MACHINE LEARNING – BASED PREDICTIVE MAINTENANCE FOR AUTOMOBILES HEALTH MONITORING : INTEGRATING RANDOM FOREST AND LSTM MODELS

Drupath S Prasad[1] and Alia Teresa T.M

[1]Scholar
Department of Computer Science
Sacred Heart College, Thevara
Ernakulam, Kerala, India
drupathsprasad@gmail.com

[2]Assistant Professor
Department of Computer Science
Sacred Heart College, Thevara
Ernakulam, Kerala, India
aliateresa@shcollege.ac.in

Abstract

In the automotive sector, pdm has come to be seen as a vital approach for minimizing unscheduled breakdowns as well as maintenance planning. Also, conventional predictive maintenance techniques allow some truck down time for vehicles inspection and maintenance, which can increase expense. This work employs machine learning (ML) models including Random Forest classifiers and Long Short-Term Memory (LSTM) networks to predict the RUL of vehicular components, as well as look for abnormal input signals in sensor readings. A web application based on Streamlit is provided where users can insert certain parameters of the car to be able to obtain real-time health scores and maintenance recommendations. This system offers a proactive approach to automobile maintenance, giving rise to increased efficiency and reliability.

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

1. INTRODUCTION

With two traditional strategies-a reactive one and a preventive one-the automotive industry has failed to ensure high levels of vehicle reliability at lower maintenance costs. Such a reactive or preventive approach has created inefficiencies, increased costs, and seized assurance of equipment. A reactive maintenance approach refers to maintenance performed after a vehicle breakdown, meaning repairs could very costly and hazardous. Preventive maintenance uses a time-specified regimen that could mean unnecessary servicing processes hence increasing the overall operational cost.

Predictive maintenance (PdM) captures the state of a component and identifies the best time within a specific operating window for maintenance. Instead of waiting for a problem to occur, PdM collects real-time, often historical, sensor data and then produces timely alerts before failure occurs. PdM systems utilize advanced analytics and machine-learning (ML) models to analyze vehicle health parameters and provide actionable insights. This system could be theoretically capable of improving component life, reducing repair costs, and enhancing safety by diminishing the occurrence of unscheduled malfunctions.

Machine learning techniques, especially LSTM networks and Random Forest classifiers, provide efficient solutions to the problem of predicting component failures and identifying anomalies. While Random Forest classifiers are very good at identifying anomalous patterns from the car sensor data, LSTMs are more suited to time-series prediction, which perfectly matches with the problem of remaining useful life. This project integrates these ML models and embeds them in a user-friendly application that would assist vehicle owners and fleet operators to make timely maintenance decisions. Timely intervention through this approach reduces vehicle downtime and ultimately improves overall efficiency.

The rest of the paper is organized as follows. Section 2 consists of Review of Literature, Section 3 describes the Problem Statement on the Predictive Maintenance. Next we provided the Objectives in Section 4. The description of dataset and detailed methodology is presented in Section 5.Finally a conclusion is given in Section 6.

