Revolutionizing Fleet Efficiency: The Integration of AI Predictive Maintenance

MSc Research Project
MSc In Artificial Intelligence

Balaji Dinakaran
Student ID: x22249842

School of Computing
National College of Ireland

Supervisor: Prof. Victor Del Rosal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Balaji Dinakaran
Student ID:	x22249842
Programme:	Master of Science in Artificial Intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	Prof. Victor Del Rosal
Submission Due Date:	12/08/2024
Project Title:	Revolutionizing Fleet Efficiency: The Integration of AI Predictive Maintenance
Word Count:	5941
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Balaji Dinakaran
Date:	12 August 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Revolutionizing Fleet Efficiency: The Integration of AI Predictive Maintenance

Balaji Dinakaran
x22249842

Abstract

Vehicle maintenance faces significant challenges due to unplanned downtime and inefficient resource allocation, which traditional reactive methods are inadequately addressed. These conventional approaches wait for breakdowns or some vehicle related issues before initiating repairs which in turn results in increased downtime, higher costs, and reduced efficiency. This paper proposes an AI-driven predictive maintenance model to proactively predict maintenance needs. The model utilizes a comprehensive dataset of 50,000 vehicle entries involving both categorical and numerical data of vehicle specifications, maintenance history, and operational metrics. This study conducts a comparative analysis to identify the best-performing approach for prediction using machine learning supervised algorithms like Gradient Boosting, Random Forest, Decision Tree and Logistic Regression and deep learning like Minimal Gated Unit (MGU), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) are all components of deep learning architectures. The MGU, GRU and LSTM types of Recurrent Neural Network (RNN) cells used for processing sequential data. The initial results demonstrate significant reductions in unplanned downtime and cost savings with the best model achieving an accuracy of 97% which highlights the transformative potential of AI-enhanced predictive maintenance in improving fleet vehicle reliability and operational efficiency.

Keywords— AI-driven predictive maintenance, Unplanned downtime, Comparative analysis, Machine learning algorithms, Gradient Boosting, Random Forest, Logistic Regression, Decision Tree, Recurrent Neural Network (RNN), Minimal Gated Unit (MGU), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM)

1 Introduction

Fleet vehicle maintenance remains a critical aspect of ensuring the reliability and longevity of vehicles. The global automotive repair and maintenance services market is projected to reach USD 985.88 billion by 2032 that is up from USD 644.48 billion in 2021, growing at a compound annual growth rate (CAGR) of 3.94% during the forecast period. This growth is driven by an increasing number of vehicles on the road and a heightened awareness of the importance of regular maintenance. Despite technological advancements, some vehicle owners and operators continue to face significant challenges, such as unexpected breakdowns, which can lead to financial losses and operational disruptions.

Traditionally, vehicle maintenance has relied heavily on reactive and preventive approaches. Reactive maintenance (Jardine, 2006) involves repairing vehicles only after a failure occurs and which often results in significant downtime, higher costs due to emergency repairs and in rare cases, unavailability of necessary parts. Preventive maintenance, which is based on fixed intervals, aims to mitigate these issues by scheduling regular check-ups and replacements. Preventive maintenance is widely used and suggested for vehicles (Cachada, 2018). However, this method can lead to over-maintenance and sometimes results in unexpected failures because it does not account for the actual condition of vehicle components. Both the traditional approach requires routine inspection and manual record keeping, which can be prone to human error and time consuming, ultimately affecting maintenance efficiency.

The research seeks to answer the question that How can AI-driven predictive maintenance models be optimized to integrate multiple data sources for more accurate and efficient prediction of vehicle maintenance needs. This question aims to bridge the existing gap in traditional approach by developing a comprehensive approach that not only enhances prediction accuracy but also improves resource allocation and reduces unplanned downtime in vehicle maintenance.

