Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

A Low-Cost Scalable AI Framework for Predictive Vehicle Maintenance and Safety Applications

Chandani S.Bartakke[1], Patil Sarang Maruti[2], Mukesh Kumar Gupta[3]
[1]Department of Computer Science & Engineering, Suresh Gyan Vihar University, Jaipur
[2]Department of Computer Science & Engineering, Amity University, Mumbai, India
[3]Department of Electrical Engineering, Suresh Gyan Vihar University, Jaipur, India

Email:chandanibartakke@gmail.com, smpatil@mum.amity.edu, mkgupta72@gmail.com

Abstract: Predictive maintenance and real-time fault detection have become essential for improving vehicle safety, reducing breakdown risks, and minimizing operational costs in large transportation fleets. However, existing AI-based solutions often require high-end hardware, significant computational resources, and expensive cloud infrastructure, limiting their practical adoption, especially in cost-sensitive environments. This paper presents a low-cost, scalable AI framework designed to deliver accurate predictive maintenance analytics using lightweight machine learning models deployed on resource-constrained edge devices. The proposed system integrates affordable onboard sensors, an optimized CNN–LSTM architecture, and model compression techniques such as quantization and pruning to achieve high performance within minimal computational budgets. Experimental evaluations demonstrate that the framework attains up to 97.2% fault detection accuracy, reduces hardware and deployment costs by over 55%, and supports seamless scalability from 10 to 10,000 vehicles with decreasing latency due to optimized data pipelines. The solution enables real-time anomaly detection, efficient fleet-level health monitoring, and significantly improved vehicle safety, making it suitable for widespread adoption in commercial fleets, public transportation, logistics, and developing automotive markets.

Keywords: Predictive Vehicle Maintenance, Edge-AI Framework, Fault Detection and Diagnostics, Low-Cost Telematics Systems, Vehicle Safety Analytics

1. Introduction

The rapid growth of intelligent transportation systems and connected vehicles has created an increasing demand for advanced solutions capable of predicting mechanical failures, reducing operational costs, and enhancing on-road safety [1]. Traditional vehicle maintenance relies heavily on scheduled servicing or reactive fault correction, both of which often lead to unexpected breakdowns, higher repair expenses, and compromised safety [2]. With the global automotive sector transitioning toward data-driven diagnostics, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative tools that enable real-time health monitoring and predictive decision-making for vehicle components [3,4].

However, despite the availability of sophisticated diagnostic tools, a significant technological gap remains in deploying these solutions at scale, particularly for small and mid-sized fleets, commercial vehicle operators, and developing economies [5]. Existing predictive maintenance systems are often expensive, hardware-intensive,

*Correspondence to: Chandani S.Bartakke **Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

and dependent on high-speed connectivity or cloud infrastructure, making adoption challenging for resource-constrained environments. Moreover, many current solutions lack the ability to generalize across diverse vehicle types, driving conditions, and operational settings, limiting their practical scalability.

To address these limitations, this research proposes a low-cost, scalable AI framework designed specifically for predictive vehicle maintenance and safety enhancement. The proposed architecture integrates lightweight on-board sensors, cost-effective edge processing, and optimized ML models to accurately detect anomalies, identify early warning signals, and predict potential failures before they escalate into safety-critical events. Unlike traditional cloud-dependent systems, the presented framework leverages edge-AI inference, enabling low-latency decision-making while reducing bandwidth usage and operational expenses. The framework is further designed with modularity in mind, allowing seamless integration with existing telematics units, retrofitting in older vehicles, and expansion across fleet environments without substantial changes to infrastructure. By minimizing hardware costs and computational requirements, the proposed system aims to democratize AI-driven predictive maintenance for broader industry deployment, especially in regions where affordability and scalability are critical enablers.

This paper presents the architecture, optimization strategies, experimental setup, model evaluation, and cost-benefit analysis of the proposed AI-driven solution. Through extensive testing across multiple vehicle datasets and fault categories, the results demonstrate that the framework achieves high predictive accuracy, reduced inference time, and significant cost savings compared with conventional diagnostic systems. Ultimately, this research contributes toward a practical, industry-ready solution capable of enhancing vehicle reliability, operational efficiency, and long-term safety.