2. LITERATURE REVIEW

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

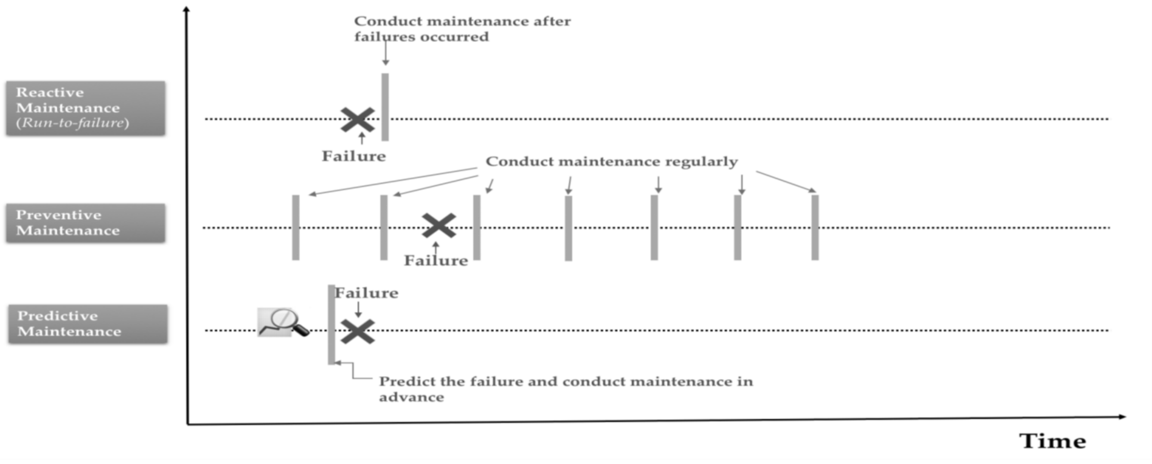

Figure 1.1

Theissler et al. have provided a complete overview of predictive maintenance (PdM): a focus is made on machine learning (ML)-based techniques used in the automotive industry. The study divides the PdM approaches into three key categories: condition-based PdM, statistical PdM, and the remaining usable life (RUL) calculation. While condition-based PdM employs real-time vehicle sensor data to detect defects or anomalies in critical components like the engine, battery, gearbox, and braking system, statistical PdM aims to detect failure tendencies using historical maintenance data collected across fleets. Besides, RUL estimate is considered by the study as machine learning models predicting the life of the dependent automobile components. Several important aspects concerning ML-based PdM, such as limited access to publicly available datasets, the need for interpretable AI models to guarantee the transparency of decision-making, and the need for labelled data for supervised learning, are highlighted. Furthermore, the study suggests that merging data based on abundant sources, such as sensor readings, diagnostic issues codes, and car telematics, may help increase prediction accuracy.

A study by Sengupta et al. [6] described a method of predictive maintenance based on machine learning models for armored vehicles. Using sensor inputs such as Light Gradient Boosting, Random Forest, Decision Tree, Extra Tree Classifier, and Gradient Boosting, this paper identifies the requirements of maintenance. Data preparation, feature selection, and testing of models were done through the method of TOPSIS analysis and K-fold cross-validation. With an ensemble strategy, the model achieved an astonishing accuracy of 98.93%, with precision and recall being 99.80% and 99.03%, respectively. It mentions the growing use of AI in predictive maintenance and suggests more data inputs--such as operator behavior and some environmental conditions--to improve resilience in the future.

The predictive maintenance framework for automobiles based on IoT and machine learning for reliability and cost reduction was proposed by Khune et al. [7]. The study was such as to collect real-time sensor data (accelerometer, GPS, temperature, and oil pressure), process it, and then apply machine learning models like decision trees, random forests, and neural networks to predict faults. Accuracy, precision, and recall to test models help in optimizing the schedule of maintenance and avoiding unexpected failures. The MQTT communication protocol as such and its implementations Eclipses Mosquitto and Eclipse Paho are also discussed by the authors for online communication 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] used various kinds of machine learning algorithms on predictive maintenance to identify the most accurate algorithm. Various machine learning techniques like random forests, support vector machines, and neural networks, can predict the faults that occur in automotive engine components. The performance data of the accuracy ratings for all five different machine learning methods were pooled together after 90 runs, with results shown in a below Fig. 1.2.

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 provided by an F-score of 97% on multi-class classifications, while a score of 85% on binary classifications. An important conclusion made from the study is stressed upon the unfailing efficiency and safety of train operations with the presence of analog sensors over digital ones. The study builds on previous research on predictive maintenance and fault detection, standing out due to its real-time applications 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] describes in detail deep learning (DL) techniques for predictive maintenance (PdM) using sensor data. The author reviewed various deep learning algorithms, which included artificial neural networks (ANN), convolutional neural networks (CNN), recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and autoencoders, commenting on their utility in fault diagnosis, anomaly detection, and remaining useful life (RUL) estimation. A case study employing the NASA turbofan engine dataset demonstrated that LSTM models substantially outperform classical machine learning techniques with a fault prediction accuracy of 99.3%.

Ersöz et al. [17] conducted a systematic review of 52 studies on predictive maintenance in transportation, including aeronautics, automotive, railways, and maritime industry. The paper pointed out the rising usage of AI-based models for fault detection and Remaining Useful Life prediction, especially LSTM networks, SVM, and Random Forests. A hot topic is digital twin technology and IoT-based predictive analytics-based implementation, keeping pace with Industry 4.0 advancement. Though there has been some advancement in the field of PdM, there is evidently a research gap in the railway and maritime industries that open avenues for high-frequency data processing, integration into the real-time scenario of sensors, and hybrid AI modeling, so as to push the boundaries of reliability and prediction accuracy.