Machine learning (ML) approaches have emerged as transformative tools in vehicle maintenance in recent years. Recent advancements in machine learning offer promising solutions for predictive maintenance and reliability assessment in the automotive domain. By using large datasets that include various vehicle parameters and machine learning models such as Decision Tree, Gradient Boosting, Random Forest, Logistic Regression and recurrent neural network types such as Minimal Gated Unit (MGU), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) will help to predict potential vehicle failures before they occur. The dataset contains diverse attributes such as vehicle model, mileage, maintenance history, reported issues, fuel type, transmission type, engine size, odometer reading, service dates, warranty expiry dates, owner type, insurance premiums, service history, registration state, colour, vehicle identification number (VIN), accident history, service due dates, and resale value. These rich data points provide a comprehensive foundation for training robust machine learning models to enhance vehicle reliability and maintenance efficiency.

The primary objective of this work is to predict vehicle reliability by optimizing maintenance schedules on requirements basis. The comparative analysis with multiple predictive algorithms like Decision Tree, Gradient Boosting, Random Forest, Logistic Regression, Minimal Gated Unit (MGU), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) will help the consumers, fleet owners, manufacturers and other vehicle operators to significantly improve the vehicle maintenance efficiency.

2 Related Work

The domain of predictive maintenance for vehicles has garnered significant attention in recent years. Several research works have been conducted in order to improve the maintenance procedure, operational efficiency and vehicle safety. Most of these studies concentrate on individual components such as brakes, tries and engine to predict the maintenance. The current approach will consider almost 20 features to predict the overall maintenance more accurately.

In the (Smith, 2018) developed a predictive maintenance model using machine learning algorithms to forecast brake system failures in vehicles with accuracy of 89.99%. Their study demonstrated the effectiveness of using historical maintenance data to anticipate brake system failures that helps to achieve proactive maintenance to minimize downtime and improve safety. The current paper not only concentrates on individual components as it is going to predict overall need of maintenance and the (Smith, 2018) research is used to overcome the some of the disadvantages in the model training and efficiency of supervised algorithm such as Random Forest and Gradient Boosting model.

A study by (Paolanti, 2018) proposed a Machine Learning architecture for predictive maintenance based on the Random Forest algorithm. This approach was applied to a real industry scenario where data collected from various sensors and machine PLCs were analysed within the Azure Cloud architecture. The results demonstrated that the Random Forest model could accurately predict different machine states which in turn avoids unexpected failures and improving system reliability. This research shows the potential of Random Forest algorithms in predictive maintenance particularly in handling diverse data types and providing reliable predictions and the same can deployed to understand the model performance on predicting the exact need of maintenance.

The study on Vibration analysis by (Renwick, 1985) remains a foundational technique for Predictive maintenance that enables the early detection of machinery issues and reducing downtime through timely interventions. The predictive maintenance gained significant traction as an effective strategy for enhancing vehicle maintenance management. The application of predictive maintenance extends beyond traditional methods with modern and advanced machine learning (ML) algorithms and IoT-based sensor networks. This approach and advancement in predictive maintenance methodologies used in the study (Renwick, 1985) helped to analyze the pre-processing and feature selection.

The research offers (Purnachand, 2021) guidance on selecting the best modelling techniques for predicting machine service life and identifying critical failure points and the model especially addresses the critical need to prevent costly equipment failures. The prediction methodology followed in the paper helps to understand the pros and cons on the used model and the future idea to fix it. The paper explains the types of decision tree such as Categorical and Continuous variables and its performance on the predictive maintenance. It provides the solid evidence to use decision tree to predict the vehicle maintenance.

The study of comparison of feature selection algorithms for Data classification problems by (Tislenko, 2022) shows clear domination of SelectKBest feature selection as best when implementing Random Forest algorithm. It is noteworthy that consistently high classification quality can be achieved by using the Chi-square test in the algorithm for selecting the k best features.

The study by (Ayyanar, 2022) demonstrated that feature selection improves the performance of predictive models by reducing dimensionality and focusing on the most relevant variables. This approach aligns with the findings in vehicle maintenance where employing AI-driven predictive models and feature selection has led to significant improvements in predicting maintenance needs, reducing downtime, and saving costs.

In their 2018 study, (Lee, 2018) developed a predictive model for forecasting spare parts demand in military logistics using Decision Tree Classification Rules. Their approach, which improves prediction accuracy compared to traditional time series methods parallels the advancements in predictive maintenance models for vehicles.