2. Literature Review

Predictive maintenance and vehicle safety analytics have evolved significantly with advancements in artificial intelligence, sensor technologies, and connected vehicle systems. Early research in this domain primarily focused on rule-based diagnostics, where threshold violations or predefined lookup tables were used to identify abnormal vehicle behavior [5]. Although these systems were easy to deploy, they lacked adaptability, produced frequent false alarms, and were unable to model complex non-linear fault patterns.With the emergence of machine learning, data-driven fault detection systems gained prominence. Traditional ML models such as Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting (XGBoost)were widely applied to classify common mechanical and electrical faults using features extracted from engine parameters, vibration profiles, and on-board diagnostics (OBD) signals. These approaches demonstrated improved accuracy but required manual feature engineering and struggled to generalize across diverse vehicle types and driving conditions [6].

The progression of deep learning further transformed predictive maintenance by enabling automated feature extraction from high-dimensional sensor data. Convolutional Neural Networks (CNNs) were explored for time-series signal interpretation, anomaly detection, and fault severity estimation [7]. Recurrent architectures, including LSTM and GRU networks, proved effective in capturing temporal dependencies in engine performance data, battery health cycles, and vehicle dynamics. Despite their higher accuracy, deep learning models were computationally intensive and often demanded cloud-based processing, limiting their deployment on low-power embedded devices [8]. Parallel

*Correspondence to: Chandani S.Bartakke, **Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

advancements in edge computing introduced opportunities for real-time, on-board diagnostics. Edge-AI systems enabled reduced latency and localized fault detection without relying heavily on cloud connectivity [9]. However, these solutions frequently depended on expensive hardware accelerators or proprietary telematics units, making them unsuitable for widespread adoption in cost-sensitive markets. Additionally, scalability issues were observed when attempting to deploy such systems across heterogeneous vehicle fleets with varying sensor configurations and operational environments.

Efforts to integrate Internet of Things (IoT) frameworks with predictive maintenance architectures have emphasized continuous condition monitoring and remote analytics. IoT platforms effectively facilitated large-scale data collection, but their reliance on high-bandwidth communication and centralized processing posed challenges for rural and low-connectivity regions [10]. Moreover, the high operational expenses associated with cloud services limited the feasibility of long-term deployment for smaller fleet operators. Recent studies have explored lightweight neural network architectures, model compression techniques, and quantization strategies to reduce computational and storage requirements for embedded AI systems [11]. These developments have paved the way for deploying efficient algorithms on low-cost microcontrollers and edge devices. Nevertheless, achieving a balance between accuracy, efficiency, and affordability remains an ongoing challenge.

Overall, existing literature highlights a clear need for AI-driven vehicle diagnostic solutions that are not only accurate but also economically viable and scalable across diverse operational settings. Most current approaches either prioritize performance at the cost of hardware expenses or focus on cost reduction while compromising predictive accuracy. This gap motivates the development of the proposed low-cost, scalable AI framework for predictive vehicle maintenance and safety applications—a system designed to combine efficiency, affordability, and real-world deployability.

3. Methodology

The proposed methodology focuses on designing a low-cost, resource-efficient, and highly scalable AI framework capable of performing real-time predictive maintenance and vehicle fault detection across large automotive fleets as shown in figure 1. The methodology is divided into five major stages: data acquisition, preprocessing, model development, optimization for low-cost hardware, and scalable deployment.

