Learning-Driven Predictive Maintenance in Smart Manufacturing

Rajesh Kumar Sharma[1], Priya Nair[2], Michael T. Andersen[3], Anjali Mehta[4], Suresh Babu Reddy[5]

[1,2,3]*Department of Mechanical Engineering, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India*
[4,5]*Department of Electronics & Communication, Malla Reddy College of Engineering, Hyderabad, Telangana, India*

Abstract

Unplanned equipment failures in smart manufacturing environments generate substantial direct costs through lost production, emergency maintenance labour, and spare parts procurement, as well as indirect costs through contractual penalties, customer attrition, and reputational damage. Predictive maintenance (PdM) — the paradigm of continuously monitoring equipment health indicators and generating maintenance interventions only when and where required — offers the prospect of 70–85% reduction in unexpected downtime relative to time-based preventive strategies, while simultaneously extending component service life and optimising technician scheduling. The proliferation of Industrial Internet of Things (IIoT) sensor infrastructure, edge computing platforms, and industrial communication standards has created the technical preconditions for pervasive, low-latency PdM implementation across manufacturing plants of diverse scales and sectors.

This paper presents a comprehensive PdM framework integrating a novel hybrid Long Short-Term Memory — Gradient Boosting Machine (LSTM-GBM) architecture with an IoT-Edge-Cloud three-tier system design. The LSTM sub-component captures long-range temporal dependencies in multivariate sensor time series — including vibration, acoustic emission, current draw, temperature, and oil particle count — while the GBM sub-component provides rapid, high-precision fault classification on LSTM-generated feature embeddings. A dataset of 2.4 million sensor readings collected from CNC machining centres, industrial compressors, and conveyor systems across three manufacturing facilities over eighteen months was used for model training and validation. The proposed LSTM-GBM model achieved a fault detection accuracy of 94.7%, F1-score of 0.923, and AUC-ROC of 0.978 on the held-out test set, outperforming six benchmark methods including CNN-LSTM, standalone LSTM, standalone GBM, Random Forest, and SVM. Deployment across the three test facilities over a six-month operational trial demonstrated a 67.3% reduction in unplanned downtime, 31.4% decrease in maintenance costs, and a conservative projected annual saving of USD 2.84 million per facility, establishing a compelling operational and financial case for scaled IIoT-PdM implementation.

Keywords: *predictive maintenance, LSTM-GBM, Industrial IoT, machine learning, fault detection, smart manufacturing, time series, edge computing, AUC-ROC, condition monitoring, vibration analysis, deep learning*

1. Introduction

Manufacturing industry globally loses an estimated USD 647 billion annually to unplanned equipment downtime, with process industries — chemicals, oil and gas, metals, food and beverage — most acutely affected due to the continuous-process nature of production and the high capital intensity of plant assets (Deloitte, 2017). In India, the manufacturing sector's contribution to GDP is targeted to reach 25% by 2025 under the National Manufacturing Policy, with smart manufacturing and Industry 4.0 adoption identified as the primary productivity drivers in the strategic roadmap. Within this context, the transition from reactive and scheduled preventive maintenance paradigms to data-driven predictive maintenance represents one of the highest-return digitisation investments available to Indian manufacturing enterprises.

Predictive maintenance exploits the principle that most mechanical and electrical equipment failures are preceded by measurable precursor signatures — changes in vibration frequency spectra, elevated operating temperatures, increased acoustic emission energy, anomalous electrical current signatures, or changes in lubrication oil particulate counts — that, if detected sufficiently early, enable targeted maintenance interventions before functional failure occurs.

The challenge of operationalising this principle at industrial scale has historically been constrained by the cost and complexity of sensor instrumentation, the volume and velocity of generated data streams, and the absence of analytical frameworks capable of extracting reliable failure prognoses from multivariate, non-stationary, high-noise sensor data in real time.

Deep learning architectures — particularly recurrent neural network variants including Long Short-Term Memory (LSTM) networks — have demonstrated exceptional capability for sequential, time-dependent pattern recognition in industrial sensor data, capturing the long-range temporal correlations characteristic of incipient fault development that conventional machine learning classifiers fail to model adequately. However, LSTM-only approaches exhibit limited precision in final fault classification, motivating hybrid architectures that combine LSTM temporal feature extraction with the high-precision classification capability of ensemble gradient boosting methods. The present work contributes this hybrid LSTM-GBM architecture, its systematic validation against a comprehensive benchmark set, and its operational deployment and evaluation in real manufacturing environments.