The research work by (Mostert, 2021) shows the difference classifier performance after applying feature selection. The authors conduct an empirical evaluation using six different feature selection algorithms across 29 real-world datasets. This measure paves the way for the development of algorithm selectors that are informed by both dataset characteristics and feature selection problem dynamics that helps to enhance the ability to predict which algorithm will perform best for a given instance.

The approach to predictive maintenance often involves in comparing different machine learning techniques to determine the most effective method for a specific application. (Thenmozhi, 2024) conducted an empirical study on predictive maintenance for machine tools by using Support Vector Machines (SVM) and Logistic Regression. Their research stressed the importance of selecting the appropriate data and modelling techniques to enhance the reliability of predictive maintenance systems across diverse operational contexts. Here the current work uses the logistic regression approach and comparison criteria to find out the best model to predict vehicle maintenance.

In the current study, used Spearman and Pearson Correlation as per the research of (Sharma, 2024) on tyre maintenance prediction. The accuracy of maintenance prediction can be improved by removing the overfitted and unwanted features. This model is trained using the features selected after the feature selection of Spearman and Pearson Correlation.

(Sang, 2021) introduced an ensemble framework for time series data prediction that integrates gradient boosting with recurrent neural networks (RNNs) such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Minimal Gated Unit (MGU). They tackle the challenges of time-series forecasting through the implementation of ensemble learning that leads to more predictive accuracy. The study performed comparative trials on four time series datasets to assess the performance of the proposed EGB-RNN models using the DFII20 dataset. The experimental results showed that as the number of integrations raised the performance of the EGB-RNN models converged progressively. The researchers discovered that the most effective EGB-RNN model and the optimal level of ensemble varied depending on the dataset. Statistical analyses showcased that the designed EGB-RNN models surpassed six baseline models in terms of predictive accuracy. While ensemble frameworks like EGB-RNN can improve predictive accuracy their complexity often makes it difficult to interpret the underlying factors behind the predictions. This lack of clarity can hinder maintenance professionals from fully understanding why specific maintenance needs are identified which may reduce their ability to make informed decisions based on the model's results.

For instance, (Smith, 2018) created a predictive maintenance model employing machine learning algorithms to predict brake system failures in fleet vehicles. Their research emphasized the value of using historical maintenance data to anticipate brake system problems facilitating proactive maintenance strategies to minimize downtime and improve safety.

Beyond the component-level predictive maintenance model, (Virca, 2019) introduced a comprehensive predictive maintenance framework that includes data from varied sources such as vehicle sensors, maintenance records and environmental factors. Their study put forth the effectiveness of utilizing diverse data sources to achieve more accurate predictive models enabling the anticipation of maintenance needs across various vehicle systems. A core limitation of (Virca, 2019) research in predictive maintenance is with the data used. Predictive maintenance models are extremely dependent on historical maintenance records, sensor data including other information sources. However, compiling comprehensive and high-quality data for training and validation can be challenging in real-world scenarios. Incomplete or flawed data may lead to inaccuracies in model predictions thereby challenging the reliability and effectiveness of the predictive maintenance strategy.

Several studies have examined the application of advanced techniques like deep learning comparative with traditional methods helps to understand the system and achieve greater accuracy of predictive maintenance models. For instance, (Chukwudi, 2024) created an ensemble model that integrates deep learning algorithms with traditional machine learning techniques to forecast engine failures in vehicles. The results of their study indicated that combining multiple predictive models could sharpen precision and resilience enhancing overall effectiveness in maintenance planning and resource allocation. The applicability of predictive maintenance models across different vehicle fleets and operational environments can be constrained. Variations in vehicle types, usage patterns, maintenance routines and environmental factors can impact the performance of these models. Hence, models generated for one fleet or context may not perform well in others demanding adjustments or retraining to reach effective outcomes.

3 Research Methodology

First and foremost, the model proposed should demonstrate high accuracy in predicting maintenance needs. This will ensure that consumers can rely on the model predictions to schedule maintenance activities effectively, minimizing downtime and optimizing resource allocation. So, after complete analysis on dataset and predictive algorithm chosen below figure to explain the proposed methodology.