3. PROBLEM STATEMENT

The maintenance of vehicles is relied upon a fixed schedule or reactive repair works, both of which lead to sub-optimal utilization of the available resources as well as an unexpected failure on many occasions. Conventional approaches fail to consider the real-time condition of the components in the vehicle and as such, one of these broad techniques can lead to excessive spending on maintenance and the risk of breakdowns. Information on its predictive existence does enable vehicle owners and fleet operators to make data-led driving decisions on maintenance, with a notable effect on operational and safety issues.

There is a need for a preventive maintenance system that provides timely estimating of Remaining Useful Life (RUL) of vehicle components, with warnings of potential failure beforehand. The data-driven approach calls for the Union of Machine Learning models like LSTMs for time-series analysis, while Random Forests are employed in detecting anomalies by providing an early warning on possible problem escalation.

Additionally, vehicle sensor data will need to be processed and analyzed for prediction accuracy. There will be challenges presented in creating a robust maintenance prediction system, including noisy data, missing data, and variability in the sensors used. This project will work on developing an ML-based system that enables predictive maintenance recommendations based on real-time vehicle telemetry data, thus enhancing vehicle performance and minimizing unexpected breakdowns and in turn optimizing maintenance costs.

4. OBJECTIVES

After several review of literatures on the topic : predictive maintenance in automobiles, the following objectives were derived :

- To develop a predictive model that can predict whether a vehicle needs maintenance or not using sensor data.

- To implement an LSTM model for predicting the RUL(Remaining Useful Life) of the vehicle.
- 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 reliability of vehicle and reduce maintenance cost through data-driven insights.

5. METHODOLOGY

5.1 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

5.2 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 split into training (80%) and testing (20%) subsets.

Training the Models :

The predictive maintenance system applies two machine-learning models: an LSTM model for predicting the Remaining Useful Life feature and a Random Forest model for anomaly detection. The LSTM model in use is suited for RUL prediction as it processes sequential time-series data. The architecture consists of one initial LSTM layer with 128 units by applying ReLu activation, followed by dropout regularization on it applied at 30% to prevent overfitting. Batch normalization is applied ahead of the second LSTM layer of 64 units and dropout is applied after this layer too. The features extracted go to the fully connected dense layer containing 32 neurons and feeding into the final output layer to predict RUL based on a linear activation function.

The LSTM Model training is carried out using the Adam optimizer, with a learning rate of 0.001, while the loss function is defined as mean squared error (MSE), and performance is evaluated based on MAE. Use of early stopping with a patience of 15 epochs was included in order to avoid the overfitting and also ReduceLROnPlateau is used to dynamically alter the learning rate when a plateau in loss is reached.

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.

