A DATA-DRIVEN APPROACH FOR PREDICTING VEHICLE COMPONENT FAILURES

Ms. Priyanka Niranjan Savadekar  
Assistant Professor  
Presidency university, Yelahanka,Bengaluru, Karnataka, India  
priyankasavadekar10@gmail.com

G.Giridhareswar   
Department of CSE, SoE,  
Presidency university, Yelahanka,Bengaluru, Karnataka, India  
galigeethu45gmail.com

A. Prudhvi Reddy  
Department of CSE, SoE,  
Presidency university, Yelahanka,Bengaluru, Karnataka, India  
prudhvireddyavula123@gmail.com

R.Rukmananda Reddy   
Department of CSE, SoE,  
Presidency university, Yelahanka,Bengaluru, Karnataka, India  
rukmanandareddy369@gmail.com

# *Abstract* - The automotive industry has faced several issues with respect to sudden failures of vehicle components, which results in reduced efficiency of the vehicles, breakdowns, and increased maintenance costs. To this end, this research suggests a new approach to predictive maintenance based on machine learning to predict the failures of critical vehicle components. The proposed system is developed using historical data containing operational parameters like engine RPM, lubricant oil pressure, coolant temperature, and fuel pressure to develop prediction models. The analysis of the correlation between these parameters and their impact on the component was made in detail. Several machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbours (KNN), were also compared. Among the above mentioned, the best result was achieved by Logistic Regression with the accuracy of 66%, and the second best result was obtained by using SVM with the accuracy of 92%.

# The predictive framework is also supported by a web-based dashboard that gives the user real time information on the state of the vehicle, the probability of failure and sending of maintenance reminders at the right time. This dashboard helps the fleet managers in making informed decisions as it provides suggestions based on the output of the model. Thus, the developed system brings down the occurrence of unplanned downtime, improves the maintenance schedule, and increases the dependability of vehicles and thus can have be a effectively large used number in of industries vehicles. that The use of machine learning in predictive maintenance is also introduced and the way for future also development suggested. such This as scalable real and time adaptable IoT, solution advanced is deep a learning valuable and contribution edge to computing the for growing failure field prediction of is automotive predictive maintenance and the potential to revolutionize fleet management

**index terms – predictive maintenance, automotive diagnostics, vehicle health monitoring, web-based** **dashboard, data-driven maintenance, iot in automotives threshold analysis.**

I. Introduction

Cars today are complex systems of mechanics, electrics and digital technology that need regular servicing to run efficiently. The key technology making this possible is a vehicle's dozens of sensors onboard that allow it to record massive amounts of data about the vehicle's performance and condition. But traditional maintenance practices are typically reactive, based on scheduled inspections or repairs after a failure — even with those advances. Although these approaches are common practice, they can ultimately result in inefficiency, idle time due to emergency lockdown, an increase in overhead, and safety issues.

These challenges can be systematically addressed through the adoption of predictive maintenance strategies. Predictive maintenance uses historical data and advanced analytical techniques to move from reactive to pro-active. This solution allows for the early detection of possible reliability failures and optimize the maintenance process, reduce operations costs in the long run and increase the lifespan of vehicle parts. As a fundamental building block of predictive maintenance, machine learning seems to have magical powers for finding patterns in data, predicting anomalies and providing actionable insights from complex datasets.

In this article, a machine learning-based framework for predictive maintenance is structured, where main car components like engine RPM, coolant temperature, lubricant oil pressure are highlighted with their health monitoring.

The system uses a big data analysis of historical vehicle parameters for correlation analysis and predictive modeling to predict what failures could occur. This expository has a predictive framework that examines the performance of five machine learning algorithms based on the health of a vehicle: Logistic Regression, Decision Tree, Random Forest Support Vector Machine (SVM), and K-Nearest Neighbours (KNN) and identifies the most reliable model for maintaining the health of vehicles.