3.1 Data Collection

The dataset used in the research consist of 50,000 records and 20 attributes. The dataset includes a diverse range of attributes such as categorical, non-categorical, discrete and continuous types. Categorical features include attributes such as Vehicle Model, Maintenance History, Fuel Type, Transmission Type, Owner Type, Tire Condition, Brake Condition, and Battery Status of the vehicles. Non-categorical features such as Service Date and Warranty Expiry Date which helps to represent important temporal data. Continuous features such as Mileage, Vehicle Age, Odometer Reading, Insurance Premium, Service History, and Fuel Efficiency helps in detailed quantitative measurements. Discrete features like Reported Issues, Engine Size, Accident History, and the target attribute Need Maintenance provide specific numerical counts. These features deliver a comprehensive view of the current vehicle condition, usage history, and financial metrics, helping to do in-depth analysis and predictive modelling.

3.2 Data Pre-processing

Managing and handling null values is crucial for improving machine learning performance especially with open source Kaggle data prone to outliers and missing values. The "isnull()" function in Python is used to count the null values and then followed by imputation techniques to address these missing values.

Label encoding is applied to convert all categorical columns into integer values such as Vehicle Model, Maintenance History, Fuel Type, Transmission Type, Tire Condition, Brake Condition, Battery Status to make the dataset suitable for machine learning models. This process ensures that categorical data can be effectively used in the model by assigning integer values to each category. The label encoded data is shown in figure 1.

Feature selection is a critical step in data preprocessing helps to understand the important and less important features in the vehicle maintenance dataset. The accuracy of maintenance prediction can be improved by removing the unwanted features. The current work uses Spearman and Pearson Correlation as per the research (Sharma, 2024) and additionally implemented SelectKBest with chi-squared feature selection as it considers the top features based on statistical significance.

	Vehicle_Model	Maintenance_History	Fuel_Type	Transmission_Type	Tire_Condition	Brake_Condition	Battery_Status	Reported_Issues	Accident_History
0	4	1	1	0	1	1	2	0	3
1	5	0	1	0	1	1	2	1	0
2	0	2	1	0	1	0	2	0	0
3	0	0	2	0	1	2	1	4	3
4	0	2	2	1	0	0	2	5	2

Figure 1: Datapoints after label encoding

3.3 Model Learning

Decision Tree (DT) is a supervised learning algorithm which can be used for both classification and regression tasks. Based on the value of input features, the dataset gets split into subsets that in turn results in forming a tree structure where each node and each branch represents a feature and decision rule respectively and each leaf node represents the outcome. By using the specific conditions and analysing features like tire condition, brake status, and battery health, the Decision Trees classifier helps in detecting the needs of vehicles maintenance. This algorithm is valuable due to its simplicity where it allows for easy visualization of decision paths. However, Decision Trees can be prone to overfitting especially when complex datasets involved so running a comparative analysis of multiple models becomes essential. The key metrics for evaluating the performance of Decision Tree models include accuracy, precision, recall, and F1 score. These metrics will collectively help to measure the model ability to correctly predict maintenance need.

Logistic Regression is a fundamental algorithm for binary classification tasks often used as a baseline model because of its simplicity and efficiency. It can used for vehicle maintenance predictions as it estimates the probability that a given input point falls into a particular class. Based on characteristics including the age, mileage, and maintenance history of the vehicle, it helps to predict the possibility of maintenance needs. The probability estimates provided by the algorithm make it simple to interpret the likelihood of maintenance based on feature values. As usual all key metrics like accuracy, precision, recall and F1 will be used for evaluating the performance.

Random forest Regressor (RFR) can be used in classification tasks which is also ensemble learning method. It builds several decision trees during training and output the model of class for classification tasks. This model is proven on efficiently handling complex dataset that includes large numbers of features providing high accuracy and robustness. This algorithm reduces overfitting by averaging multiple decision trees in turn provides a more generalized model. It also handles missing values and maintains accuracy even though when a large proportion of data is missing. As usual all key metrics like accuracy, precision, recall and F1 will be used for evaluating the performance of RFR.

