

Article

Not peer-reviewed version

Framework for Next-Generation Predictive Maintenance of Software-Defined Vehicles Using Cloud, Edge Computing and Modern AI

[Kamal Pandey](#)*

Posted Date: 28 July 2025

doi: 10.20944/preprints202507.2205.v1

Keywords: predictive maintenance; automotive; cloud computing; machine learning; explainable AI; generative AI; digital twin; predictive maintenance (PdM)



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Framework for Next-Generation Predictive Maintenance of Software-Defined Vehicles Using Cloud, Edge Computing and Modern AI

Kamal Pandey

IT Corp Functions, Strat & Architecture, Rivian utomotive Inc (of Aff.), Irvine and United States of America;
kamalkismca@gmail.com

Abstract

The evolution of the automotive industry toward connected software-defined vehicles demands a shift from reactive to proactive, data-driven maintenance strategies. Predictive Maintenance (PdM) addresses this need but faces challenges in terms of data quality, model accuracy, and stakeholder trust. This study presents a next-generation, cloud-native PdM framework designed to overcome these barriers. It introduces an end-to-end reference architecture that integrates secure data ingestion, unified analytics, and scalable ML Ops to manage the entire AI lifecycle. We benchmark classical machine learning models (Random Forest, XGBoost) and deep learning approaches (LSTM and transformer) for key automotive use cases, including high-accuracy fault classification and Remaining Useful Life (RUL) estimation, using industry-standard datasets. To enhance transparency, Explainable AI (XAI) techniques, such as SHAP, are applied to convert complex model outputs into technician-friendly insights. The framework further leverages generative AI for synthetic data augmentation, automated report generation, and LLM-powered virtual assistants to support diagnostics and decision-making. The novel contributions of this study include the integration of federated learning with differential privacy, component-level digital twins, and an adaptive orchestration layer for over-the-air (OTA) model updates in zonal ECU architectures. This study provides a technical blueprint for building a scalable, intelligent, and privacy-preserving PdM solution that improves vehicle reliability, reduces unplanned downtime, and aligns with modern enterprise strategies in the electric vehicle sector.

Keywords: predictive maintenance; automotive; cloud computing; machine learning; explainable ai; generative ai; digital twin; predictive maintenance (PdM)

1. Introduction

The automotive industry is undergoing a profound transformation that renders traditional reactive and preventive maintenance models economically unfeasible. These legacy approaches result in excessive downtime, unnecessary costs, and an inability to preemptively address catastrophic failures. Predictive Maintenance (**PdM**) is a proactive strategy that leverages the vast data generated by modern vehicles to forecast component degradation and estimate the Remaining Useful Life (RUL). Studies have shown that PdM can reduce maintenance costs by **25–35%** and eliminate up to **75%** of unexpected breakdowns.

To achieve this, a scalable, cloud-native platform capable of managing the entire AI lifecycle is required. A modern cloud ecosystem provides powerful data analytics and machine learning services that are ideal for this purpose. This study presents a technical framework for designing and implementing a next-generation PdM solution. We detail a reference architecture, benchmark predictive models, and integrate both Explainable AI (XAI, ChatGPT) for transparency and generative AI for automation and decision support.

A successful PdM system is built on a scalable architecture that seamlessly manages data from the vehicle to the cloud and translates it into actionable intelligence. The overall architecture of the proposed Predictive Maintenance System is shown in Figure 1. This system is designed to manage the data flow from data acquisition at the edge to actionable insights at the edge. Our proposed reference architecture orchestrates key cloud services within an end-to-end framework.

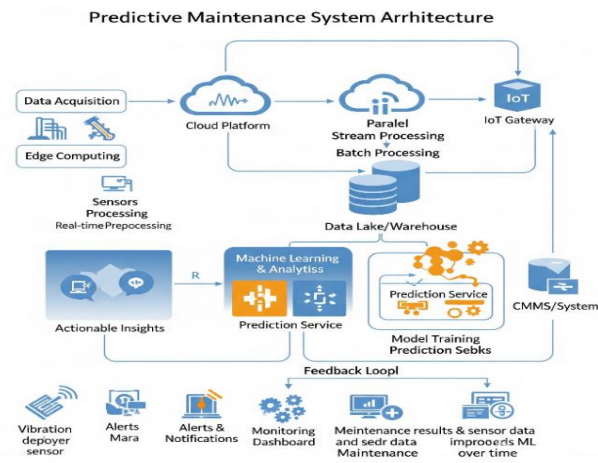


Figure 1. Example of PdM System.