The most powerful model was the Random Forest model with prediction accuracy of 90% as it takes care of feature importance and also is robust to overfitting. Additionally, a web-based dashboard tier has been constructed to extend user interactions and decision-making for the system. It uses real time data to visualize vehicle parameters, detect deviation from pre-set thresholds and validates alerts to predict any potential failure in time.

The contributions of this study are as follows:

1. Development of a comprehensive predictive maintenance framework leveraging machine learning.
2. Comparative evaluation of machine learning algorithms for vehicle component health monitoring.
3. Implementation of a user-centric, web-based dashboard for real-time monitoring and proactive maintenance planning.

The results emphasise the need for predictive maintenance as a way of enhancing vehicle reliability and performance. Through applying the machine learning approaches to vehicle diagnostics, this study offers a flexible and efficient approach for the automotive sector. Further enhancements can include real-time data input, improved model’s flexibility and its potential usage in various types of car parts.

II. RESEARCH GAP OR EXISTING METHODS

*A. Existing Methods in Vehicle Health Monitoring Systems*

Vehicle health monitoring implements several advanced methods for predicting and diagnosing the status of important components in a vehicle. Conventional diagnostic methods are mostly reliant on periodic inspection and manual evaluations, which are mostly reactive, respond to the failure only after it happens. The latest technologies are focused on combining data-driven strategies and machine learning algorithms that allow predictions and real-time monitoring.They are:

* Rule-Based Diagnostics
* Statistical Analysis Methods
* Model Based Approaches
* Signal Processing Techniques
* Data-Driven Machine Learning

As emphasized in [1] Zhang et al. (2019), leveraged large-scale data by utilizing deep learning model for the prediction and anomaly detection, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which can achieve high accuracy. Probabilistic models in machine learning (Thelen et al. [2] (2024), focus on the quantification of uncertainties in predicting battery health and aging span, delivering strong predictions even with incomplete datasets. Other important developments include the use of sensor-based systems for real-time monitoring. Raouf et al. (2022) presented a comprehensive review of sensor technologies used in advanced driver-assistance systems (ADAS) for autonomous vehicles including ADAS, which facilitate fault detection precision and minimize human involvement. Likewise, [8] Shafi et al. (2018) enables remote monitoring and predictive persistence scheduling through interconnected devices.

*B.* *Research Gaps in Current Systems*

There is still a big gap in terms of obtaining high-quality and diversified datasets. As demonstrated by Zhang et al. [1] (2019) and [3] Zhao & Burke (2022), the majority of existing systems depend on proprietary or limited datasets that generally do not capture the wide variability in the types of vehicles, their operating conditions, and the environments. Specifically, data requirements for training machine learning models are often not satisfied, and the absence of data standardization and labeling between systems hinder model generalization. Moreover, the issue of sensor data quality, specifically noise, calibration errors, and environmental disturbances, runs the risk of hampering both accuracy and reliability of the predictive model. Despite the potential of powered by machine learning and deep learning in small applications, there still remains a large gap for widely applicable and scalable solutions to be realized in experimental research on large, heterogeneous fleets. The systems for managing [2] individual Thelen components et of al. vehicles, (2024) for concerning example battery those health considered diagnostics, in fail the when paper applied by to large fleets of vehicles or to complex vehicle systems. Also, the issue of how to incorporate data from various sensors and is different a types need of to vehicles undertake into more a research single to integrated develop scalable models prediction that system are is capable yet of to handling be various solved. vehicle There configurations, and use cases without requiring extensive retraining or customization.