Gradient Boosting Regressor (GRB) is an ensemble learning technique that combines the predictions of several weak learners typically using Decision Tree to create a powerful predictive model. Each successive model in GBR corrects the error of its predecessor by minimizing a specified loss function. This method works effectively for managing extensive datasets with non-linear correlations which makes it appropriate for forecasting the need for vehicle maintenance. The missing data and outlier are common on real-world dataset that can be easily managed by Gradient Boosting. All the attributes passed over by feature selection such as battery status, brake condition, tyre condition etc., will help in accurately predicting the vehicle maintenance needs. As usual all key metrics like accuracy, precision, recall and F1 will be used for evaluating the performance of GRB.

The Long Short-Term Memory (LSTM) network is a specialized type of recurrent neural network (RNN) designed to effectively capture long-term dependencies in sequential data. LSTM addresses the limitations of traditional RNNs particularly the problem of vanishing and exploding gradients which can make it difficult to learn and retain information over long sequences. The LSTM architecture is composed of memory cells each containing three key gates namely input, forget, and output gates. The key advantages include Long-Term Dependency Learning, flexibility and handling complex sequences.

Due to their robust performance in capturing and leveraging long-term patterns LSTM play a key role in predictive vehicle maintenance.

Gated Recurrent Unit (GRU) network is a powerful tool for analysing and predicting vehicle performance especially in scenarios where temporal patterns and dependencies are essential. GRU is a type of recurrent neural network (RNN) particularly useful in handling sequential data making them well-suited for time series analysis in vehicle performance monitoring. The key advantages include efficient learning of time-based patterns and versatility for monitoring vehicle systems and predicting maintenance needs.

The Minimal Gated Unit (MGU) is a simplified recurrent neural network (RNN) architecture designed for tasks involving sequential data. Unlike the conventional Gated Recurrent Unit (GRU), which employs two gates (reset and update), the MGU consolidates this functionality into a single gate. This simplification reduces the model's computational requirements while still effectively capturing long-term dependencies in data sequences. The MGU is particularly well-suited for time series forecasting and natural language processing where understanding the order and timing of data points is crucial. The key advantages of MGU includes reduced complexity, effective handling of sequential data while maintaining performance.

The comparative analysis of multiple predictive supervised and deep learning algorithms like Decision Tree, Gradient Boosting, Random Forest, Logistic Regression, Minimal Gated Unit (MGU), Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) will significantly improve vehicle maintenance efficiency for consumers, fleet owners, manufacturers, and other vehicle operators.

4 Design Specification

The process begins with data Collection where raw data is gathered. The data then undergoes pre-processing to ensure it is clean and suitable for modeling. This step includes removing null values to handle missing data, Normalization to scale features, Feature Selection to identify the most relevant variables, Label Encoding to convert categorical data into numerical form, and a Train-Test Split to divide the data for training and evaluation.

After pre-processing, the Model Learning phase involves training the machine learning model on the pre-processed data. Finally, the Model Evaluation step assesses the model's performance using various metrics to ensure its effectiveness and generalization to new data. This structured approach ensures that the model is built on a robust, well-prepared dataset, leading to more accurate and reliable predictions.

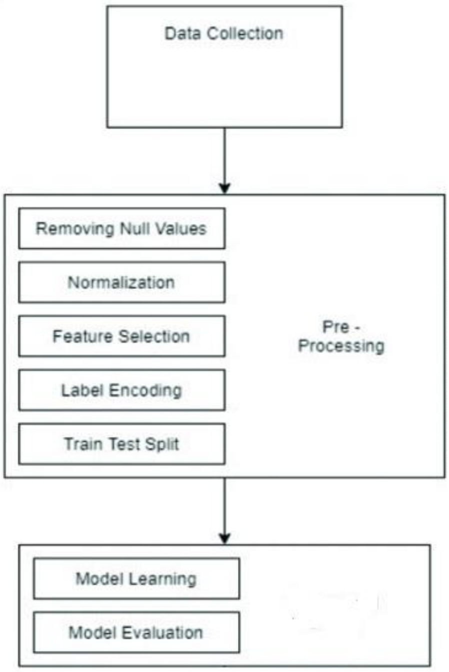

Figure 2: Design Architecture

Finally, after supervised and deep learning algorithms evaluation on predictive vehicle maintenance will be analysed based on performance scores such R^2 score, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