The data journey begins with secure ingestion from globally dispersed vehicles using services for device management and authentication and a scalable real-time messaging buffer. [1–5] Data are then processed by a serverless service executing unified transformation pipelines for both historical batch data and live streams, a critical practice to prevent training-serving skew. [6]

At the heart of the framework lies a *scalable, cloud-native data warehouse* that serves as a unified analytics hub. [8] All ML-ready data are consolidated here, where embedded ML capabilities allow for rapid model prototyping directly within the data warehouse using SQL. 17 For high-throughput raw time-series data, a NoSQL wide-column database can serve as a complementary storage layer. 12 A critical aspect of modern manufacturing is the unification of Operational Technology (OT) and Information Technology (IT) data, which can be achieved through solutions such as a Manufacturing Data Engine with a Cortex Framework, creating a holistic view of factory and business operations.

For enterprise-grade AI, a unified ML Ops platform orchestrates the entire machine learning lifecycle. 19 It provides managed services for custom training, experiment tracking, a centralized model registry for governance, and scalable endpoints for both batch and online predictions. [7]

This framework is designed for hybrid intelligence, combining cloud power with edge responsiveness. Using edge computing software and hardware accelerators, models can be deployed on vehicles for real-time inference, enabling immediate safety alerts, ensuring operation with intermittent connectivity, and enhancing data privacy by processing sensitive information locally. 21 The cloud remains essential for computationally intensive tasks such as training large-scale models on fleet-wide data. 22 This creates a cyclical feedback loop in which predictions and technician feedback are captured to continuously retrain and improve the models over time. [9–11]

1.1. Background and Problem Statement

The paradigm shift toward software-defined vehicles (SDVs) has dramatically increased the complexity of automotive systems. While enabling unprecedented features and connectivity, this evolution also exposes the limitations of the traditional maintenance strategies. Reactive maintenance, or "fix-it-when-it-breaks," leads to unplanned downtime and high operational costs, while scheduled preventive maintenance often results in the unnecessary replacement of functional components.

Predictive Maintenance (PdM) offers a solution, but existing frameworks face significant gaps that hinder their enterprise-wide adoption. This study proposes a framework that directly addresses these gaps.

- Scalability and Data Management: Legacy PdM systems struggle to ingest, process, and analyze terabytes of high-velocity data generated by a global fleet of connected vehicles.
- Lack of Explainability: Many advanced machine learning models operate as "black boxes," making it difficult for technicians to trust their predictions without understanding the underlying reasoning. This trust deficit is a major barrier to the adoption of AI.
- Data Scarcity for Failures: Critical component failures are rare events by design, leading to highly imbalanced datasets that complicate the training of exact failure-prediction models.
- Privacy and Security: Transferring raw vehicle and user data to a central cloud raises significant data privacy and sovereignty concerns, requiring robust, privacy-preserving techniques

1.2. The Cloud-Native Predictive Maintenance Framework

As illustrated in Figure 2, the updated cloud-native predictive maintenance (PdM) framework integrates federated learning with differential privacy, component-level digital twins, embedded generative AI diagnostic agents, and an Adaptive PdM Orchestrator layer. This orchestrator dynamically coordinates over-the-air (OTA) model updates across the zonal ECU architecture, aligning with Rivian’s next-generation platform strategy.

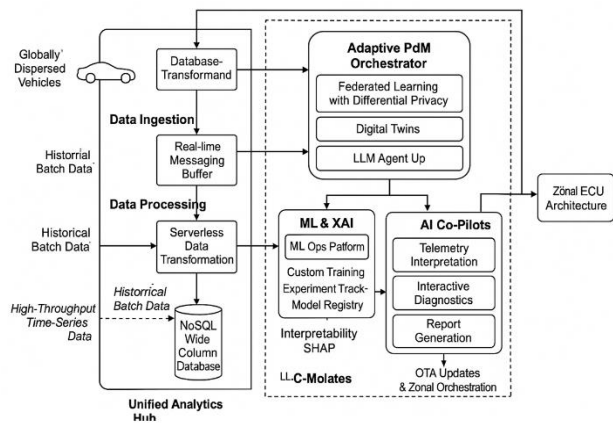


Figure 2. Cloud-native PdM architecture integrating federated learning, digital twins, LLM agents, and adaptive zonal orchestration. Cloud-native predictive maintenance (PdM) architecture integrating federated learning, digital twins, LLM diagnostic agents, and adaptive OTA orchestration for zonal ECU deployment.