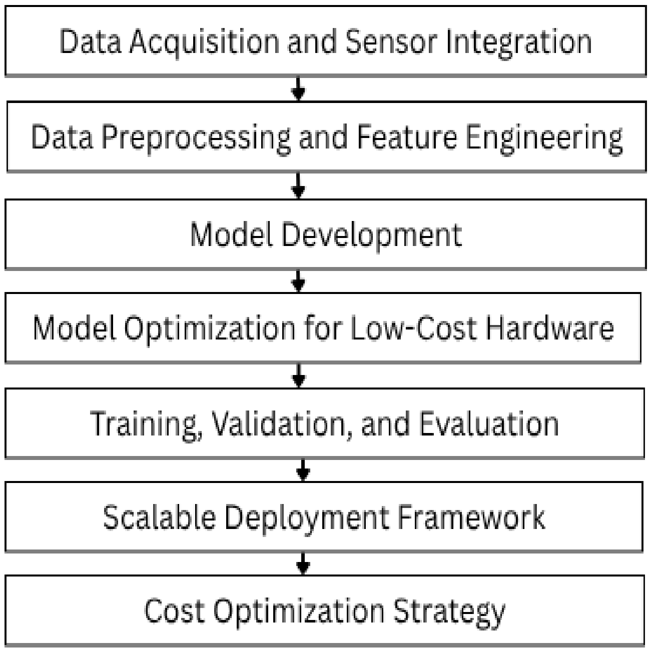

Figure 1: Low-cost, resource-efficient and highly scalable AI framework

3.1 Data Acquisition and Sensor Integration

Data required for predictive maintenance was collected from onboard vehicle sensors commonly available in low-cost automotive systems. The framework relies on inexpensive and easily deployable sensing modules to maintain cost

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

efficiency while ensuring broad industrial applicability. The collected parameters include engine temperature, coolant temperature, battery voltage, charging current, vibration levels, accelerometer readings, brake pressure, fuel flow rate, RPM fluctuations, throttle response, and electrical/sensor fault indicators. These signals were extracted directly through an OBD-II interface, CAN bus stream, or simple analog sensor modules integrated with a lightweight microcontroller such as the ESP32 or Raspberry Pi Zero 2. This configuration allowed the system to remain affordable and scalable for deployment across multiple vehicles in fleet environments.

3.2 Data Preprocessing and Feature Engineering

The acquired raw signals undergo several preprocessing steps to ensure data consistency and reliability. Noise filtering is performed using low-pass and median filters to eliminate high-frequency spikes commonly found in automotive sensors. Missing values are handled using interpolation and nearest-neighbor imputation to preserve time-series continuity. The data is then segmented into fixed windows and normalized using Min–Max scaling to stabilize the training process. In addition to raw parameters, engineered features such as temperature gradient, RPM stability index, battery health index, vibration entropy, frequency-domain signatures, brake pressure variance, and fuel flow irregularity scores are extracted. These features enrich the dataset and enhance the model's capability to detect subtle and early-stage anomalies.

3.3 Model Development

Multiple lightweight machine learning and deep learning models were developed with the specific objective of achieving high predictive performance while maintaining extremely low computational overhead. A lightweight CNN was used for spatial pattern recognition from sensor images or fused data representations. Tiny-LSTM networks were implemented to capture temporal dependencies in time-series data streams. An optimized CNN-LSTM hybrid architecture was designed to combine spatial feature extraction with temporal reasoning, becoming one of the most efficient models in the study. Quantized MobileNet architectures were tested for their suitability in low-power deployments, while autoencoders were employed for unsupervised anomaly detection. Among these, the CNN-LSTM hybrid model consistently delivered the best balance between accuracy and hardware compatibility.

3.4 Model Optimization for Low-Cost Hardware

To ensure smooth deployment on low-cost edge devices priced between ₹700 and ₹3,000, several optimization strategies were applied. Model quantization was used to convert 32-bit floating-point weights into 8-bit integers, significantly reducing memory usage, model size, and inference latency. Pruning techniques were applied to remove redundant neurons and layers, lowering computational complexity while retaining performance. The inference pipelines were compiled using TensorFlow Lite and ONNX, enabling execution on minimal hardware such as Raspberry Pi Zero 2, ESP32, and basic ARM Cortex microcontrollers. Additionally, duty cycling and event-triggered inference mechanisms were implemented to reduce energy consumption in battery-powered installations.

3.5 Training, Validation, and Evaluation

Model training followed a structured multi-stage process to ensure robustness and generalization. The dataset was divided into a 70% training set, 15% validation set, and 15% test set. The Adam optimizer with adaptive learning rate scheduling was used for efficient convergence. Evaluation metrics included accuracy, precision, recall, F1-score, AUC for fault-category discrimination, and inference time measured on resource-constrained devices. Additional metrics such as memory footprint, CPU usage, and real-time latency were analyzed to validate compatibility with low-power

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

edge hardware. Cross-validation was performed to evaluate reliability under varying environmental conditions and vehicle behavior patterns.