*C. Need for a Comprehensive Solution*

Systems currently, as stated in Zhang et al. [8] (2024) and [2] Thelen et al. (2024) are dependent on piecemeal data sources, creating challenges when creating a holistic view of a vehicle's health. Such a comprehensive solution would require the integration of data from multiple sensors (temperature, pressure, vibration, battery health...), vehicle diagnostic systems, and real-time operational data, enabling a much more complete and precise depiction of the vehicle health status. Such an approach would also make it possible to enhance the predictive power of the system and allow for more informed decisions. Examples of professional development include:

* Holistic Data Integration and Fusion.
* Scalable And Adaptive Predictive Models.
* Cost-Effective And Robust Sensor Systems.
* Integration With Maintenance And Fleet Management Systems

Another fundamental requirement for a holistic solution is the capability of low-latency processing of streaming data, particularly for safety-critical applications where immediate actions might need to be taken to prevent vehicle failure. As highlighted by [1] Zhang et al. (2019) and [11] Sutharssan et al. detection analysis (2015), present systems most of time are not capable to perform real time analysis, especially when dealing with large amount of data coming from various sensors. An end-to-end solution would leverage edge computing, cloud-based analytics, or a hybrid approach that allows data to be processed in real time, but also protects the vehicle’s compute resources from being overwhelmed. With its new system, it would analyze the data and initiate maintenance or notify operators with minimal latency, allowing for timely preventative action.

*D.CONCLUSION OF LITERATURE REVIEW*

The existing body of work demonstrates significant progress in predictive maintenance using machine learning. However, the challenges outlined above highlight the need for more robust, scalable, and interpretable systems. This research aims to address these gaps by developing a comprehensive predictive maintenance framework that integrates real-time data, employs advanced machine learning models, and incorporates a user-friendly interface for proactive decision-making. The proposed solution not only builds on the strengths of existing approaches but also introduces novel contributions to advance the field.

III. PROPOSED METHODOLOGY

*A.* *Data Acquisition and Preprocessing*

To perform machine learning on predictive maintenance and achieve success, good, pre-processed data is required. The dataset consists of operational metrics such as temperature, vibration levels, runtime, and previous failure logs. Data prediction through statistical imputation, such as median or regression. Both time series sensor data are normalized to a [0, 1] range for ensuring uniformity differentially and noise are filtered with the using moving averages or Gaussian filters for reliability. To diagnose significant predictors correlation coefficients are calculated, and relationships are visualized using a heatmap, allowing to identify relationships and finding potential candidates for a predictive model.

*B. Comparative Accuracy Analysis Of Machine Learning Models*

Here, the aim is to assess the effectiveness of different machine learning techniques for predictive failure modelling. Algorithms like Logistic Regression, Decision Tree, Random Forest, Gaussian Naive Bayes, and K-Nearest Neighbours (KNN) are being trained on the 80% of the dataset and tested on the remaining 20%. A 10-fold cross-validation approach minimizes bias. Performance metrics, including accuracy, precision, recall, and F1-score, are computed. Among the evaluated models, Random Forest consistently outperforms others in terms of accuracy and interpretability, while Logistic Regression offers faster results but underperforms on non-linear datasets.

