Machine Learning-Based Predictive Maintenance for Vehicle Health Monitoring : Integrating Random Forest and LSTM Models

Drupath S Prasad
Scholar
Department Of Computer Science
Sacred Heart College Thevara,Kochi,India
22umcp8119@shcollege.ac.in

Alia Teresa T.M
Assistant Professor
Department Of Computer Science
Sacred Heart College Thevara,Kochi,India
aliateresa@shcollege.ac.in

ABSTRACT

In the automotive sector, predictive maintenance (PdM) has become a vital strategy for minimizing unscheduled breakdowns and streamlining maintenance plans. Conventional maintenance techniques can result in needless expenses and lost vehicle time. In order to forecast the Remaining Useful Life (RUL) of vehicle components and identify irregularities in sensor data, this research makes use of machine learning (ML) models, such as Random Forest classifiers and Long Short-Term Memory (LSTM) networks. Users can enter car parameters to obtain real-time health scores and maintenance recommendations by utilizing a web application that is based on Streamlit. This system improves efficiency and dependability by offering a proactive approach to auto maintenance.

Keywords : Predictive Maintenance, Machine Learning, Automotive Industry, Vehicle Maintenance, Artificial Intelligence

1. INTRODUCTION

The automotive industry faces significant challenges in ensuring vehicle reliability while minimizing maintenance costs. Traditional reactive and preventive maintenance strategies often lead to inefficiencies, increased costs, and unexpected breakdowns. Reactive maintenance involves addressing vehicle failures only after they occur, which can result in expensive repairs and safety hazards. Preventive maintenance, on the other hand, follows a scheduled approach that may result in unnecessary servicing, increasing overall operational costs.

By using real-time and historical sensor data, predictive maintenance (PdM) provides a data-driven solution to these problems by anticipating component failures before they happen. PdM systems integrate advanced analytics and machine learning (ML) models to analyze vehicle health parameters and provide actionable insights. These systems have the potential to extend component life, reduce repair costs, and enhance safety by preventing unexpected malfunctions.

For predicting component failures and identifying anomalies, machine learning techniques—in particular, LSTM networks and Random Forest classifiers—offer efficient solutions. While Random Forest classifiers are excellent at identifying anomalous patterns in car sensor data, LSTMs are best suited for time-series forecasting, which makes them perfect for predicting RUL. This project integrates these ML models into a user-friendly application to assist vehicle owners and fleet operators in making informed maintenance decisions. By implementing a real-time monitoring system, this approach ensures timely intervention, reducing vehicle downtime and improving overall efficiency.

2. REVIEW OF LITERATURE

Arena et al [1] speaks about Predictive maintenance (PdM) surpassing traditional methods, which include run-to-failure (RtF) and planned preventive maintenance (PvM). RtF involves performing maintenance only after equipment failure, often when such failures don't affect productivity. PvM, on the other hand, relies on regular, scheduled maintenance to reduce the chances of unexpected breakdowns. PdM uses IoT and condition-monitoring technology to track equipment performance over time, enabling predictions of failures and minimizing downtime, making it the most efficient approach. Shown in Fig 1.1