A successful PdM system requires a scalable architecture to manage the end-to-end data pipeline from the vehicle to the cloud and translate raw data into actionable intelligence. Our proposed reference architecture orchestrates key cloud services within a cohesive framework.

The architecture begins with secure data ingestion from globally dispersed vehicles using services for device management, authentication, and scalable real-time messaging buffers. Data are then processed by serverless services executing unified transformation pipelines for historical batch data and live streams, which is a critical practice to prevent training-serving skew.

The core of the framework is a petabyte-scale, serverless data warehouse that serves as a unified analytical hub. All ML-ready data are consolidated here, enabling rapid model prototyping directly within the data warehouse. For high-throughput time-series data, a NoSQL wide-column database complements the layer 13. A modern approach unifies operational technology (OT) and information technology (IT) data, creating a holistic view of factory and vehicle operations. 8 For enterprise-grade

AI, a unified ML Ops platform orchestrates the entire machine learning lifecycle by providing managed services for custom training, experiment tracking, centralized model registry, and scalable endpoints for batch and online predictions. This hybrid architecture combines the cloud scale with edge responsiveness. Using edge computing and hardware accelerators, models can be deployed on vehicles for real-time inference, enabling immediate alerts and local privacy-preserving processing. The cloud handles computationally intensive training on fleet-wide data, creating a continuous feedback loop for model retraining and improvement.

2. Methodology: An End-to-End Workflow

The proposed framework transforms raw vehicle data into predictive intelligence using a sequence of integrated stages.

Secure Data Ingestion and Processing The process begins with the secure data ingestion from a globally dispersed fleet. Data is transmitted using lightweight IoT protocols like

2.1. Secure Data Ingestion and Processing

The process begins with the secure data ingestion from a globally dispersed fleet. Data is transmitted using lightweight IoT protocols like MQTT (Message Queuing Telemetry Transport), which are ideal for environments with intermittent connectivity. This stage leverages cloud services for robust device management, mutual TLS (mTLS) authentication for security, and a real-time messaging buffer to handle high-volume data streams without losses. Once ingested, the data undergoes transformation in a unified analytics pipeline that processes both historical batch data and live streams to prevent training-serving skew.

2.2. Unified Analytics Hub

The core of the framework is a petabyte-scale, serverless data warehouse that serves as a unified analytics hub. Multiple forms of analytics are performed in concert.

- **Descriptive Analytics:** To visualize historical performance and operational trends.
- **Diagnostic Analytics:** To perform root cause analysis on past anomalies and failures.
- **Predictive Analytics:** To forecast future failures and estimate the RUL using ML models.
- **Prescriptive Analytics:** To recommend specific maintenance actions and inventory planning based on predictive outputs.

For high-throughput time-series data, a NoSQL wide-column database can serve as a complementary storage layer

2.3. MLOps and Hybrid Intelligence

An enterprise-grade MLOps platform orchestrates the entire machine learning lifecycle, providing managed services for experiment tracking, model versioning, and governance in a centralized registry, and scalable endpoints for both batch and online prediction. The architecture was designed for hybrid intelligence, combining cloud power with edge responsiveness. Optimized models are deployed on vehicles using edge computing accelerators for real-time inference, enabling immediate safety alerts while preserving privacy by processing sensitive data locally. This creates a continuous feedback loop in which predictions and technician feedback from the field are used to retrain and improve the models over time.

3. Benchmarking Predictive Models

The selection of the appropriate ML model depends on the specific task, data availability, and interpretability requirements.

3.1. Feature Engineering for Automotive Data

The performance of any predictive model is highly dependent on the quality of its input feature. For automotive PdM, features are engineered from various sources, including

- Time-series sensor data: Readings such as battery temperature, voltage, coolant pressure, motor RPM, and vibration sensor data are collected.
- The vehicle operational parameters included vehicle speed, acceleration/deceleration patterns, charging cycles, and ambient temperature.
- Historical records: Diagnostic Trouble Codes (DTCs), past repair orders, and component replacement history.

These raw data streams are processed to create meaningful features, such as rolling averages, standard deviations, and frequency-domain features via Fourier transforms, which are then fed into the models

3.2. Performance on Standardized Datasets

The model performance was validated using public benchmarks. The

The NASA C-MAPSS dataset, a well-known benchmark for prognostics, consists of simulated run-to-failure turbofan engine data, making it ideal for validating RUL estimation models such as LSTMs and Transformers. In contrast, the

The Scania truck dataset focuses on classifying air pressure system failures from real-world operational data and is characterized by a severe class imbalance, making it suitable for evaluating cost-sensitive classifiers such as Random Forest and XGBoost.