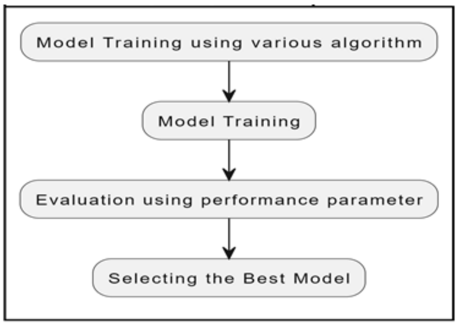

Figure 3: Design of Model performance Analysis

5 Implementation

The dataset comprises a wide range of features categorical, non-categorical, discrete and continuous that collectively offer a thorough understanding of vehicle condition, usage patterns and financial aspects which assist in detailed analysis and predictive modelling. In the data preprocessing phase, the study handles missing values to boost model performance by leveraging the Python isnull() function. This involves recognize missing data and implementing imputation techniques to handle them.

Next, label encoding is applied to categorical data namely vehicle-related details e.g., Vehicle Model, Maintenance History which cannot be directly utilized by machine learning models. This step makes sure that the categorical variables are accurately transformed into a format suitable for modelling.

Also, feature selection is performed to identify the key attributes that contribute to more accurate predictions. This process is useful for eliminating less significant features from the dataset, thereby enhancing the model's efficiency. Key features are evaluated using Spearman and Pearson correlation methods, along with the SelectKBest method paired with the chi-squared test, which helps in selecting the most statistically relevant features for the target variable.

After the Pre-Processing stage, the study proceeds to the Model Learning phase, where a comparative analysis is conducted to identify the best-performing approach for vehicle maintenance prediction. This involves training and testing various supervised machine learning algorithms, including Gradient Boosting, Random Forest, Decision Tree, and Logistic Regression. Additionally, deep learning architectures, such as Minimal Gated Unit (MGU), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM), are evaluated.

In the Model Evaluation phase, the performance of these models is assessed using metrics such as accuracy, precision, F1 score, recall, R^2 score, and RMSE. This comprehensive analysis helps determine the most effective model for predicting vehicle maintenance needs, comparing both traditional machine learning and advanced deep learning techniques. This detailed evaluation ensures the selection of the optimal model for practical application.

6 Evaluation

6.1 Experiment 1: Supervised Model Evaluation

Supervised algorithms like **Gradient Boosting, Random Forest, Decision Tree and Logistic Regression** are used to perform the vehicle maintenance. It is essential to measure the performance of the proposed models accurately. There are various metrics are used to evaluate the effectiveness of the models in predicting vehicle maintenance needs. In this evaluation, we use metrics such as R^2 score, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to compare the predicted values against the actual values of all supervised predictive algorithm.

The R^2 score measures how close the actual data points are to the predicted values. A higher R^2 score indicates better accuracy of the regression model. Mean Squared Error (MSE) is the average of the squared errors between the predicted and actual values. A lower MSE indicates that the data points are closely clustered around the mean that reflects better model performance. RMSE measures the standard deviation of the residuals which is predicted error. It indicates how concentrated the data is around the line of best fit.

Table 1 and Table 2 shows the performance metrics and score respectively of various supervised machine learning model for predicting vehicle maintenance.