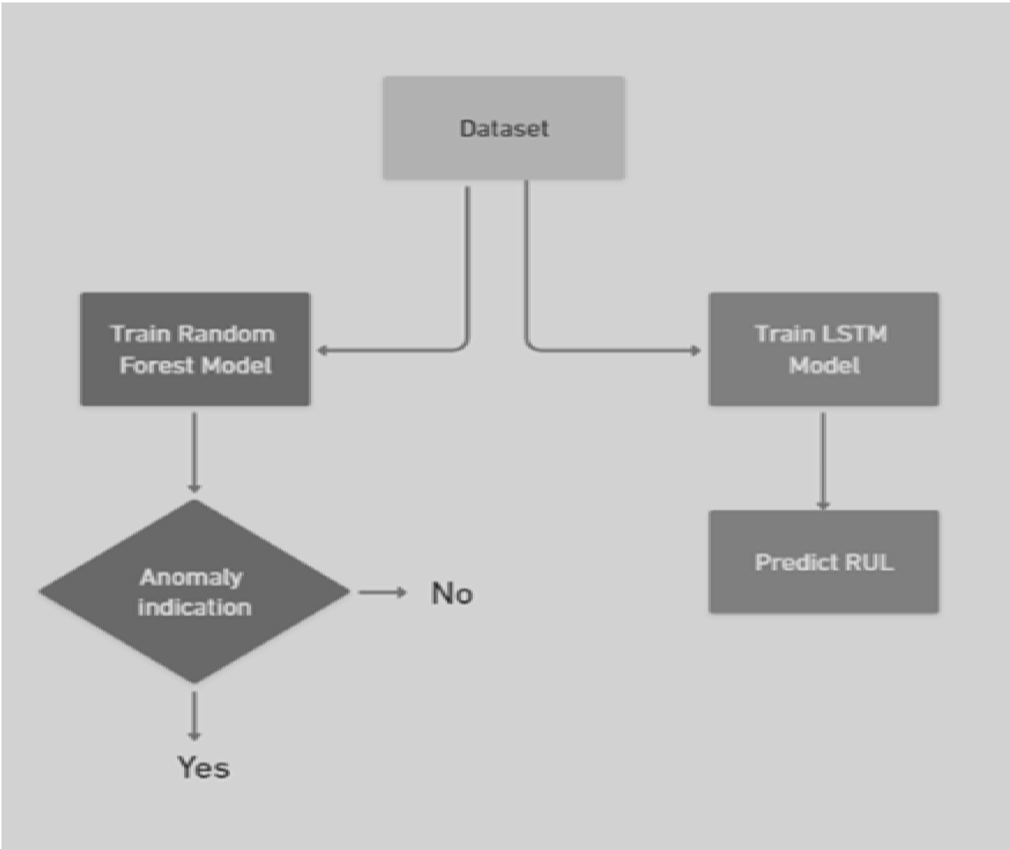

Figure 1.4 . Predictive maintenance framework

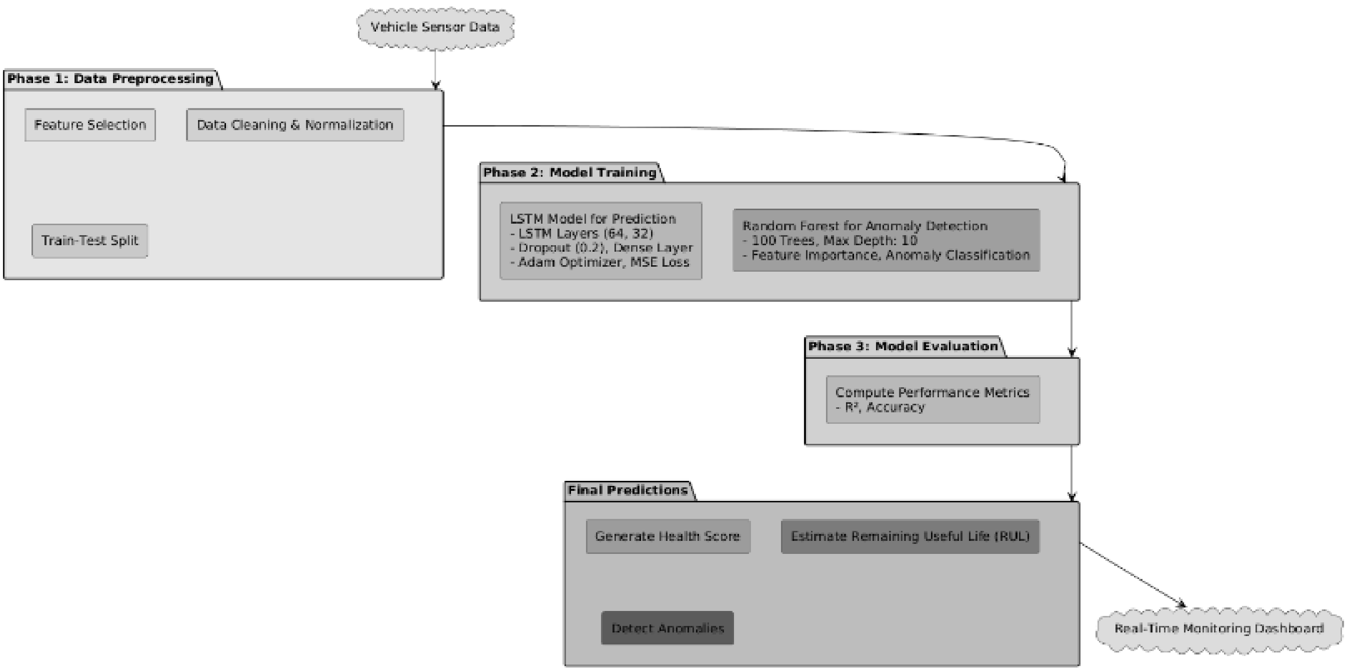

Figure 1.5 .Detailed Methodology Representation of Proposed Model

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. In order to improve interpretability, visualizations display health score trends, RUL projections, and probability of anomalies; all of these elements take usability to another level, enabling vehicle owners and fleet managers to use predictive maintenance as a decision-support tool. An example interface, made with Streamlit, is shown in Figure 1.6 below: The methodology of this project follows a structured workflow comprising data preprocessing, feature extraction and model implementation discussed below:

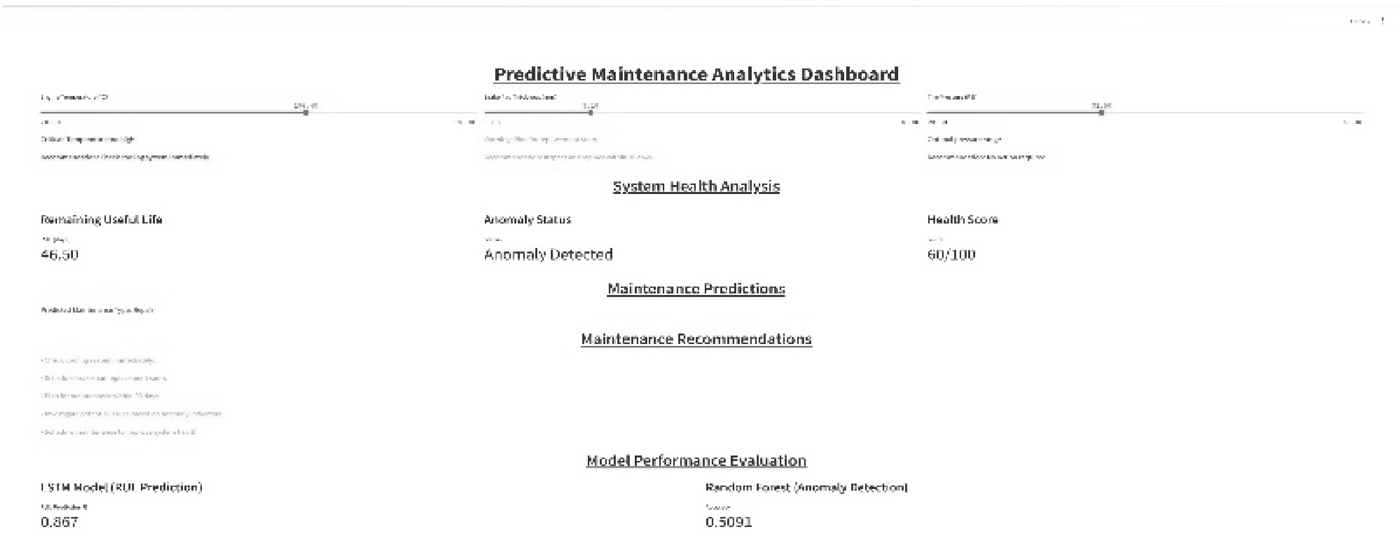

Figure 1.6

Evaluation Metrics :

The effectiveness of the predictive maintenance system is verified through model evaluation metrics. LSTM model achieved an R^2 value of 0.867, indicating accuracy in predicting Remaining Useful Life. Random Forest classifier performed with an accuracy of 0.5091, showcasing reliability in anomaly detection. The evaluation results compare favorably with live applications of the chosen predictive maintenance framework for vehicle monitoring.

6. CONCLUSION

This project successfully implements an actual predictive maintenance system using machine learning to improve the reliability of vehicles. The system recommends maintenance using LSTM networks for predicting RUL and Random Forest classifiers for detecting anomalies. The developed web application, which is based on Streamlit, enables vehicle health monitoring in real time, enabling users to act proactively. This approach improves safety and optimizes maintenance costs and thus provides a viable solution for the automotive industry.

Although such approaches like RNNs have shown great potential, challenges of data quality as well as algorithm selection still exist. Overall, PdM is going to be a game changer for vehicle maintenance especially for autonomous and connected vehicles but more researches are required to optimize it for its broader applications.

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.

[18] D Divya, Bhasi Marath, and MB Santosh Kumar. Review of fault detection techniques for predictive maintenance. Journal of Quality in Maintenance Engineering, 2022.

[19] Guo, J.; Lao, Z.; Hou, M.; Li, C.; Zhang, S. Mechanical fault time series prediction by using EFMSAE-LSTM neural network.Measurement **2021**, 173, 108566.

[20] Safavi, S.; Safavi, M.A.; Hamid, H.; Fallah, S. Multi-Sensor Fault Detection, Identification, Isolation and Health Forecasting for Autonomous Vehicles. Sensors 2021, 21, 2547.