3.3. Evaluating Model Performance

Although accuracy is a primary metric for classification, a comprehensive evaluation requires a broader set of metrics. For imbalanced fault classification tasks, we also consider

Precision, recall, and F1-score were used to ensure that the model effectively identified rare failure events. For RUL estimation, the root mean squared error (RMSE) and mean absolute error (MAE) were used to quantify the accuracy of time-to-failure predictions. Furthermore, inference time and computational cost are critical metrics, especially for models intended for on-vehicle deployment on edge devices

3.4. Foundational Models for Fault Classification

Tree-based ensemble models, such as Random Forest (**RF**) and XGBoost, are powerful for classifying component failures. They excel at capturing nonlinear relationships in structured data. Studies show that RF achieves ~98% accuracy in EV fault diagnosis and XGBoost achieves ~96% on bearing vibration data, establishing strong baselines. [24,25]

3.5. Advanced Models for RUL Estimation

Tree-based ensemble models, such as **Random Forest (RF)** and **eXtreme Gradient Boosting (XGBoost)**, are powerful and effective for classifying component failures. They excel at capturing nonlinear interactions in structured data and consistently achieve high accuracy. Studies show that RF achieves **98.18%** accuracy in EV fault diagnosis, outperforming other classifiers, and XG Boost achieves **96.61%** accuracy with high efficiency on bearing vibration data.¹² They are ideal for proving strong baselines.

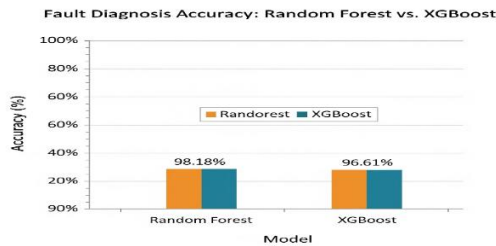


Figure 3. Comparative performance of ML models on automotive PdM datasets.

3.6. Performance on Standardized Datasets

The performance was validated using public benchmarks. On the **NASA C-MAPSS dataset for RUL estimation**, deep learning approaches such as **LSTMs** consistently outperform classical models. In contrast, on the Scania truck dataset, which focuses on cost-sensitive fault classification with severe class imbalance, tree-based ensembles such as RF, combined with effective imbalance handling, achieve excellent results. A new Scania dataset provides a richer resource for a broader range of PdM tasks. This highlights that a mature PdM framework requires a portfolio of models tailored to specific business objectives.^{24, 26}

Table 1. Comparative Analysis of Predictive Models.

Model	Models and Use Cases		
	PdM Task	Dataset	Result
Random Forest (RF)	Fault Classification	EV/Scania	Accuracy ~98%/High Cost-Efficiency
XGBoost	Fault Classification	Bearings/Scania	Accuracy ~96%/High Cost-Efficiency ²²
LSTM	RUL Estimation	C-MAPSS	High RUL Score (outperforms classical ML)
CNN-LSTM	RUL Estimation	C-MAPSS	Higher RUL Score (outperforms LSTM): ²⁵
Transformer	RUL Estimation	C-MAPSS	State-of-the-Art Forecasting Accuracy ²⁶

Comparative Analysis of Predictive Models Tasks, Data, and Results.

4. Building Trust with Explainable ai (xai)

High model accuracy is meaningless if the outputs are not trusted by technicians. **XAI** tackles the "black box" problem by providing human-understandable explanations for predictions, which is critical for debugging, compliance, and stakeholder trust. [14–16]

Modern cloud tools enable XAI using methods such as **Sampled Shapley (SHAP)** and **Integrated Gradients**²¹ for deep neural networks. These produce feature attributions that indicate how each input affects the prediction. However, numerical explanations alone are insufficient. The

framework must include a translation layer that converts attributions into clear and actionable guidance, transforming a cryptic model output into a credible diagnostic starting point.