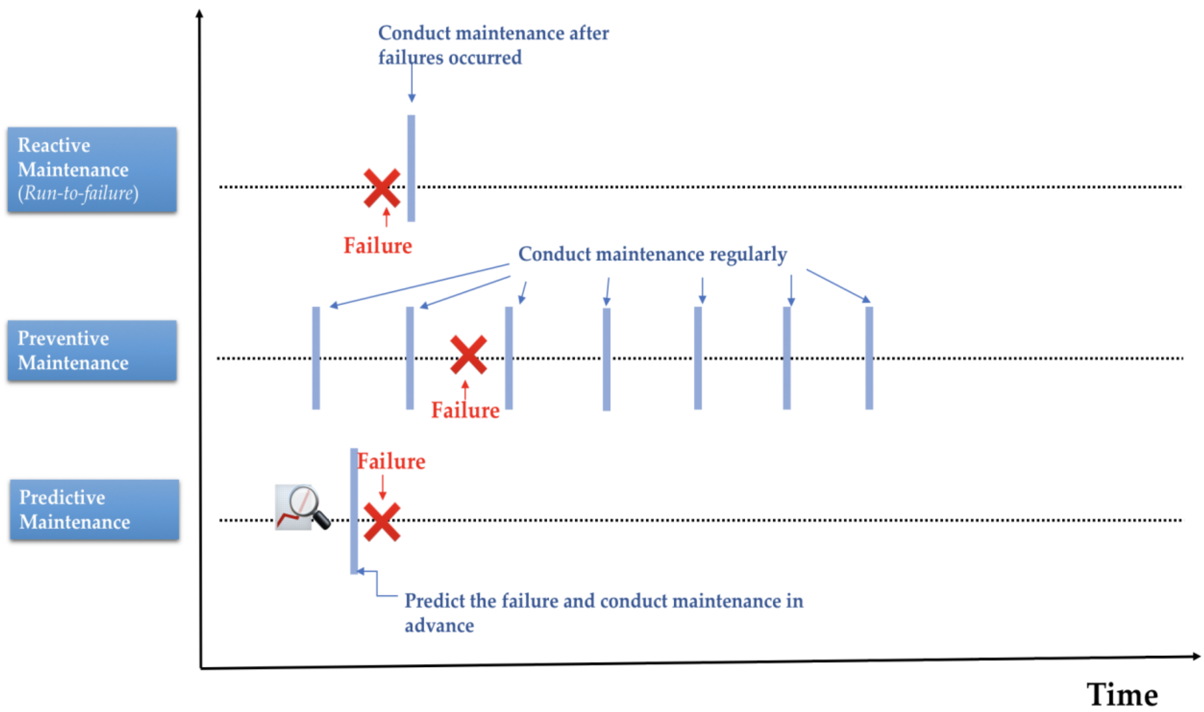


Figure 1.1

A comprehensive examination of predictive maintenance (PdM) in the automotive industry with an emphasis on machine learning (ML)-based techniques is provided by Theissler et al [2]. The study divides PdM approaches into three primary categories: condition-based PdM, statistical PdM, and remaining usable life (RUL) calculation. While condition-based PdM uses real-time vehicle sensor data to find defects or anomalies in important parts like engines, batteries, gearboxes, and braking systems, statistical PdM uses fleet-wide historical maintenance data to find failure trends. In order to optimize maintenance schedules, the study also investigates RUL estimate, in which machine learning models forecast the lifespan of vehicle components. They draw attention to several important issues with ML-based PdM, such as the lack of publicly accessible datasets, the requirement for interpretable AI models to guarantee decision-making openness, and the reliance on labeled data for supervised learning. Furthermore, according to the study, combining data from many sources (such as sensor readings, diagnostic issue codes, and car telematics) may improve prediction accuracy.

Sengupta et al [6] used an ensemble of machine learning models to offer a predictive maintenance strategy for armored vehicles. To forecast maintenance requirements based on sensor data, the study combines Light Gradient Boosting, Random Forest, Decision Tree, Extra Tree Classifier, and Gradient Boosting. Data preparation, feature selection, and model evaluation using TOPSIS analysis and K-fold cross-validation are all part of the methodology. Their combined strategy produced a remarkable 98.93% accuracy rate, with precision and recall of 99.80% and 99.03%, respectively. The study highlights the expanding use of AI in predictive maintenance and makes recommendations for future research, such as adding more data sources including operator behavior and environmental conditions to strengthen the model's resilience.

Khune et al [7] proposed a predictive maintenance framework for automobiles using IoT and machine learning to enhance reliability and reduce costs. The study collects real-time sensor data (e.g., accelerometers, GPS, temperature, oil pressure), processes it, and applies machine learning models like decision trees, random forests, and neural networks for fault prediction. The models are evaluated using accuracy, precision, and recall to optimize maintenance schedules and prevent unexpected failures. The authors also discuss the MQTT communication protocol and its implementations, Eclipse Mosquitto and Eclipse Paho, for real-time data transmission in connected vehicles.