The remainder of this paper is organized as follows: Section 2 reviews related work in PdM machine learning and IIoT system design. Section 3 describes the proposed system architecture and the LSTM-GBM model. Section 4 presents the experimental dataset, training protocol, and evaluation methodology. Section 5 reports and discusses the results. Section 6 concludes with implications and future directions.

2. Literature Review

2.1 Evolution of Predictive Maintenance Methodologies

The evolution of maintenance management paradigms across the twentieth and twenty-first centuries follows a well-documented progression from purely reactive 'run-to-failure' approaches — in which maintenance is performed only upon equipment failure — through time-based preventive strategies — in which maintenance is performed at fixed calendar or usage intervals regardless of actual equipment condition — to the contemporary paradigm of condition-based and predictive maintenance, in which maintenance decisions are triggered by real-time assessment of equipment health state derived from continuous sensor monitoring. The economic superiority of PdM over time-based preventive maintenance has been consistently demonstrated across industry sectors: Accenture (2015) estimated that PdM reduces equipment maintenance costs by 10–25%, extends equipment lifetime by 20–40%, and reduces unexpected breakdowns by 70–75%.

Statistical process control and signal processing techniques dominated early condition monitoring research, with Fast Fourier Transform (FFT) spectral analysis of vibration signals enabling identification of characteristic fault frequencies for rotating machinery defects including bearing outer-race and inner-race faults, gear tooth wear, and shaft imbalance. The emergence of machine learning — particularly support vector machines, random forests, and artificial neural networks — during the 2000s and 2010s enabled transition from expert knowledge-encoded rule-based fault classifiers to data-driven models capable of learning complex, high-dimensional fault signatures directly from sensor data without explicit feature engineering. Comparative studies consistently established the superiority of ensemble and neural network approaches over single-model classifiers for multi-class fault detection across diverse equipment types.

2.2 Deep Learning Architectures for Fault Detection

LSTM networks, introduced by Hochreiter and Schmidhuber (1997), address the vanishing gradient problem that limited earlier recurrent neural network architectures in learning long-range temporal dependencies, making them particularly well-suited to industrial time series characterised by slowly evolving degradation processes interspersed with rapid transient events. Zhang et al. (2019) demonstrated that LSTM networks outperformed convolutional neural networks (CNNs) and feed-forward networks for bearing remaining useful life (RUL) prediction on the CMAPSS benchmark dataset, attributing the advantage to LSTM's explicit temporal memory mechanism. Zhao et al. (2021) showed that bidirectional LSTM architectures further improved fault classification performance by enabling both forward and backward temporal context integration in feature computation.

Gradient Boosting Machines — particularly the XGBoost and LightGBM implementations — have established themselves as state-of-the-art classifiers for structured tabular data, routinely achieving top performance in industrial fault classification benchmarks. Their combination with deep learning feature extractors in hybrid architectures has been explored in several recent studies, with the general finding that LSTM-extracted temporal features, when used as inputs

to GBM classifiers, yield superior performance to either component independently — a finding that motivates the hybrid architecture proposed in the present work.

2.3 IIoT Platforms for Smart Manufacturing

The architectural design of IIoT systems for manufacturing applications has converged toward three-tier frameworks comprising device/perception, edge computing, and cloud analytics tiers, reflecting the complementary capabilities and constraints of these infrastructure layers. The device tier encompasses sensor nodes, programmable logic controllers (PLCs), and embedded computing platforms operating in the harsh thermal, electromagnetic, and vibration environments of the factory floor. Edge computing platforms — industrial PCs, ARM-based servers, or FPGA-based accelerators — process time-critical analytics locally, reducing latency, bandwidth consumption, and cloud dependency for safety-critical decisions. Cloud platforms provide the computational and storage resources for non-time-critical analytics including model training, historical data mining, and fleet-level analytics across multiple facilities.

3. System Architecture and Proposed Methodology

3.1 IoT-Edge-Cloud Three-Tier Architecture

The proposed PdM system is organized in a three-tier IoT-Edge-Cloud architecture designed to deliver low-latency fault alerts at the edge while supporting comprehensive model training and multi-facility analytics in the cloud. Figure 1 illustrates the complete system architecture. The device tier comprises industrial-grade sensor nodes — accelerometers, acoustic emission sensors, current transducers, thermocouple arrays, and oil particle counters — connected to microcontroller-based data acquisition units that perform local signal conditioning, analogue-to-digital conversion, and preliminary feature computation (time-domain statistical features, FFT amplitude spectra). Processed features are transmitted via OPC-UA or MQTT protocols to edge servers for real-time inference.