Our framework utilizes Shapley Additive exPlanations (SHAP) to generate feature attributions for each prediction. In practice, SHAP calculates the contribution of each input feature to the model output. For instance, a high positive SHAP value for 'Battery Temperature' in a fault prediction indicates that it is a primary driver of the predicted failure. However, raw numerical values are insufficient for the technicians. The framework includes a translation layer that converts these attributions into actionable guidelines. For example, the output

[BatteryTemperature=+0.32, ChargeCycles=+0.27] is translated into the natural language recommendation: 'The model predicts a high risk of failure primarily due to high battery temperatures and a high number of charge cycles. Please inspect the battery cooling system and review the vehicle's recent charging history,' as demonstrated in the LLM prompt template

5. Leveraging Generative AI for Enhanced Maintenance

Generative AI addresses key data challenges and reshapes human-computer interaction in maintenance.

5.1. Synthetic Data Generation

To address the scarcity of failure data, we use

Generative Adversarial Networks (GANs), in which a generator network creates synthetic sensor data and a discriminator network attempts to distinguish it from real data, iteratively improve data fidelity. For AI co-pilots, we propose a

The Retrieval-Augmented Generation (RAG) architecture allows the LLM to retrieve information from a knowledge base of technical manuals, repair logs, and real-time data before generating a response. This grounds the model in factual information, enabling it to accurately respond to complex diagnostic queries.

Despite its potential, the use of generative AI in safety-critical domains requires caution. Key challenges include mitigating the risk of model 'hallucination' (generating incorrect or fabricated information), ensuring the privacy of data used in LLM training, and validating the accuracy of synthetic data. Robust validation, human-in-the-loop oversight, and grounding techniques, such as RAG, are essential to ensure the reliability and safety of these systems.

5.2. AI Co-Pilots and Automated Reporting

Large Language Models (LLMs) Large Language Models (LLMs) can power sophisticated virtual assistants for technicians. These AI co-pilots can engage in natural language diagnostics, interpret fault codes, and provide interactive, step-by-step repair guidance based on technical manuals and historical data. 56 Automakers, such as Volkswagen and Mercedes-Benz, have already implemented such assistants. LLMs can also automate the generation of human-readable maintenance reports by synthesizing the outputs of predictive and XAI models. Effective prompt engineering, which structures queries with a clear context, tasks, and constraints, is key to unlocking actionable and domain-specific responses. [16]

AI assistants (e.g., ChatGPT, Gemini, Grok Ai) enable AI contextual prompt engineering to interpret fault codes, guide repairs, and auto-generate clear maintenance reports from [16–19]

Table II summarizes the key applications of various generative AI models in the context of automotive maintenance

Table 2. Applications of Generative AI in Automotive Maintenance.

Model	Generative Ai Applications in Automotive Maintenance	
	Capability	Use Case
LLMs	Natural language generation	Virtual Assistants, Reports
Generative Adversarial Networks (GANs)	Synthetic data	Rare failure event augmentation
Variational Autoencoders (VAEs)	Unsupervised anomaly detection	Data generation, representation learning
Large Language Models (LLMs)	Natural language generation	Virtual Assistants, Reports

Generative AI Models and Use Cases in Automotive Maintenance.

6. Enhancements and Future Directions

To extend the practical impact and novelty of the proposed cloud-native predictive maintenance (PdM) framework for software-defined electric vehicles, this section outlines concrete enhancements grounded in recent industry trends, Rivian’s platform strategy, and cross-OEM scalability requirements. These enhancements aim to advance the framework into a next-generation, enterprise-ready solution.

6.1. Federated Learning with Differential Privacy

To address privacy and data sovereignty concerns, we augmented the framework with Federated Learning (FL). This approach enables model training across leading manufacturing company fleets without transferring raw vehicle data to the cloud. Differential Privacy (DP) can be layered to inject calibrated noise, making it mathematically improbable to re-identify individual data points, thereby strengthening compliance with emerging automotive data regulations.

6.2. Component-Level Digital Twins

The integration of digital twins of key EV components supports high-fidelity "what-if" scenario testing. For example, a battery pack twin can simulate degradation under extreme weather conditions or driving patterns. These twins are continuously synchronized with live sensor data, allowing for a highly accurate RUL estimation.

6.3. Embedded Generative AI Agents

We propose deploying on-vehicle and cloud-based Large Language Model (LLM) agents as diagnostic co-pilots. These agents will:

- Interpret telemetry and XAI outputs in real time.
- Generate human-readable maintenance reports,
- Provide step-by-step fault resolution guidance to technicians.
- Learn iteratively from technician feedback to improve over time.