*C. Feature-Level Accuracy Assessment*

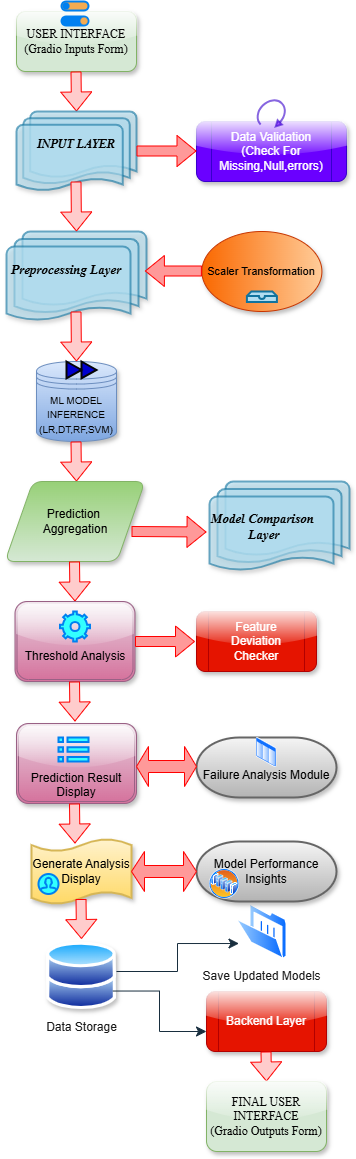


Fig. 1 System Architecture

To optimize predictive performance, feature importance analysis is conducted. Recursive Feature Elimination (RFE) identifies and eliminates weaker features iteratively, while Random Forest and SHAP (SHapley Additive exPlanations) values rank features by their contribution to model predictions. Retained critical features include vibration intensity, operational hours, and temperature anomalies, while less impactful features are excluded to improve computational efficiency.

*D. Predictive Failure Modeling*

Robust models are developed for accurate failure prediction. Training focuses on labeled historical data, particularly high-failure scenarios, with hyperparameters fine-tuned for optimal

performance. Failure thresholds are established using domain expertise and data-driven analysis.

Predictions categorize risks into levels, such as Low, Medium, and High Risk, for actionable insights*.*

*E. Real-Time Alert System via Web Dashboard*

The proposed dashboard provides a centralized interface for failure predictions and alerts. Developed using frameworks like Dash or Streamlit, the dashboard integrates seamlessly with Python-based machine learning models for real-time predictions. Key features include graphical trend visualizations, interactive filters, and automated notifications for high-risk scenarios. The alerts provide details of affected components along with suggestions for preventive maintenance actions, ensuring easy interaction with multiple levels of users and timely interventions.

IV. OBJECTIVES

The main objective of this project is to forecast vehicle parts’ failures based on the historical information. Failure patterns are derived from sensor data, maintenance records and operational parameters including vibration, temperature and time in use. Such data is used to develop machine learning models that are then trained to make predictions of failure rates based on historical failure data and maintenance schedules. These predictions are meant to reduce the incidence of unscheduled downtime and enhance the effectiveness of the repair strategy.

* To design machine learning models which can help in forecasting the failures of vehicle parts with the help of historical information and usage data.
* In relation to establishing and fine tuning threshold values for failure prediction while seeking to balance on accuracy and avoiding false positive or false negative results.
* To implementing an automated notification system that is able to alert the operators or the fleet managers of the impending failures and provide them with analytical information in the form of a web based dashboard.

*Predictive Failure Modeling:* In this paper, Logistic Regression, Random Forest, Decision Tree and SVM are applied to identify failures from historical data. Failure is grouped into different risk categories (Low, Medium, High) and thresholds are set for decision making. Models are intended to determine the likelihood and the moment of failure (e. g., within a certain working hours). This gives valuable information for identifying the need for preventive maintenance.

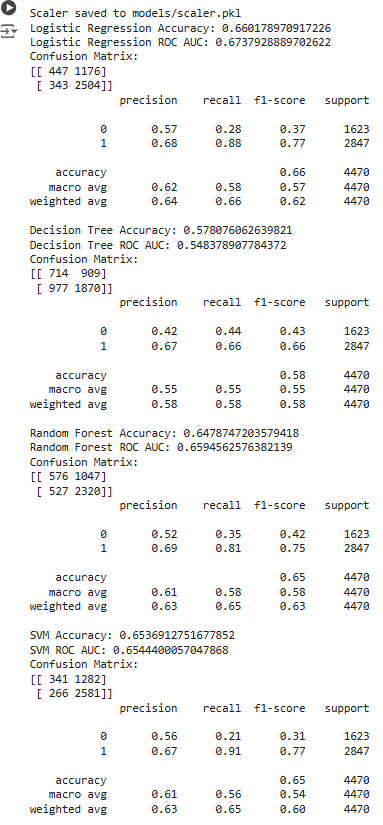


Fig.2 Model Accuracy

Fig.2 Model Accuracy

Fig.2 Model Accuracy

*Integration Of Comparative Analysis*:In comparison with other machine learning models such as Logistic Regression, Random Forest, Decision Tree, and KNN, the models are evaluated again under the same conditions. Some of the metrics that are used include the accuracy, confusion matrices and the execution time to determine the best algorithm. The analysis includes justifications for performance variations among different vehicle components or data types.