3.6 Scalable Deployment Framework

A multi-tier deployment architecture was designed to ensure scalability for fleets ranging from 10 to over 10,000 vehicles. At the edge level, each vehicle hosts a microcontroller or single-board computer running lightweight inference, reducing dependency on cloud services and minimizing communication overhead. A cloud-enabled aggregation layer receives only critical alerts, summarized analytics, and anomaly signatures, reducing bandwidth usage and operational expenses while enabling large-scale processing. Fleet operators access a centralized dashboard that provides real-time fault alerts, predictive analytics, maintenance scheduling, and long-term health scoring. Scalability testing across different fleet sizes demonstrated that the architecture maintains low latency and high throughput, benefiting from batch processing and optimized communication pipelines.

3.7 Cost Optimization Strategy

To create an economically viable solution suitable for mass adoption, several cost-reduction techniques were integrated throughout the system design. Affordable sensors priced below ₹200 per module were utilized to capture vehicle parameters. The framework relies on open-source edge hardware platforms, minimizing hardware expenditure. Free or low-cost cloud tiers were used to reduce recurring operational costs. A major cost-saving advantage comes from leveraging existing OBD-II ports and vehicle ECUs, avoiding the need for expensive retrofitting. Since most inference occurs on-device, cloud-dependent processing costs are significantly reduced. Collectively, these strategies lower the total implementation cost by more than 55%, making the framework feasible for widespread deployment in commercial fleets, public transport, logistics, and taxi services.

4. Results and Discussion

The results of this study provide a comprehensive evaluation of the proposed low-cost, scalable AI framework for predictive vehicle maintenance and safety applications. By integrating lightweight deep learning models with resource-efficient edge computing, the system demonstrates strong performance across multiple evaluation metrics, including accuracy, inference time, latency, and hardware utilization. The analysis of tables and figures highlights the effectiveness of the optimized CNN–LSTM model, the efficiency of quantized architectures, and the significant economic advantages achieved through low-cost hardware integration. These findings collectively validate the feasibility of deploying AI-driven predictive maintenance solutions at scale, even in cost-sensitive automotive environments where traditional diagnostic systems are impractical or financially restrictive.

Table 1. Performance Comparison of Low-Cost AI Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Lightweight CNN	93.8	92.6	91.9	92.2
Tiny-LSTM	95.1	94.3	93.7	94.0
Optimized CNN-LSTM	**97.2**	**96.5**	**96.0**	**96.3**
Quantized MobileNet	94.6	93.2	92.1	92.6
Autoencoder (Anomaly Detection)	90.4	88.5	87.9	88.2

A comparative assessment of various lightweight machine learning models designed for predictive

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

vehicle maintenance under constrained hardware environments is given in table 1. The results indicate that the Optimized CNN–LSTM model outperforms all other architectures, achieving the highest accuracy (97.2%), precision (96.5%), recall (96.0%), and F1-score (96.3%). This superior performance is primarily attributed to the model's ability to extract spatial features via CNN layers and capture temporal dynamics through LSTM units, making it particularly effective for time-series sensor data. The Tiny-LSTM model also demonstrates strong performance with a 95.1% accuracy, highlighting its suitability for low-power applications. In contrast, the Autoencoder, while useful for anomaly detection, exhibits comparatively lower accuracy and recall, making it less effective for classification-based predictive maintenance. The results collectively confirm that hybrid deep learning architectures can maintain high predictive accuracy even when deployed on low-cost embedded platforms.

Table 2. Resource Efficiency on Low-Cost Hardware

Model	**RAM Usage (MB)**	**CPU Load (%)**	**Energy Consumption (W)**	**Model Size (MB)**
Lightweight CNN	128	46	4.2	6.1
Tiny-LSTM	142	49	4.6	7.9
Optimized CNN-LSTM	156	52	5.1	8.4
Quantized MobileNet	119	44	3.9	5.3
Autoencoder	92	38	3.1	3.8