6.4. Adaptive OTA & Zonal Orchestration

An "Adaptive PdM Orchestrator" is a cloud service that monitors the on-vehicle model performance. It dynamically triggers OTA updates based on model drift, new data availability, or the release of an improved model architecture, ensuring the fleet benefits from the latest intelligence

6.5. Enterprise KPIs and Fleet Metrics

Enhancing the framework with enterprise-aligned KPIs enables quantifiable business values. Example metrics include the following:

- Reduction in unexpected breakdowns,
- Decrease in technician dispatches,
- Average latency from anomaly detection to mitigation,
- OTA deployment success rate,
- Improvement in the RUL prediction accuracy.

6.6. Cross-OEM Scalability and JV Alignment

To support the Rivian–Volkswagen joint venture and future cross-OEM deployments, we propose designing the PdM layer to be modular and OEM-agnostic. This enables faster adaptation to new vehicle platforms, reducing the time-to-market and total cost of ownership (TCO).

6.7. XAI Translation with Natural Language Layer

To expand explainable AI capabilities, we recommend adding a translation layer powered by LLMs. This layer converts technical feature attributions (e.g., SHAP values) into technician-friendly explanations and recommended actions, thereby enhancing usability and stakeholder trust.

6.8. Updated Framework and Pseudo-Code

The updated architecture diagram (Figure 2) illustrates the integration of FL, digital twins, LLM agents, and the Adaptive PdM Orchestrator. Algorithm 1 outlines the high-level federated PdM training pipeline with privacy preservation.

6.9. Federated Learning with Differential Privacy:

The 'calibrated noise' injected to achieve Differential Privacy is typically applied using the

6.10. Laplace or Gaussian Mechanisms

The amount of noise is controlled by a privacy budget, epsilon (ϵ), which represents the trade-off between privacy and model accuracy. A smaller epsilon provides stronger privacy guarantees but may lead to a less accurate global model. Determining the optimal epsilon is a critical tuning step for balancing regulatory compliance and model utility."

6.11. Component-Level Digital Twin

These digital twins enable 'what-if' scenario testing. For example, a digital twin of a battery pack can simulate its degradation under extreme weather conditions or aggressive driving pattern scenarios that are difficult and expensive to replicate. These twins are continuously synchronized with live sensor data from their physical counterparts using real-time data streams, allowing for a dynamic and highly accurate estimation of the RUL.

6.12. Adaptive OTA and Zonal Orchestration

The 'Adaptive PdM Orchestrator' is a cloud-based service that checks the performance of on-vehicle models. It dynamically triggers OTA updates based on several factors, including model **drift**

(performance degradation over time), the availability of significant new training data from the fleet, and the release of an improved model architecture. This ensures that the entire fleet benefits from the latest intelligence in a seamless and controlled way.

Algorithm 1: Federated PdM Training with Differential Privacy

Input: Local datasets $\{D_1, D_2, \dots, D_N\}$ on N vehicles

Output: Global PdM model θ

```
Initialize global model  $\theta_0$ 
for each communication round  $t = 1$  to  $T$  do
    for each vehicle  $i \in \{1, \dots, N\}$  in parallel do
         $\theta_i \leftarrow \text{LocalTrain}(\theta_{t-1}, D_i)$ 
         $\theta_i \leftarrow \text{AddDPNoise}(\theta_i, \epsilon)$ 
    end for
     $\theta_t \leftarrow \text{Aggregate}(\{\theta_i\})$ 
end for
return  $\theta_T$ 
```

The end-to-end federated PdM training workflow with differential privacy is illustrated in Figure 4. Each vehicle trains its local model on private data, applies calibrated DP noise, and the server securely aggregates the updates to produce an improved global model. This process is repeated across multiple communication rounds, enabling fleet-wide learning without sharing raw data. 26-27

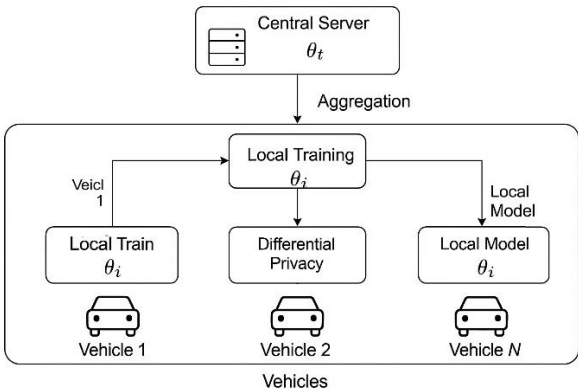


Figure 4. Federated predictive maintenance (PdM) workflow: Each vehicle trains a local model, adds calibrated differential privacy noise, and securely aggregates updates to build a global model.