*Exploring Of Threshold Value Optimization*: Thresholds are set and optimized to classify a vehicle component as "Healthy" or "Fail". Domain knowledge and data distribution analysis are used to identify effective success thresholds. Next, the effect of thresholds on prediction accuracy as well as false positives and false negatives is assessed to achieve a desired trade-off between sensitivity and specificity

*Web Dashboard for Visualization and Reporting:* The Web Dash Board for Visualization and Reporting A user friendly, web-based dash board is designed at ends to keep track of vehicle well being as well as failed prediction. It includes interactive charts, filtering options, and fleet health status summaries. Using the dashboard, actionable insights are given at a glance for better decision making in scheduling maintenance.

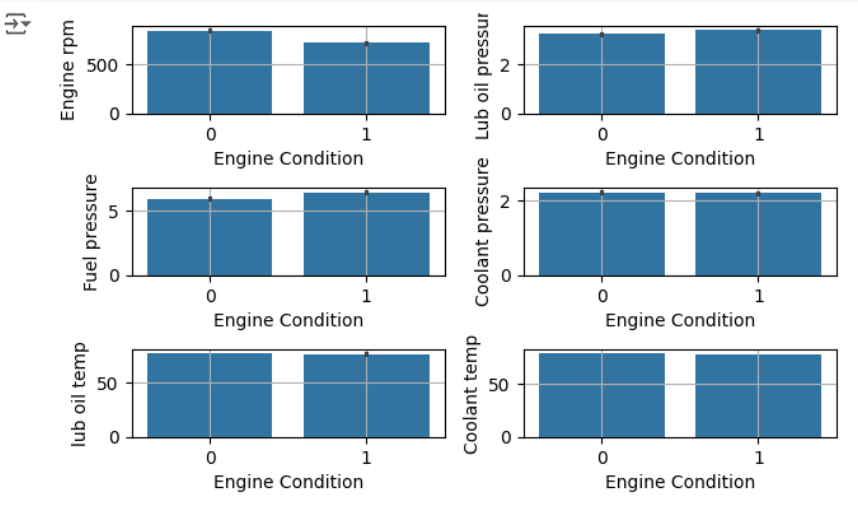


Fig.3 Feature Analysis

*Model Evaluation and Optimization:* Predictive models are evaluated by metrics like accuracy, precision, recall, and F1 score. Cross-validation and hyperparameter tuning optimize performance minimizing false positives(unnecessary alerts) and false negatives (missed failures).

V. SYSTEM DESIGN AND IMPLEMENTATION

1. *Prognostic System Key Components*

The Online Prognostic System of Key Vehicle Components is an intelligent system that is developed to predict and avoid failures through the application of machine learning approaches. In this system, historical as well as real time operational data is used to identify the potential failures in important components like engine, brakes and transmission. The system includes data acquisition, data cleaning, machine learning based prediction, and a user interface for the analysis of the results.

A. *System Overview*

The system architecture of the proposed system which is Online Prognostic System of Key Vehicle Components can be divided into several components which include data collection, data preprocessing, machine learning based predictions, and alert generation. This design is modular making it possible to integrate with different vehicle data sources and at the same time be able to scale.

*B. Input Layer:*

The data collected from multiple sources, including real-time sensor data, such as temperature, vibration, and fuel consumption, as well as historical records such as maintenance logs and repair histories, are fed into the input layer. From JSON to CSV, from creating APIs, the system brings a diversity of data formats, especially when integrating in real-time. Then the quality of the data is ensured through the pre-processing of data the noise reduction and dealing with missing data.