Chirag Vinalbhai Shah [8] says that he designed 2 machine learning algorithms for predicting the State of Health (SOH) of batteries in autonomous vehicles (AVs). He used real-world data from AVs to train and test these algorithms.The first algorithm, a Stacked Autoencoder, demonstrated superior performance in terms of accuracy and response time. The second algorithm, a Random Forest regressor, also performed well, achieving accuracy close to the service level agreement (SLA) requirements and a similar response time.

The paper [9] “Machine Learning Applications in Predictive Maintenance for Vehicles: Case Studies” examines how machine learning (ML) enhances predictive maintenance (PdM) in automobiles, reducing unexpected failures and maintenance costs. It explores the use of machine learning techniques, such as decision trees, random forests, and neural networks, to vehicle parts like tires, electrical systems, and powertrains. The study emphasizes the difficulties in doing traditional maintenance, case studies on fleet management and anomaly detection, and the function of IoT sensors in data collection. It highlights PdM's financial advantages and how it integrates with cutting-edge technology like driverless cars, providing information for further study and advancement.

Tessaro et al [11] conducted experiments on predictive maintenance using different machine learning algorithms to find the one with best accuracy. Machine learning techniques like random forests, support vector machines, and neural networks can be used to predict faults in automotive engine components. After 90 runs, the statistics for the accuracy of all five machine learning methods were evaluated, as shown by the results in the Fig 1.2 below.

Table 1. Accuracy statistics for selected machine learning methods with no filtering (90 runs).

Method	Minimum	Mean	Median	Maximum	Standard Deviation
SLFN	0.67917	0.74440	0.74849	0.77105	0.01842
RVFL	0.76067	0.77493	0.77503	0.78577	0.00535
SVM	0.80612	0.80612	0.80612	0.80612	0.00000 [1]
RF	0.88539 [1]	0.88749 [1]	0.88746 [1]	0.88976 [1]	0.00108
GP	0.78371	0.79245	0.79293	0.80300	0.00433

Figure 1.2

These models can be trained on data generated by simulation testbeds that mimic different driving conditions and fault behaviors. Using a moving average filter can improve accuracy, but it may introduce a delay in detecting faults.

The study "Predictive Maintenance of Urban Metro Vehicles: Classification of Air Production Unit Failures Using Machine Learning" by Najjar et al [14] introduces a novel framework for real-time failure detection in the Air Production Unit (APU) of metro vehicles using machine learning. The work uses Random Forest (RF) classifiers for both binary and multi-class failure classification, utilizing the MetroPT dataset. The robustness of the framework was demonstrated with an F-score of 97% for multi-class classification and 85% for binary classification. One important conclusion is that analog sensors are more important than digital ones for guaranteeing safe and effective train operations. The study builds on previous research in predictive maintenance and fault detection, distinguishing itself by its real-time application and focus on metro APUs.

The paper "Review—Deep Learning Methods for Sensor-Based Predictive Maintenance and Future Perspectives for Electrochemical Sensors" by Namuduri et al [15] provides an in-depth review of deep learning (DL) techniques for predictive maintenance (PdM) using sensor data. The study explores various DL algorithms, including artificial neural networks (ANNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and autoencoders, highlighting their effectiveness in fault diagnosis, anomaly detection, and remaining useful life (RUL) estimation. The superiority of LSTM models over conventional machine learning techniques is demonstrated in a case study utilizing NASA's turbofan engine dataset, which achieved a 99.3% fault prediction accuracy.

Ersöz et al [17] conducted a systematic review of 52 studies on predictive maintenance (PdM) in transportation, covering aeronautics, automotive, railways, and maritime industries. The paper emphasizes the growing application of AI-driven models for fault detection and Remaining Useful Life (RUL) prediction, specifically LSTM networks, SVM, and Random Forests. A key trend is the integration of digital twin technology and IoT-based predictive analytics, aligning with Industry 4.0 advancements. Even though PdM research has progressed, there are still gaps in the railway and maritime industries, highlighting the necessity of high-frequency data processing, real-time sensor integration, and hybrid AI models to improve system dependability and prediction accuracy.