As shown in Figure 2 and Figure 5, the high-level cloud-native PdM architecture is shown, and Fig. 5 focuses specifically on the federated training workflow with differential privacy. "21"

6.13. Example Prompt Template for LLM Diagnostic Co-Pilot

"You are an AI diagnostic assistant. Given the following fault code and SHAP explanation:
Fault Code: P0A1C
Key Features Impact: [BatteryTemperature=+0.32, ChargeCycles=+0.27, AmbientTemp=-0.10]
Generate a plain-language diagnostic summary and recommend next inspection steps for a technician."

By integrating these enhancements, the framework evolves into a scalable, intelligent, and privacy-aware PdM system tailored for next-generation EV platforms and enterprise deployments.

7. Advanced Paradigms and Future Frontiers

7.1. Federated and Transfer Learning

Federated Learning (FL) is a distributed approach that enables collaborative model training on decentralized data without the data ever leaving its source, preserving privacy and data sovereignty. To provide stronger guarantees, [20]

Differential Privacy can be layered on top, injecting calibrated noise to make it mathematically impossible to re-find any single data point. [64]

Transfer learning is also crucial for adapting pretrained models to new vehicle types with limited data, significantly reducing training time and data requirements. [21]

7.2. Automotive Digital Twins

The convergence of these technologies results in the Automotive Digital Twin dynamic, a virtual replica of a physical asset that is continuously refreshed with real-time operational data. Digital twins support high-fidelity simulations for testing innovative components, advancing predictive maintenance techniques [22,23] and conducting cyber-physical resilience assessments in a secure virtual environment. This capability transforms predictive maintenance from reactive and preventive approaches to a fully data-driven, adaptive practice.

8. Conclusion

Transitioning to a data-driven predictive maintenance framework is an essential strategy for modern automotive manufacturers. This study presents a robust, cloud-native framework that integrates end-to-end data management with advanced AI. We benchmarked the foundational and advanced models for fault classification and RUL estimation. Critically, this work addresses key industry challenges through several novel contributions: the integration of

federated learning with differential privacy for data sovereignty; the use of component-level digital twins for high-fidelity simulations; and an **adaptive orchestration layer** for dynamic, over-the-air model updates. By incorporating explainable AI for trust and generative AI for automation, this framework provides a comprehensive implementation blueprint. Successful adoption requires a cultural shift toward data-driven decision-making, allowing organizations to transform maintenance from a cost center into a proactive driver of reliability, safety, and operational excellence.

Acknowledgment: The author, Dr. Kamal Pandey, would like to thank Dr. Ben Bellash, Ph.D., IEEE Fellow, and Mr. Anoop Narang, Senior Director, IT Corporate Functions, Strategy & Architecture, Rivian Automotive Inc., United States of America, for their valuable time, constructive feedback, and expert review, which have significantly strengthened the quality and clarity of this paper.

References

1. Y. Hu, "Explainable Program Synthesis by Localizing Specifications," *OOPSLA*, 2023. <https://www.yifei-h.com/files/OOPSLA23.pdf>. Accessed: Jun. 24, 2025.
2. F. Zhai, et al., "Towards Explainable Decision Making with Neural Program Synthesis and Library Learning," *CEUR Workshop Proceedings*, vol. 3432, 2023. <https://ceur-ws.org/Vol-3432/paper30.pdf>. Accessed: Jun. 24, 2025.
3. C. Cai et al., "Learning to Synthesize Programs as Interpretable and Generalizable Policies," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. <https://papers.nips.cc/paper/2021/file/d37124c4c79f357cb02c655671a432fa-Paper.pdf>.
4. L. Chen et al., "Modular Machine Learning: An Indispensable Path towards New-Generation Large Language Models," *arXiv preprint*, arXiv:2504.20020v1, 2025. <https://arxiv.org/html/2504.20020v1>.
5. X. Sun et al., "Neuro-symbolic approaches in artificial intelligence," *National Science Review*, vol. 9, no. 6, nwac035, 2022. [Online]. Available: <https://academic.oup.com/nsr/article/9/6/nwac035/6542460>.