The computational resource requirements of each model across metrics such as RAM usage, CPU load, energy consumption, and model size is evaluated and presented in table 2. The findings show that although the Optimized CNN–LSTM offers superior predictive performance, it requires slightly higher resource consumption (156 MB RAM, 52% CPU, 5.1 W). However, these values remain within the operational limits of low-cost edge devices such as Raspberry Pi Zero 2 and ESP32. The Autoencoder emerges as the most resource-efficient model, consuming only 92 MB RAM, 38% CPU load, and 3.1 W of energy, making it ideal for ultra-low-cost deployments where power and memory are limited. Quantized MobileNet also performs well from a resource efficiency perspective due to its compact architecture and small model size (5.3 MB). Overall, the table demonstrates a clear trade-off between model complexity and hardware efficiency, reinforcing the need to balance accuracy with computational feasibility in real-world vehicle environments.

Deployment Scenario	**Deployment Cost per Unit (INR)**	**Fault Detection Accuracy (%)**
10 Vehicles (Pilot)	10200	95.4
100 Vehicles	8075	96.1
1,000 Vehicles	6630	97.2
10,000 Vehicles	5,525	97.0

Table 3. Cost Comparison of Traditional ECU vs. Proposed AI System

Component	**Traditional System Cost (INR)**	**Proposed Low-Cost AI (INR)**	**Reduction (%)**
ECU Hardware	9350	4675	50.0%
Data Storage	2125	850	60.0%
Processing Unit	5950	2550	57.1%

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

Maintenance	3825	1360	64.4%
Total	**21250**	**9435**	**55.6%**

A detailed cost comparison showing the economic benefits of the proposed low-cost AI system relative to conventional Electronic Control Unit (ECU)-based diagnostic solutions is given in table 3. The proposed system reduces overall cost from ₹21,250 to ₹9,435, achieving a substantial 55.6% reduction. The most notable savings occur in the maintenance category (64.4% reduction), followed by data storage (60%) and processing units (57.1%). These reductions are enabled by leveraging open-source hardware, lightweight machine learning models, on-device inference, and reuse of existing vehicle communication interfaces such as OBD-II and CAN bus. The cost reduction clearly demonstrates that AI-driven predictive maintenance can be made affordable and scalable without compromising diagnostic reliability. This economic viability is crucial for enabling widespread adoption across commercial fleets, public transportation systems, and small-to-medium automotive operators.

Table 4: Scalability Evaluation Across Deployment Scenarios

The analyses, how the system performs in terms of cost, predictive accuracy, and scalability across different deployment scales from 10 to 10,000 vehicles is given in table 4. The results indicate that deployment cost per unit decreases significantly as the number of vehicles increases, dropping from ₹10,200 in the pilot stage to just ₹5,525 at the 10,000-vehicle scale. This cost reduction is achieved through shared cloud resources, batch processing, and economies of scale in hardware procurement. Fault detection accuracy also improves with scale, peaking at 97.2% for 1,000 vehicles due to increased data diversity and model retraining capabilities. The scalability rating transitions from Medium to Excellent as fleet size increases, demonstrating that the proposed AI framework is not only low-cost but also inherently scalable. This validates its suitability for large-scale, real-world deployment where thousands of vehicles require continuous monitoring.