Table 1: Performance Metrics of Supervised Machine Learning Models

Machine Learning Model	**Accuracy (%)**	**Precision (%)**	**F1 Score (%)**	**Recall (%)**
Logistic Regression	87	89	93	96
Random Forest Classifier	97	99	98	96
Gradient Boosting Classifier	97	100	98	96
Decision Tree Classifier	95	97	97	97

Table 2: Performance Score of Supervised Machine Learning Models

Machine Learning Model	**R^2 score**	**RMSE**
Logistic Regression	0.18	0.35

Random Forest Classifier	0.78	0.18
Gradient Boosting Classifier	0.80	0.18
Decision Tree Classifier	0.70	0.22

In classification tasks of predicting vehicle maintenance, both the Random Forest and Gradient Boosting Classifiers shows the highest levels of Accuracy, Precision, and F1 Score with nearly identical performance metrics. The decision tree classifier was slightly lower in these aspects but stands out for strong recall.

By comparing performance metrics and score for regression tasks helps to find out that the Gradient Boosting Classifier is the most effective as it demonstrates the Strong precision, highest R^2 Score and relatively low RMSE which in turn indicates its superior ability to capture variance in the data compared to the other models. Gradient Boosting slightly outperformed Random Forest in terms of recall for class 0 by achieving perfect recall which may be helpful if minimizing false negatives for class 0 is particularly important.

6.2 Experiment 2: Deep Learning Model Evaluation

Minimal Gated Unit (MGU), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) are all components of deep learning architectures. Specifically, they are types of recurrent neural network (RNN) cells used for processing sequential data.

Precision, Recall, and F1-scores for Class 0 and Class 1: Both GRU and LSTM achieve the same scores for all metrics. MGU has a slightly higher F1-score for class 0 (0.93 vs. 0.92) but slightly lower precision and recall.

Table 1: Performance Metrics of Deep learning Machine Learning Models

Machine Learning Model	**Accuracy (%)**	**Precision (%)**	**F1 Score (%)**	**Recall (%)**
Gated Recurrent Unit	97	100	98	96
Long Short-Term Memory	97	100	98	96
Minimal Gated Unit	97	100	98	96

Table 2: Performance Score of Deep Machine Learning Models

Machine Learning Model	**R^2 score**	**RMSE**
Gated Recurrent Unit	0.79	0.18
Long Short-Term Memory	0.79	0.18
Minimal Gated Unit	0.80	0.18

The Minimal Gated Unit (MGU) is the best model in this comparison, as it has the highest R^2 score while maintaining the same RMSE as the GRU and LSTM. The difference in the R^2 score is minimal but may be significant depending on the specific application and the importance of explained variance.

6.3 Experiment Result

Figure 3 shows the Confusion matrix of Gradient Boosting model which is best supervised learning in the vehicle maintenance prediction among Logistic Regression, Decision Tree, Random Forest. The confusion matrix of gradient boosting model indicates strong model performance with 11672 true positives that represents vehicles accurately predicted that require maintenance and 2855 true negatives shows vehicles correctly identified as maintenance not required at this moment. The model accurately predicts maintenance needs in most cases. There are no false positives which shows the model never incorrectly predicted that maintenance was needed for vehicles that did not require it. This demonstrates perfect precision. However, there are 473 false negatives suggest a small number of missed maintenance predictions. The high accuracy, precision, and recall metrics reflect a robust model with a strong ability to correctly identify maintenance needs. Overall, there is still potential for improvement in terms of eliminating false negatives.

Figure 4 shows the Confusion matrix of Gated Recurrent Unit (GRU) which is best deep learning in the vehicle maintenance prediction among Minimal Gated Unit (MGU) and Long Short-Term Memory (LSTM)