6. H. Jiang et al., "Large Language Models Are Neurosymbolic Reasoners," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 13, pp. 15644–15651, 2024. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/29754>.
7. Pandey, Kamal. (2025). Designing Ethical AI for Development: Challenges and Opportunities in Humanitarian Engineering and Electric Vehicles, Designed for Adventure. 1. 1-8.
8. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), <https://doi.org/10.1609/aaai.v34i07.6999>
9. Pandey, Kamal. (2025). AI-Powered Low-Code App Development: A Benchmark Review for 2025 Digital Transformation. 10.13140/RG.2.2.35587.62244/1.
10. Pandey, Kamal & Narang, Anoop & Journal, Ijeasm. (2025). AI-Powered Transformation in the Modern Digital Workplace: A Roadmap for the Future of Work and Ethical Consideration. 6. 2582-6948. 10.13140/RG.2.2.30934.36161.
11. Pandey, Kamal. (2025). The Intelligent Workplace: AI and Automation Shaping the Future of Digital Workplaces. *International Journal of Scientific Research in Computer Science and Engineering*. 13. 10.26438/ijsrcse.v13i1.607.
12. L. Ma et al., "Unlocking the Potential of Generative AI through Neuro-Symbolic Architectures – Benefits and Limitations," *arXiv preprint*, arXiv:2502.11269v1, 2025. [Online]. Available: <https://arxiv.org/html/2502.11269v1>.
13. X. Zhou and Y. Xie, "Recent Developments in Causal Inference and Machine Learning," *Scholars at Harvard*, 2023. [Online]. Available: https://scholar.harvard.edu/files/xzhou/files/brand-zhou-xie2023_causal.pdf. Accessed: Jun. 24, 2025.
14. Y. Wang et al., "Causal Inference Meets Deep Learning: A Comprehensive Survey," *PubMed*, 2024. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/39257419/>. Accessed: Jun. 24, 2025.
15. B. Zhang et al., "Causality, Machine Learning, and Feature Selection: A Survey," *Sensors*, vol. 25, no. 8, p. 2373, 2025. [Online]. Available: <https://www.mdpi.com/1424-8220/25/8/2373>. Accessed: Jun. 24, 2025.
16. K. Chen et al., "A Survey on Causal Inference," *arXiv preprint*, arXiv:2002.02770, 2020. [Online]. Available: <https://arxiv.org/abs/2002.02770>. Accessed: Jun. 24, 2025.
17. J. Lin et al., "Interpretable Preferences via Multi-Objective Reward Modeling and Mixture-of-Experts," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 7882–7897. [Online]. Available: <https://aclanthology.org/2024.findings-emnlp.620.pdf>.
18. Pandey, Kamal. (2025). Abstract Cloud Computing: High-Performance Algorithms and AI Innovations for High-Resolution Image and File Processing.
19. Pandey, Kamal. (2024). Artificial Intelligence (AI) in Electric Vehicle Ecosystems: Challenges, Opportunities, and Models for Accelerated Adoption. *International Journal of Applied Science and Engineering Review*. 5. 1-19. 10.13140/RG.2.2.32810.32962.
20. J. Lin et al., "Interpretable Preferences via Multi-Objective Reward Modeling and Mixture-of-Experts," *arXiv preprint*, arXiv:2406.12845v1, 2024. [Online]. Available: <https://arxiv.org/html/2406.12845v1>. Accessed: Jun. 24, 2025.
21. A. Santoro et al., "A Simple Neural Network Module for Relational Reasoning," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: <http://papers.neurips.cc/paper/7082-a-simple-neural-network-module-for-relational-reasoning.pdf>. Accessed: Jun. 24, 2025.
22. Singh, Trilok & Pandey, Kamal. (2024). The Future of Automotive Industry: AI and Cloud-Driven Digital Transformation in the US Region -A Holistic Examination. 10.13140/RG.2.2.31644.19846.
23. K. Pandey, "The Intelligent Workplace: AI and Automation Shaping the Future of Digital Workplaces ", *Int. J. Sci. Res. Comp. Sci. Eng.*, vol. 13, no. 1, pp. 1–10, Feb. 2025.
24. Pandey, K. 2025 "Designing Ethical AI for Development: Challenges and Opportunities in Humanitarian Engineering and Electric Vehicles, Designed for Adventure" Preprints. <https://doi.org/10.20944/preprints202507.0141.v1>

25. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12),. <https://doi.org/10.1609/aaai.v35i12.17325>
26. Besta, M., Blach, N., Kubicek, A., Gerstenberger, R., Podstawski, M., Gianinazzi, L., Gajda, J., Lehmann, T., Niewiadomski, H., Nyczyk, P., & Hoefler, T. (2024). Graph of Thoughts: Solving Elaborate Problems with Large Language Models. Proceedings of the AAAI Conference on Artificial Intelligence.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.