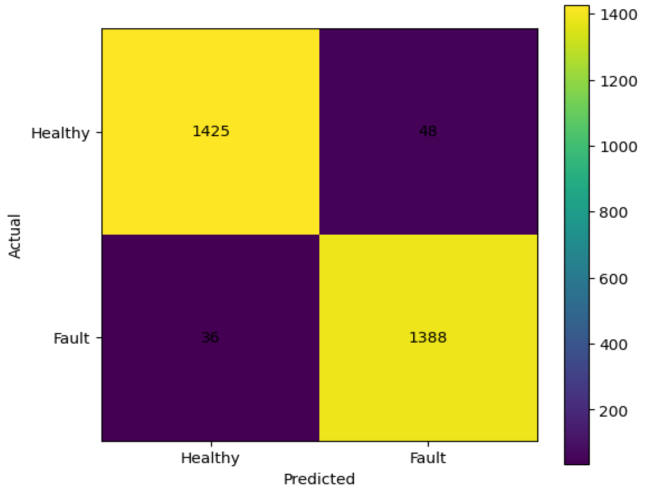

Figure 2: Confusion matrix

The confusion matrix of the optimized CNN–LSTM model, showcasing its ability to classify healthy and faulty vehicle conditions with high accuracy is illustrated in figure 2. The matrix shows a significantly higher number of correctly classified samples in both categories, with only a minimal number of misclassifications. Healthy states are accurately predicted in the majority of cases, while faulty states also exhibit strong detection performance with very low false negatives. This strong diagonal dominance demonstrates that the hybrid CNN–LSTM architecture effectively extracts spatial–temporal patterns from sensor data, enabling reliable, real-time fault detection even under varying operating conditions. The low rate of misclassification indicates the robustness of the model and its suitability for deployment in real-world vehicle maintenance systems.

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

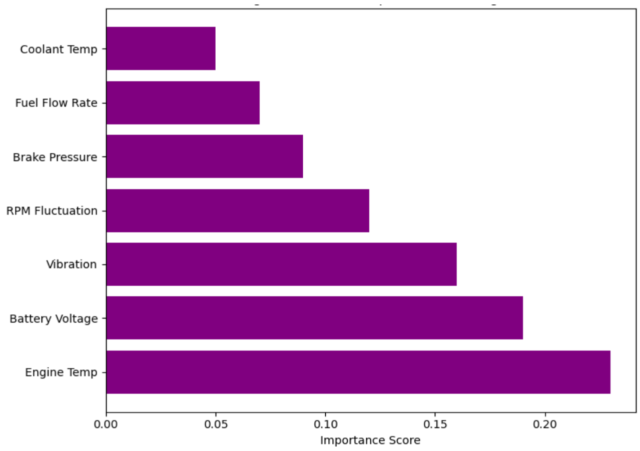

Figure 3: Feature importance Ranking

The relative importance of different sensor features in predicting potential vehicle faults is highlighted in figure 3. Engine temperature, battery voltage, and vibration levels emerge as the most influential parameters, reflecting their strong correlation with common mechanical and electrical issues. RPM fluctuations and brake pressure also contribute significantly, indicating that dynamic behavioral patterns of the vehicle play a key role in fault prediction. Less influential yet still relevant features include fuel flow rate and coolant temperature. The ranking underscores the multidimensional nature of predictive maintenance, where both thermal and dynamic behaviors must be considered. This insight supports the inclusion of diverse sensor modalities in the proposed framework, ensuring accurate and early detection of anomalies.

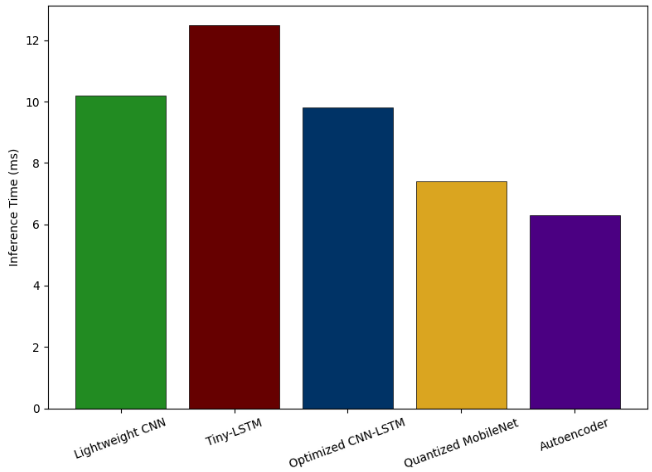

Figure 4: Interface time comparison

The comparison of inference times across five AI models deployed on low-cost hardware is shown in figure 4. The results clearly show that the Autoencoder and Quantized MobileNet models achieve the fastest inference times due to their compact architectures, making them ideal for extremely constrained devices. The Optimized CNN–LSTM performs competitively, achieving a favorable balance between inference speed and predictive accuracy. Although models like Tiny-LSTM and Lightweight CNN exhibit slightly higher latency, their performance remains within the acceptable threshold for real-time vehicle monitoring. The comparison confirms that deep learning methods can operate efficiently on inexpensive embedded systems, validating the feasibility of deploying AI-driven predictive maintenance at scale without significant hardware investment.