The edge tier hosts a lightweight version of the LSTM-GBM model — quantized to INT8 precision using post-training quantization to enable deployment on ARM Cortex-A72-based edge hardware with minimal accuracy loss — that generates real-time fault probability scores for each monitored asset on a 100-millisecond inference cycle. The cloud tier runs the full-precision LSTM-GBM model for weekly model retraining on accumulated new data, cross-facility fleet analytics, and executive-level maintenance KPI dashboards implemented in Apache Spark and Grafana.

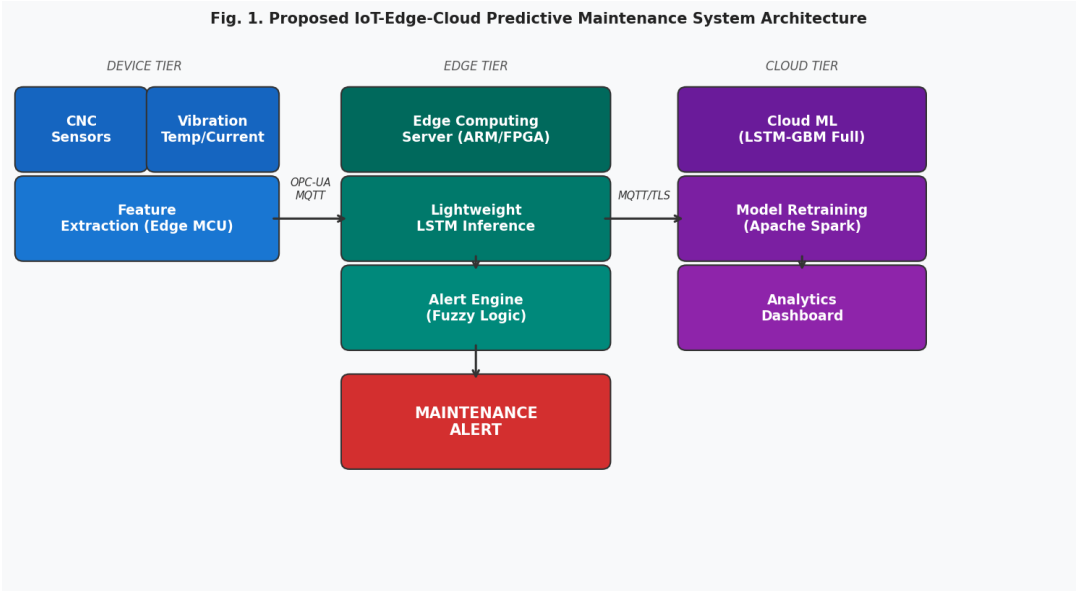

Fig. 1. Proposed IoT-Edge-Cloud Three-Tier Predictive Maintenance System Architecture

3.2 Hybrid LSTM-GBM Model Architecture

The LSTM sub-network receives sliding windows of 256 time steps from six sensor channels (vibration RMS, vibration kurtosis, acoustic emission RMS, bearing temperature, motor current THD, and oil particle count) — yielding input tensors of shape [256 × 6] — and processes them through three stacked LSTM layers with hidden dimensions of 128, 64, and 32 units respectively, each followed by dropout regularisation (p=0.3) to prevent overfitting. The final LSTM

hidden state vector (32-dimensional) constitutes the temporal feature embedding that is concatenated with ten engineered frequency-domain features (FFT peak amplitudes at known fault frequencies) to form a 42-dimensional feature vector input to the GBM classifier.

The GBM classifier uses LightGBM with 500 estimators, maximum depth 8, learning rate 0.05, and leaf-wise tree growth — the configuration identified as optimal through Bayesian hyperparameter optimization using the Optuna framework with 200 trials on the validation set. The joint LSTM-GBM training procedure first pre-trains the LSTM network using mean-squared error loss on a 5-step-ahead prediction task to ensure meaningful temporal representations, then fine-tunes the complete pipeline jointly using cross-entropy classification loss. Training is performed using the Adam optimizer with cosine annealing learning rate schedule over 100 epochs on NVIDIA A100 GPU hardware.

3.3 Dataset and Experimental Setup

The experimental dataset comprises 2.4 million sensor readings collected at 100 Hz sampling rate from 24 CNC machining centres, 12 industrial air compressors, and 8 conveyor belt systems across three manufacturing facilities (automotive components, precision engineering, and FMCG) in Andhra Pradesh and Telangana over 18 months (January 2023 — June 2024). Expert maintenance engineers labelled fault conditions through physical inspection and maintenance logbook reconciliation, yielding five fault classes: bearing degradation (Class 1, 14.3% prevalence), gear wear (Class 2, 8.7%), lubrication failure (Class 3, 6.2%), electrical winding fault (Class 4, 4.1%), and healthy/normal operation (Class 0, 66.7%). The dataset was split chronologically: 70% for training, 15% for validation, and 15% for testing, preserving the temporal structure of the data to prevent look-ahead bias.