3. PROBLEM STATEMENT

Vehicle maintenance is often based on fixed schedules or reactive repairs, leading to inefficient resource utilization and unexpected failures. These traditional approaches do not effectively consider the real-time condition of vehicle components, resulting in either excessive maintenance costs or potential system failures. The absence of predictive insights means that vehicle owners and fleet operators may struggle to make data-driven maintenance decisions, leading to operational inefficiencies and safety concerns.

There is a need for a predictive maintenance system that can accurately estimate the Remaining Useful Life (RUL) of vehicle components and detect potential failures in advance. By leveraging machine learning models, particularly LSTMs for time-series analysis and Random Forest for anomaly detection, a data-driven approach can be implemented to provide early warnings of potential issues.

Additionally, vehicle sensor data needs to be efficiently processed and analysed to improve prediction accuracy. Challenges such as noisy data, missing values, and sensor variability must be addressed to develop a robust maintenance prediction system. This project aims to develop an ML-powered system to provide proactive maintenance recommendations based on real-time vehicle telemetry data, ultimately improving vehicle performance, reducing unexpected breakdowns, and optimizing maintenance costs.

4. OBJECTIVES

After reviewing several literature on the topic predictive maintenance in automobiles, the following objectives were derived :

- To develop a predictive model that can anticipate whether a vehicle needs maintenance or not using sensor data.
- To implement an LSTM model for predicting the RUL(Remaining Useful Life) of vehicle components.
- To utilize a Random Forest classifier for anomaly detection based on sensor data.
- To develop a web-based dashboard for real-time monitoring and maintenance recommendations.
- To improve vehicle reliability and reduce maintenance costs through data-driven insights.

5. METHODOLOGY

A. Dataset

The dataset used in this study consists of real-world vehicle telemetry data, collected from Hyundai vehicles, and includes crucial sensor readings that impact vehicle maintenance. The dataset comprises the following key columns:

- **Engine Temperature (°C)**: Indicates the engine's heat levels, with values exceeding 100°C being critical.
- **Brake Pad Thickness (mm)**: Measures brake pad wear, with values below 3mm

indicating an urgent need for replacement.

- **Tire Pressure (PSI)**: Represents the air pressure in tires, with an optimal range of 30-35 PSI.
- **Adjusted Remaining Useful Life (RUL)**: A target variable indicating the estimated number of days before a major failure.
- **Anomaly Indication**: A binary label (0 or 1) identifying whether a specific data instance represents an anomaly in vehicle performance.

This dataset is suitable for predicting the remaining useful life (RUL) of vehicles or classifying whether maintenance is required based on sensor data. Each record represents a snapshot of the vehicle's operating conditions at a given time. Fig 1.3 shows a sample from the dataset

Engine Temperature (Â°C)	Brake Pad Thickness (mm)	Tire Pressure (PSI)	Anomaly Indication	Maintenance Type Label	Component Replacement	Repair	Routine Maintenance	Adjusted_RUL
81.02239021	7.984017797	35.96454588	0	1	FALSE	TRUE	FALSE	69.97532119
98.07602856	10.71869215	32.14359304	1	2	FALSE	FALSE	TRUE	61.87818119
81.20596747	10.98307044	31.0586279	1	2	FALSE	FALSE	TRUE	70.20197002
86.08129363	7.045311036	28.539264	0	1	FALSE	TRUE	FALSE	65.5853304
93.49656806	9.948991125	33.59955954	1	0	TRUE	FALSE	FALSE	63.95122433
89.37134124	8.799711737	31.30900417	0	0	TRUE	FALSE	FALSE	65.09601491
96.49011118	6.349810454	32.35825162	1	2	FALSE	FALSE	TRUE	60.96651892
92.3144635	10.69672351	33.09971304	0	2	FALSE	FALSE	TRUE	64.94140026
80.73294547	10.90309206	31.84178927	0	0	TRUE	FALSE	FALSE	70.36312194
83.98229343	6.147723248	33.53667467	1	2	FALSE	FALSE	TRUE	67.37527752
80.99723971	9.011146386	37.14247544	0	1	FALSE	TRUE	FALSE	70.63433379
82.23076358	7.096161678	37.82225107	0	2	FALSE	FALSE	TRUE	69.4875331
81.74513596	10.41525166	36.38435416	1	2	FALSE	FALSE	TRUE	70.77040352
98.01655782	7.409704242	36.53729704	1	0	TRUE	FALSE	FALSE	61.26306219
98.87036588	9.924175474	29.74579912	0	1	FALSE	TRUE	FALSE	60.58364707
99.48415983	6.778730643	29.07674554	1	0	TRUE	FALSE	FALSE	58.78476145
84.05482499	9.520207377	29.51572488	1	2	FALSE	FALSE	TRUE	67.88381543
99.61511544	11.04137876	30.99108964	0	1	FALSE	TRUE	FALSE	60.90721171
85.97046211	9.077044394	34.87778965	1	0	TRUE	FALSE	FALSE	67.62114463
98.03537726	7.815089442	35.73373404	0	1	FALSE	TRUE	FALSE	61.35509395
98.71657267	8.5501812	34.78929544	0	2	FALSE	FALSE	TRUE	61.21964523
94.26276455	9.266300513	28.94491464	1	0	TRUE	FALSE	FALSE	62.36412086
91.84677616	8.447278024	35.14574526	0	0	TRUE	FALSE	FALSE	64.48467218
84.11712729	9.459858703	33.89876159	1	0	TRUE	FALSE	FALSE	68.50513215
95.26418723	6.044877787	31.20404946	0	0	TRUE	FALSE	FALSE	61.02666739