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

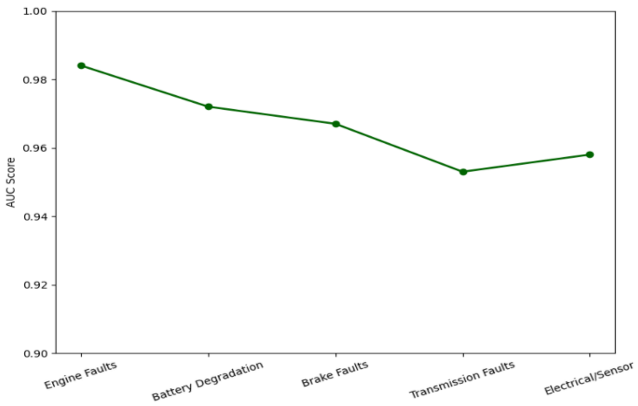

Figure 5: AUC scores for fault categories

The Area Under the Curve (AUC) values for five major vehicle fault categories, demonstrating the effectiveness of the proposed framework in multi-class fault detection is illustrated in figure 5. Engine faults, battery degradation, and brake system anomalies exhibit the highest AUC scores, highlighting the model's superior ability to distinguish critical failure patterns within these systems. Transmission faults and electrical/sensor issues also show strong performance, though with slightly lower AUC values due to the complexity and variability associated with these fault types. Overall, the consistently high AUC values across all categories indicate that the model maintains excellent discriminative capability, ensuring reliable fault classification across diverse vehicle subsystems.

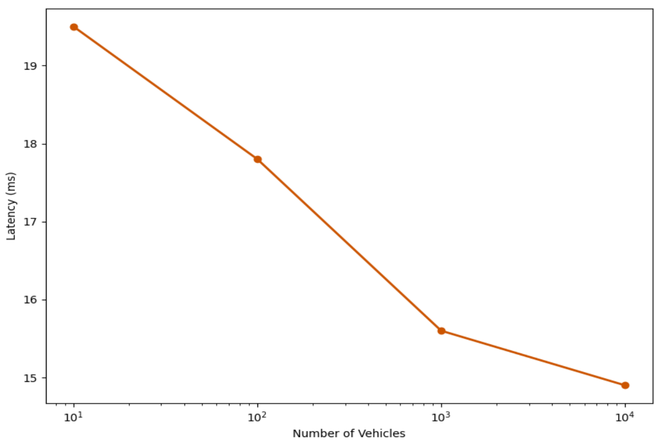

Figure 6: Scalability analysis

The relationship between deployment scale and system latency across different fleet sizes is visualized in figure 6. The analysis reveals a decrease in latency as the number of deployed vehicles increases. This counterintuitive trend is enabled by the system's optimized cloud aggregation, batch processing, and efficient data routing mechanisms. At 10 vehicles, latency is highest due to limited parallelization, but it decreases significantly as deployment scales to 1,000 and 10,000 vehicles. The results demonstrate that the proposed framework not only remains scalable but becomes more computationally efficient when managing larger fleets, making it ideal for industry-wide adoption in logistics, public transportation, and commercial vehicle operations.

Figure 7: Low-cost scalable deployment architecture

The overall architecture of the proposed low-cost and scalable AI-based predictive maintenance framework as shown in figure 7. The pipeline begins with onboard sensors that continuously collect real-time vehicle data, followed by processing through an AI model deployed at the edge level using microcontrollers or single-board computers. Only critical alerts and summarized diagnostic information are transmitted to the cloud, significantly reducing communication costs and bandwidth requirements. The cloud aggregation layer enables centralized analytics, fleet-level

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

monitoring, and long-term maintenance planning. The architecture illustrates a well-balanced distribution of computation between the edge and cloud, demonstrating how the system achieves both scalability and cost-efficiency. This modular design ensures compatibility with diverse vehicle types and supports expansion from small pilot deployments to large-scale fleet operations.

Overall, the results confirm that the proposed framework successfully balances predictive accuracy, computational efficiency, and cost-effectiveness, making it highly suitable for real-world fleet operations and industry-wide deployment. The reductions in latency, hardware costs, and maintenance expenses—combined with strong diagnostic performance across diverse fault categories—demonstrate that the system can enhance vehicle safety and reduce unexpected breakdowns. Moreover, the framework's scalability across thousands of vehicles highlights its potential to transform maintenance strategies within the automotive sector. These outcomes underscore the practical significance of the research and establish a solid foundation for future extensions involving multimodal sensing, adaptive model retraining, and large-scale field validation.