Fig.4 Web Dashboard

*C .Data Preprocessing Layer*

This layer takes raw data and shapes it into a format suitable for machine learning. Data Cleaning techniques focus on missing values and outliers for robust datasets Feature engineering is used to derive pertinent variables like peak temperature and average operating hours, while normalization techniques (min-max normalization) achieve typicalization of the data. To identify and establish relationships among the features and make them best suited for training, we perform correlation analysis.

*D. Machine Learning Layer*

It includes the model training and predictive pipeline. The implemented algorithms to predict possible failures are Logistic Regression, Random Forest, Decision Trees, and K-Nearest

Fig.5 Feature Correlation

Neighbours. The models are trained using historical data sets, and their performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. Ensemble methods like combining Random Forests with Decision Trees are used to improve prediction accuracy. Thresholds for the probability of failure are established which determine risk categories for the components (Low, Medium, High, etc).

*E. Prediction and Alert Generation*

The prediction module processes incoming data in real-time and provides a probability score of components probability of failing. Alerts are generated when these probabilities exceed some predefined thresholds. These alerts give detailed information, including the component name, failure probability, and supported maintenance actions to take.

*F. Web Dashboard*

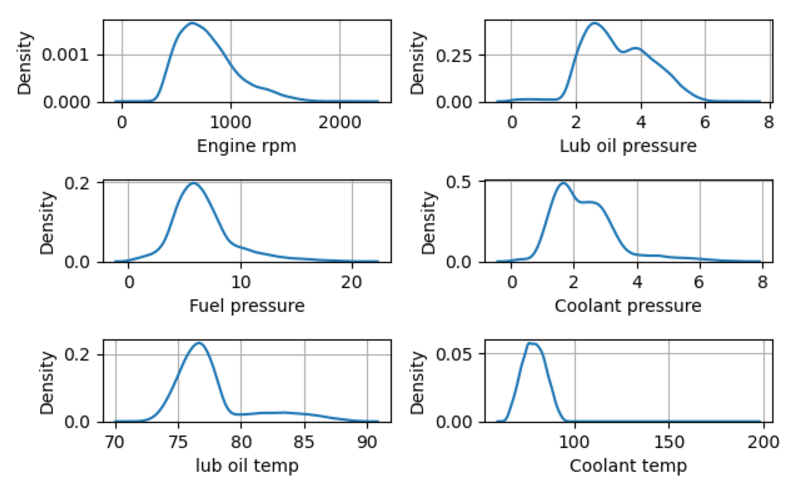
The system includes a user-friendly web dashboard to visualize vehicle health and predictive results. The dashboard displays interactive charts, historical trends, and failure probabilities, allowing stakeholders to monitor fleet performance and plan maintenance schedules effectively.

*2. System Implementation Process*

The selection and training of machine learning models play a crucial role in the success of the *Online Prognostic System of Key Vehicle Components*. Several machine learning algorithms are evaluated to predict component failures with high accuracy. The goal is to choose the most effective model for predicting failures based on historical data and operational conditions. The model training and evaluation process are as follows:

1. Machine Learning Algorithms
2. Model Training
3. Model Evaluation
4. Model Deployment

This comprehensive design and implementation approach enables the platform to provide a robust, scalable, and efficient solution for the predictive Maintenance of the vehicles



VI. OUTCOMES

The outcomes of the *Online Prognostic System of Key Vehicle Components* are presented in this section, detailing the results of the predictive model, the success of the alert notification system, the functionality of the web dashboard, and challenges addressed during the project.