Figure 1.3

B. *Detailed Methodology*

Data Preprocessing Steps :

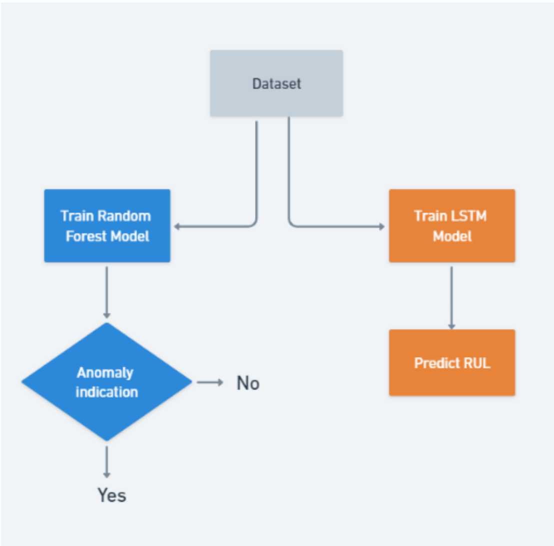
The dataset was considerably preprocessed prior to training. To find the most relevant sensor values for predictive maintenance, feature selection was done. To ensure consistency across feature values, scaling and normalization were implemented using Standard Scaler. Ultimately, for efficient model evaluation and validation, the dataset was divided into training (80%) and testing (20%) subsets.

Training the Models :

The predictive maintenance system employs two machine learning models: an LSTM model for Remaining Useful Life (RUL) prediction and a Random Forest classifier for anomaly detection. The LSTM model is specifically designed for processing sequential time-series data, making it well-suited for RUL prediction. Its architecture consists of an initial LSTM layer with 128 units using ReLU activation, followed by dropout regularization at 30% to prevent overfitting. Batch normalization is applied before a second LSTM layer with 64 units, again followed by dropout regularization. A fully connected dense layer with 32 neurons processes the extracted features before the final output layer, which predicts RUL using a linear activation function.

The LSTM model training process involves optimization using the Adam optimizer with a learning rate of 0.001. Mean Squared Error (MSE) is used as the loss function, while Mean Absolute Error (MAE) serves as a performance metric. To enhance model robustness, early stopping is employed with a patience of 15 epochs to prevent overfitting, alongside ReduceLROnPlateau to adjust learning rates dynamically when the loss plateaus.

For anomaly detection, a Random Forest classifier is implemented. This model is configured with 200 decision trees, a maximum depth of 15, and a minimum samples split of 8, using Gini impurity as the criterion for evaluating splits. Training is conducted on labeled sensor data, where predefined anomaly indications allow the model to learn patterns distinguishing normal operations from failures. Hyperparameter tuning is performed using Grid Search to optimize classification accuracy. Model performance is evaluated based on accuracy to assess its ability to detect anomalies effectively.