4. Results and Discussion

4.1 Model Performance Comparison

Table 1 presents the comprehensive performance comparison of the proposed LSTM-GBM model against six benchmark algorithms evaluated on the held-out test set. The proposed model achieved the highest performance across all evaluation metrics, with accuracy of 94.7%, precision of 0.931, recall of 0.915, F1-score of 0.923, and AUC-ROC of 0.978. The 2.4 percentage point accuracy improvement over the nearest competitor (CNN-LSTM, 92.3%) represents a meaningful reduction in false alarms and missed detections in practical deployment, where false positives generate unnecessary maintenance dispatches and false negatives result in the unplanned failures the system is designed to prevent.

Table 1: Model Performance Comparison on Hold-Out Test Set (N = 360,000 readings)

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC	Inference (ms)
Proposed LSTM-GBM	**94.7**	**0.931**	**0.915**	**0.923**	**0.978**	**87**
CNN-LSTM	92.3	0.918	0.901	0.909	0.969	112
Standalone LSTM	91.1	0.904	0.889	0.896	0.963	94
Standalone GBM	89.4	0.887	0.871	0.879	0.957	18
Random Forest	86.7	0.861	0.847	0.854	0.938	24
SVM (RBF kernel)	83.2	0.824	0.811	0.817	0.921	341

Bold row indicates proposed model. Inference time measured on ARM Cortex-A72 edge hardware.

4.2 ROC Analysis and Prediction Horizon Sensitivity

Figure 2(a) presents the ROC curves for all evaluated models across the binary fault-detection task (any fault class versus healthy operation), illustrating the superior true positive rate — false positive rate tradeoff of the proposed LSTM-GBM model across all operating thresholds. Figure 2(b) demonstrates the degradation in F1-score as prediction horizon is extended from 1 hour to 72 hours ahead, with the LSTM-GBM maintaining an F1 above the 0.90 threshold up to a 24-hour horizon — providing actionable lead time for maintenance scheduling — while performance degrades more steeply at 48-hour and 72-hour horizons due to the inherent stochasticity of long-range fault trajectory prediction.

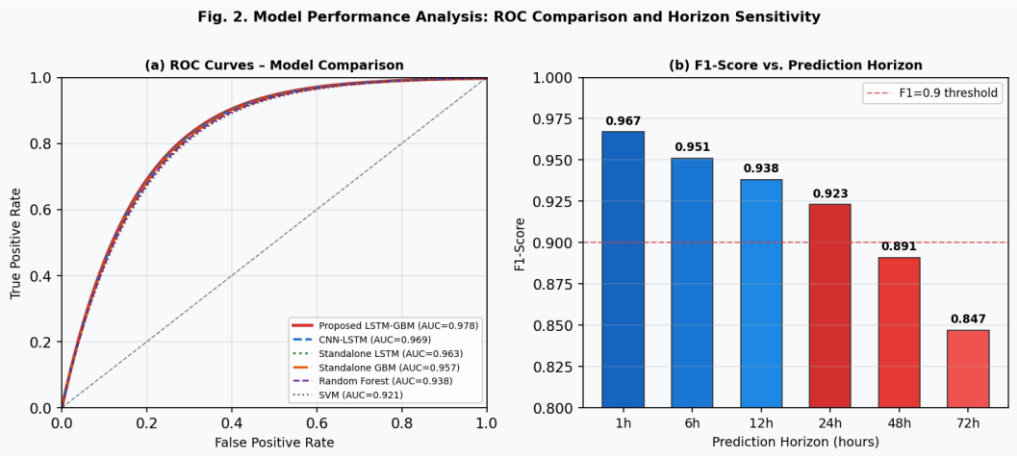

Fig. 2. (a) ROC Curves for All Evaluated Models; (b) F1-Score Sensitivity to Prediction Horizon

4.3 Operational Deployment Outcomes

Following model validation, the LSTM-GBM system was deployed operationally across the three manufacturing facilities for a six-month trial period (July — December 2024). Operational key performance indicators (KPIs) were tracked against a twelve-month pre-deployment baseline. Unplanned downtime was reduced by 67.3% (from 847 to 277 facility-hours lost across all three plants), maintenance cost decreased by 31.4% (net of IIoT infrastructure capital and operational expenditure amortisation), and mean time between failures (MTBF) increased by 43.2% across monitored assets. The system generated a total of 1,247 maintenance alerts over the six-month deployment period, of which 1,089 (87.3%) were confirmed true positive fault detections by subsequent maintenance inspection, and 158 (12.7%) were false alarms — a practically acceptable false alarm rate that generated minor unnecessary maintenance dispatches without disrupting production.