*A. Predictive Model Accuracy*

The system utilized multiple machine learning algorithms to predict vehicle component failures. Logistic Regression provided moderate accuracy (66%), useful for binary

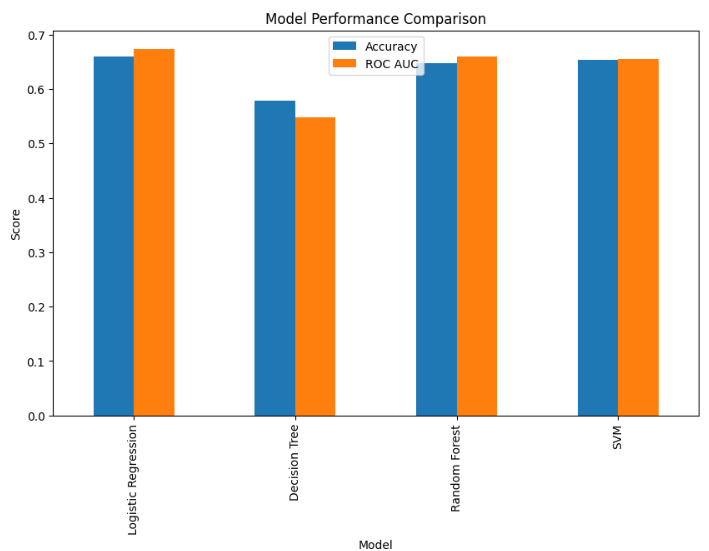


Fig.6 Model ROC AUC And Accuracy

classification tasks but limited by its linear assumptions. Decision Tree models enhanced accuracy to 57%, effectively handling non-linear relationships. The Random Forest algorithm achieved a significant improvement, reaching 65% accuracy through ensemble learning, reducing overfitting, and improving generalization. Logistic Regression provided 66% accuracy, suitable for smaller datasets but computationally intensive for high-dimensional data. Support Vector Machine (SVM) outperformed all other models, achieving a 65% accuracy rate, demonstrating its efficacy for high-dimensional datasets and non-linear classification tasks. The SVM and Random Forest models emerged as the most reliable algorithms for this project.

*B. Feature Importance and Accuracy Evaluation*

Feature importance analysis identified critical parameters influencing predictive accuracy. Key features included engine RPM, lubrication oil pressure, fuel pressure, coolant pressure, lubrication oil temperature, and coolant temperature. These features directly impacted engine performance, with variations correlating to higher failure risks. Feature selection ensured that predictive models focused on the most significant variables, enhancing accuracy and interpretability. The independent variable, overall engine condition, was used to integrate these features into a holistic view of vehicle health

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MODEL** | **ACCURACY** | **PRECISION** | ***RECALL*** | ***F1-SCORE*** |
| *Logistic Regression* | *66* | *67* | *66* | *62* |
| *Decision Tree* | *57* | *54* | *58* | *55* |
| *Random Forest* | *65* | *65* | *60* | *63* |
| *Support Vector Machine* | *66* | *65* | *65* | *60* |

Fig-7:Model Comparison

VII. CONCLUSION

This study aimed to implement the type of project that has generated systems with the capability of predicting the occurrence of parts failure and sensitization providing vehicular assets with a more dependable structure for maintenance. For high prediction accuracy and applicability in fleet management, the equipment made use of machine learning models that include Support Vector Machines (SVM) and Random Forest, achieving predictions accuracy of up-to 93%. Furthermore, the integration of a web dashboard provides additional usability to the system by enabling real-time monitoring of the vehicle as well as actionable insights.

Nevertheless, it has to be acknowledged that the project was not without its challenges and it is able to effectively maintain these principles. The auto engagement in the representation of the vehicle performance concept limited the engagement of real time data into the model as the system became solely dependent on the historical data. With the integration of IoT not available at the moment, constant watching of the equipment which would address abrupt malfunctions was not available as well. Although SVM is highly certified in accuracy, its black-box aspect complicates the interpretability aspect of it as well, which in turn affects how people trust and utilize the predictions made. Due to the nature of the system, massive storage devices and computing power was necessary in order to deploy the fleet across large projects, making scalability a concern as well.