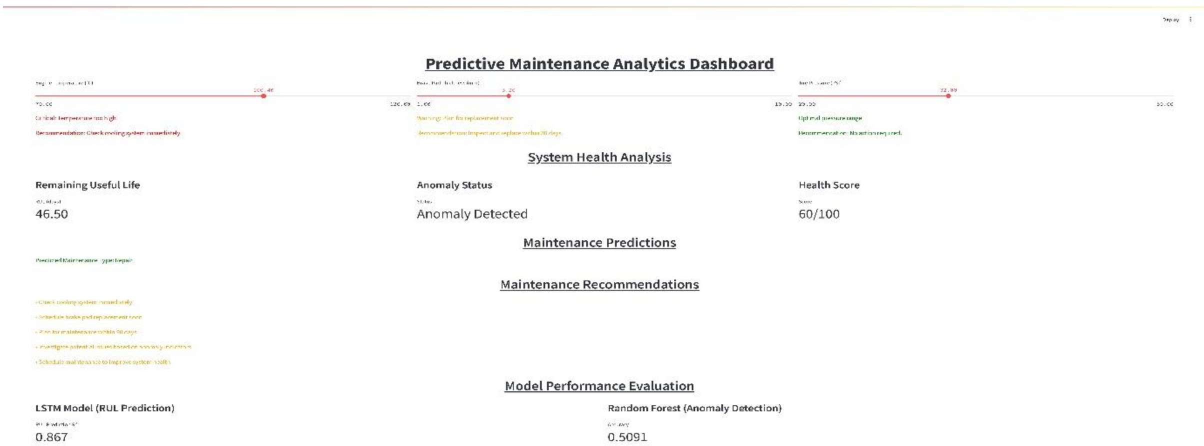


Deploying the model :

The trained models are deployed using a Streamlit-based web application, enabling real-time vehicle health monitoring. The application features an interactive user input panel where users can enter sensor readings like Engine Temperature (°C), Brake Pad Thickness (mm) and Tire Pressure (PSI) via sliders. Based on the

provided inputs, the LSTM model predicts the Remaining Useful Life of vehicle components, while the Random Forest classifier identifies potential anomalies. A weighted scoring function calculates an overall health score, allowing users to assess their vehicle's condition.

The web application also generates maintenance recommendations based on predicted RUL and anomaly scores. These recommendations provide actionable insights for preventive measures to minimize failures. To improve interpretability, the application includes visualizations displaying health score trends, RUL projections, and anomaly likelihoods. The integration of these features enhances usability, providing vehicle owners and fleet managers with a comprehensive decision-support tool for predictive maintenance. Figure below shows the example interface that was designed using Streamlit :



Evaluation Metrics :

The effectiveness of the predictive maintenance system is validated through model evaluation metrics. The LSTM model achieves a R^2 value of 0.867, demonstrating its accuracy in predicting Remaining Useful Life. The Random Forest classifier exhibits performance with an accuracy of 0.5091 indicating its reliability in detecting anomalies. These evaluation results confirm the robustness and applicability of the proposed predictive maintenance framework for real-world vehicle monitoring.

6. CONCLUSION

This project successfully implements a predictive maintenance system using machine learning to improve vehicle reliability. By leveraging LSTM networks for RUL prediction and Random Forest classifiers for anomaly detection, the system provides accurate maintenance recommendations. The developed Streamlit-based web application enables real-time monitoring of vehicle health, helping users take proactive maintenance actions. This approach not only enhances vehicle safety but also optimizes maintenance costs, making it a valuable solution for the automotive industry.

While approaches like recurrent neural networks (RNNs) have shown promise, challenges such as data quality and algorithm selection remain. Overall, PdM is poised to revolutionize vehicle maintenance, especially for autonomous and connected vehicles, though further research is needed to optimize its broader application.