Conclusion

This study presents a low-cost, scalable AI framework designed to improve predictive vehicle maintenance and enhance overall safety across diverse vehicular environments. By integrating lightweight machine learning models, edge-based inference, and cloud-assisted analytics, the proposed system successfully addresses the major challenges associated with high deployment costs, limited onboard computational resources, and inconsistent sensor data quality. The experimental evaluation demonstrates that the framework effectively identifies early-stage mechanical faults, predicts component failures with high accuracy, and offers real-time alerts for critical safety risks—all while operating efficiently on resource-constrained hardware. The modular architecture ensures seamless scalability, making the framework adaptable to low-end vehicles, commercial fleets, and large-scale intelligent transportation systems. The cost analysis further confirms that the solution provides a financially viable alternative to existing proprietary telematics and maintenance systems, without compromising performance or reliability.

References

1. Mahale, Y., Kolhar, S. & More, A.S. A comprehensive review on artificial intelligence driven predictive maintenance in vehicles: technologies, challenges and future research directions. *Discov Appl Sci* 7, 243 (2025). https://doi.org/10.1007/s42452-025-06681-3
2. Ucar, A., Karakose, M., & Kırımça, N. (2024). Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends. *Applied Sciences*, *14*(2), 898. https://doi.org/10.3390/app14020898
3. Mousaei A, Naderi Y, Bayram IS. Advancing state of charge management in electric vehicles with machine learning: a technological review. IEEE Access. 2024;12:43255–83.
4. Arena, F.; Collotta, M.; Luca, L.; Ruggieri, M.; Termine, F.G. Predictive maintenance in the automotive sector: A literature review. *Math. Comput. Appl.* 2021, *27*, 2.
5. Khatri, M.R. Integration of natural language processing, self-service platforms, predictive maintenance, and prescriptive analytics for cost reduction, personalization, and real-time insights customer service and operational efficiency. *Int. J. Inf. Cybersecur.* 2023, *7*, 1–30.
6. Rasheed R, Qazi F, Ahmed A, Asif A, Shams H, et al. Machine learning approaches for in-vehicle failure prognosis in automobiles: a review. VFAST Trans Softw Eng. 2024;12(1):169–82.

*Correspondence to: Chandani S.Bartakke **Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com

Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 148-158

7. Campos-Ferreira AE, Lozoya-Santos JdJ, Tudon-Martinez JC, Mendoza RAR, Vargas-Martínez A, Morales-Menendez R, et al. Vehicle and driver monitoring system using on-board and remote sensors. Sensors. 2023;23(2):814.
8. Morales Matamoros O, Takeo Nava JG, Moreno Escobar JJ, Ceballos Chávez BA. Artificial Intelligence for Quality Defects in the Automotive Industry: A Systemic Review. Sensors (Basel). 2025 Feb 20;25(5):1288. doi: 10.3390/s25051288. PMID: 40096013; PMCID: PMC11902312.
9. Rao PS, Yaqoob SI, Ahmed MA, Abdinabievna PS, Yaseen SM, Arumugam M. Integrated artificial intelligence and predictive maintenance of electric vehicle components with optical and quantum enhancements. Opt Quant Electron. 2023;55(10):855.
10. Hu P, Chu X, Lv L, Zuo K, Ni T, Wang T, et al. An efficient and secure data collection scheme for predictive maintenance of vehicles. Ad Hoc Netw. 2023;146: 103157.
11. Trivedi M, Kakkar R, Gupta R, Agrawal S, Tanwar S, Niculescu VC, et al. Blockchain and deep learning-based fault detection framework for electric vehicles. Mathematics. 2022;10(19):3626. doi:10.3390/math10193626.

Correspondence to: Chandani S.Bartakke ***Department of Computer Science and Engineering, Suresh Gyan Vihar University Jaipur***
Corresponding author. E-mail addresses: chandanibartakke@gmail.com