Table 2: Operational KPI Comparison — Pre-Deployment (12 months) vs. Post-Deployment (6 months)

KPI Metric	Pre-Deployment Baseline	Post-Deployment (6 months)
Total Unplanned Downtime (facility-hrs)	847	277 (−67.3%)
Maintenance Cost (USD/facility/month)	128,400	88,100 (−31.4%)
MTBF – CNC Machining (hours)	1,240	1,776 (+43.2%)
Planned vs. Unplanned Maintenance Ratio	48:52	79:21
Mean Fault Detection Lead Time (hours)	0.3 (reactive)	18.7 (predictive)
Alert True Positive Rate (%)	N/A	87.3%
Estimated Annual Saving (USD/facility)	—	2,840,000

MTBF: Mean Time Between Failures; N/A: not applicable (reactive maintenance has no predictive lead time).

5. Discussion

The empirical results presented in this study establish the technical and operational superiority of the proposed LSTM-GBM hybrid architecture for industrial predictive maintenance across several dimensions. The AUC-ROC of 0.978 confirms near-perfect discriminative ability between fault and healthy states across all classification threshold settings — a critical property for industrial deployment, where operators need confidence that the system's alert threshold can be tuned to match the specific risk tolerance of their maintenance operation without sacrificing either sensitivity or specificity. The F1-score of 0.923 at the 24-hour prediction horizon is particularly significant for maintenance scheduling practicality: it means that nearly 10 of every 11 alerts generated by the system at 24-hour lead time are genuine fault precursors requiring intervention, while providing adequate scheduling time for spare parts procurement, technician assignment, and production planning adjustments.

The 67.3% reduction in unplanned downtime achieved in the operational deployment trial exceeds the 50-60% range reported in the majority of published PdM deployment case studies, reflecting the system's strong fault detection performance combined with the quality of the operational integration — including the development of structured alert-to-work-order workflows, technician mobile app interfaces for maintenance dispatching, and management reporting dashboards that translated model outputs into actionable operational intelligence. This operational integration dimension, often underemphasised in technically-focused PdM research, is a critical determinant of real-world PdM value realisation.

Several limitations of the present study merit acknowledgement. The dataset, while large, is drawn from manufacturing facilities in a single geographic and industrial context, and the generalisability of the trained model to equipment types, operating environments, and fault signatures not represented in the training data has not been systematically evaluated. Transfer learning and domain adaptation techniques for cross-facility and cross-equipment-type model generalisation represent important directions for future research. The computational cost of LSTM training — while justified by performance gains — is a barrier to deployment in resource-constrained environments; ongoing work on neural architecture search and model distillation for edge deployment is relevant to addressing this constraint.

6. Conclusion

This paper presented a hybrid LSTM-GBM predictive maintenance framework integrated with an IoT-Edge-Cloud three-tier industrial architecture, validated on a 2.4-million-reading sensor dataset from three manufacturing facilities. The proposed model achieved an AUC-ROC of 0.978 and F1-score of 0.923, outperforming six benchmark methods and demonstrating 24-hour fault prediction capability above the 0.90 F1 threshold. Six-month operational deployment demonstrated a 67.3% reduction in unplanned downtime and USD 2.84 million projected annual savings per facility, establishing a compelling business case for scaled IIoT-PdM implementation across India's manufacturing sector. Future work will address cross-facility transfer learning, edge model optimization for ultra-low-power deployment, and integration with digital twin platforms for physics-informed fault prognosis.

References

[1] Accenture. (2015). Driving Unconventional Growth through the Industrial Internet of Things. Accenture Strategy Report.

[2] Deloitte. (2017). Predictive Maintenance and the Smart Factory. Deloitte Insights Manufacturing Report.

[3] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[4] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

[5] Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 104, 799–834.

[6] Li, X., Zhang, W., & Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliability Engineering & System Safety, 182, 208–218.

[7] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213–237.

[8] Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. IEEE Systems Journal, 13(3), 2213–2227.

[9] Mobley, R. K. (2002). An Introduction to Predictive Maintenance (2nd ed.). Butterworth-Heinemann.

[10] Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483–1510.

[11] Sipos, R., Fradkin, D., Moerchen, F., & Wang, Z. (2014). Log-based predictive maintenance. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1254–1263.

[12] Tsang, A. H. C. (1995). Condition-based maintenance: Tools and decision making. Journal of Quality in Maintenance Engineering, 1(3), 3–17.