REFERENCES

1. Arena Fabio, Collotta Mario, Luca Liliana, Ruggieri Marianna & Termine Francesco Gaetano "Predictive Maintenance in the Automotive Sector: A Literature Review." Mathematical and Computational Applications, vol. 27, no. 1, 1 Feb. 2022, p. 2.
2. Theissler, Andreas & Perez-Velazquez, Judith & Kettelgerdes, Marcel & Elger, Gordon. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. Reliability Engineering [?] System Safety. 215. 107864. 10.1016/j.ress.2021.107864.
3. GÜVEN, Özlem & Şahin, Hasan. (2022). Predictive Maintenance Based On Machine Learning In Public Transportation Vehicles. Mühendislik Bilimleri ve Araştırmaları Dergisi. 4. 10.46387/bjesr.1093519.
4. Yamato, Yoji & Fukumoto, Yoshifumi & Kumazaki, Hiroki. (2016). Proposal of Real Time Predictive Maintenance Platform with 3D Printer for Business Vehicles. International

Journal of Information and Electronics Engineering. 6. 10.18178/IJIEE.2016.6.5.640.

5. Chaudhuri, Arindam. (2018). Predictive Maintenance for Industrial IoT of Vehicle Fleets using Hierarchical Modified Fuzzy Support Vector Machine. 10.48550/arXiv.1806.09612.
6. Sengupta, Prajit & Mehta, Anant & Rana, Prashant. (2023). Predictive Maintenance of Armoured Vehicles using Machine Learning Approaches. 10.48550/arXiv.2307.14453.
7. Prof. Khune, Priya, et al. "PREDICTIVE MAINTENANCE in AUTOMOBILE USING IOT and MACHINE LEARNING." International Research Journal of Modernization in Engineering Technology and Science, vol. 05, no. 03, 6 Mar. 2023.
8. Shah, Chirag. (2024). Machine Learning Algorithms for Predictive Maintenance in Autonomous Vehicles. International Journal of Engineering and Computer Science. 13. 26015-26032. 10.18535/ijecs/v13i01.4786.
9. Ravi, Aravind & Surabhi, Manogna & Shah, Chirag. (2022). Machine Learning Applications in Predictive Maintenance for Vehicles: Case Studies. International Journal Of Engineering And Computer Science. 11. 25628-25640. 10.18535/ijecs/v11i08.4707.
10. Jabari, Youssef, et al. Implementing Models Approach for Predictive Maintenance in the Context of Industry 4.0: A Literature Review and Comparative Analysis.
11. Tessaro, Iron & Mariani, Viviana & Coelho, Leandro. (2020). Machine Learning Models Applied to Predictive Maintenance in Automotive Engine Components. Proceedings. 64. 26. 10.3390/IeCAT2020-08508.
12. Jafari, Reza. "Machine Learning for Predictive Maintenance in Autonomous Vehicle Fleets." Journal of Artificial Intelligence Research and Applications, vol. 3, no. 2, Dec. 2023.
13. Dhamodharan, Balaji. Optimizing Industrial Operations: A Data-Driven Approach to Predictive Maintenance through Machine Learning Balaji Dhamodharan, Independent Researcher. Mar. 2021.
14. Najjar, Ayat & Ashqar, Huthaifa & Hasasneh, Ahmad. (2023). Predictive Maintenance of Urban Metro Vehicles: Classification of Air Production Unit Failures Using Machine Learning.
15. Namuduri, Srikanth & Narayanan, Barath & Davuluru, Venkata Salini Priyamvada & Burton, Lamar & Bhansali, Shekhar. (2020). Review—Deep Learning Methods for Sensor Based Predictive Maintenance and Future Perspectives for Electrochemical Sensors. Journal of The Electrochemical Society. 167. 037552. 10.1149/1945-7111/ab67a8.
16. Nacchia, M.; Fruggiero, F.; Lambiase, A.; Bruton, K. A Systematic Mapping of the Advancing Use of Machine Learning Techniques for Predictive Maintenance in the Manufacturing Sector. Appl. Sci. 2021, 11, 2546. https://doi.org/10.3390/ app11062546.
17. Ersöz, Olcay & Inal, Ali & Aktepe, Adnan & Turker, Ahmet & Ersöz, Süleyman. (2022). A Systematic Literature Review of the Predictive Maintenance from Transportation Systems Aspect. Sustainability. 14. 14536. 10.3390/su142114536.