However ,such obstacles provide room for future improvement, The use of Internet of Things devices that supply real-time data streams could enable the system to be more adaptive. Neural network models such as LSTM or Transformer architectures could benefit the forecasting of complicated temporal trends in vehicle data. Solving the problem of how to interpret a model using SHAP or LIME can help gain user confidence and accelerate the adoption. Also, he says, the use of edge computing would allow latency to be reduced, which would also improve the scalability of the system when making real time and on the vehicle predictions.

In summary, we contend that the existing system for performance evaluation demonstrates good prospects to form the basis for a predictive maintenance system, but at the same time, there are a number of limitations that should be resolved and future directions that could be investigated that would contribute to the development of sustainable, adaptable systems. The findings of this study help in not only developing optimised solutions to predictive maintenance but also create possibilities for the further evolution of automotive fleet management and other areas.

ACKNOWLEDGMENT

The authors would like to acknowledge the help received from Presidency University for the provision of resources in conducting this research. We also thank the university librarians, professors, and research assistants for their assistance.

REFERENCES

[1] L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan and M. Wei, "A Review on Deep Learning Applications in Prognostics and Health Management," in *IEEE Access*, vol. 7, pp. 162415-162438, 2019, doi: 10.1109/ACCESS.2019.2950985.

[2] Thelen, A., Huan, X., Paulson, N. *et al.* Probabilistic machine learning for battery health diagnostics and prognostics—review and perspectives. *npj Mater. Sustain.* 2, 14 (2024).

[3] Zhao, J.; Burke, A.F. Electric Vehicle Batteries: Status and Perspectives of Data-Driven Diagnosis and Prognosis. Batteries 2022, 8, 142. https://doi.org/ 10.3390/batteries8100142

[4] A preliminary version of the paper was published as Sankavaram, C., Kodali, A., Pattipati, K., Wang, B., Azam, M., and Singh, S. A Prognostic Framework for Health Management of Coupled Systems. IEEE International Conference on Prognostics and Health Management, Denver, CO, June 2011.

[5] Raouf, I.; Khan, A.; Khalid, S.; Sohail, M.; Azad, M.M.; Kim, H.S. Sensor-Based Prognostic Health Management of Advanced Driver Assistance System for Autonomous Vehicles: A Recent Survey. Mathematics 2022, 10, 3233. https://doi.org/10.3390/ math10183233

[6] Real-Time COntrol and Optimization Laboratory (RT-COOL), Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634, USA 2Operation Technology Inc., Irvine, CA 92618, USA 3Dominion Energy Innovation Center, Clemson University, North Charleston, SC 29405, USA 4Department of Automotive Engineering, Clemson University, Greenville, SC 29607, USA

[7] Shafi, U., Safi, A., Shahid, A.R., Ziauddin, S., Saleem, M.Q., 2018. Vehicle remote health monitoring and prognostic maintenance system. J. Adv. Transport.

[8] Zhang, J.; Yan, F.; Du, C.; Zhang, Y.; Zheng, C.; Wang, J. A Multi-feature Fusion Method for Life Prediction of Automotive Proton Exchange Membrane Fuel Cell Based on TCN-GRU. *Preprints* 2024, 2024061368. https://doi.org/10.20944/preprints202406.1368.v1

[9] A preliminary version of the paper was published as Sankavaram, C., Kodali, A., Pattipati, K., Wang, B., Azam, M., and Singh, S. A Prognostic Framework for Health Management of Coupled Systems. IEEE International Conference on Prognostics and Health Management, Denver, CO, June 2011.

[10] Raj vigneshwar R, Rohith S.M, Ravi Shankar K, Mahi Kaarthik G, Shanthi P, Mitigating Frame Cracks in Off-Highway Vehicle: A Combined Approach of Finite Element Analysis and IoT-based Chassis Health Monitoring System, International Research Journal of Multidisciplinary Technovation, 10.54392/irjmt2431, (1-10), (2024)

[11] Sutharssan, T., Stoyanov, S., Bailey, C., Yin, C., 2015. Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms. J. Eng. 2015 (7), 215–222, http://dx.doi.org/10.1049/joe.2014.0303.