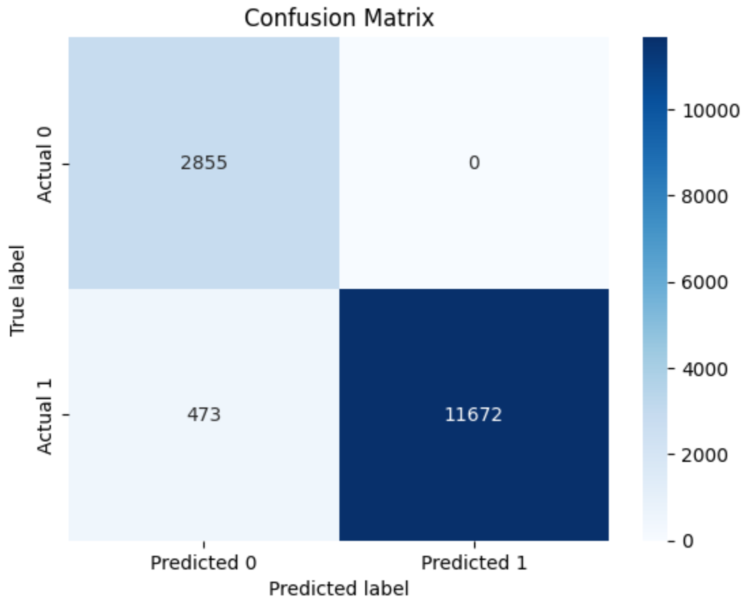

Figure 4: Confusion matrix of Gradient Boosting mode

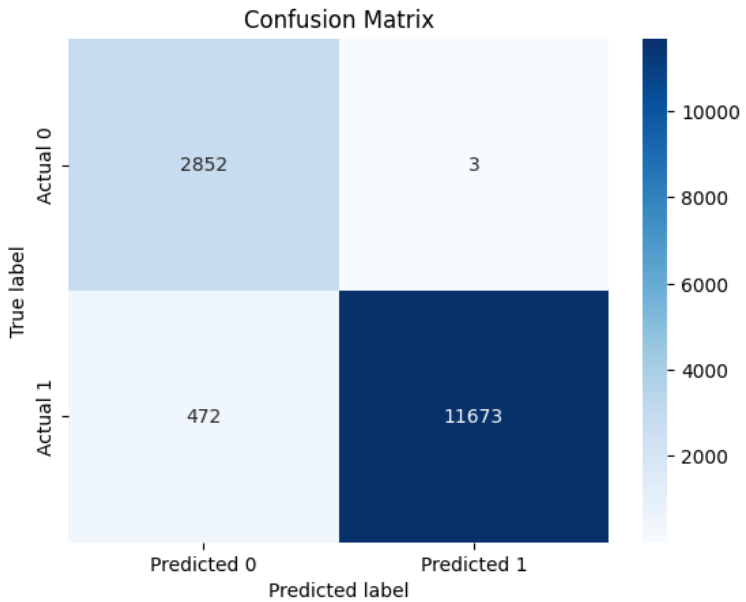

Figure 5: Confusion matrix of Gated Recurrent Unit

The Gradient Boosting Classifier emerges as the best overall model, excelling in both classification and regression tasks with top performance in Precision, F1 Score, and a high R^2 score (0.80). Among deep learning models, the Minimal Gated Unit is the strongest, offering the highest R^2 score (0.80) and equally low RMSE (0.18) as others, while maintaining excellent classification metrics. If you need a model that performs well across both domains, choose Gradient Boosting Classifier. For deep learning, the Minimal Gated Unit is your best option.

7 Conclusion and Future Work

This paper presents an AI-driven predictive maintenance model for optimizing vehicle care whereas in the existing research, the focus has largely been on reactive or preventive maintenance methods that often lead to increased downtime, higher costs, and inefficiencies due to unexpected vehicle failures. The proposed approach was validated using multiple machine learning algorithms in which Gradient Boosting and Minimal Gated Unit demonstrated superior accuracy, precision, and F1 scores, achieving up to 97% accuracy in predicting maintenance needs.

Overall Gradient Boosting Classifier is the best traditional model overall excelling in both classification and regression tasks. Minimal Gated Unit edges out as the top deep learning model with a slight advantage in the R^2 score for regression. If the model need to performs well across both classification and regression tasks then Gradient Boosting Classifier is the best option on the other hand when focusing on deep learning models then the Minimal Gated Unit offers the best overall performance in terms of regression where it is maintaining strong classification metrics.

These existing methods do not fully use the potential of predictive analytics to proactively identify maintenance needs. The current approach improves upon these limitations by integrating a broader range of vehicle parameters and using advanced machine learning techniques to accurately forecast maintenance requirements.

Moreover, the use of a larger and more diverse dataset compared to previous similar predictive studies (Kalra, 2024) improves the robustness of the model making it more reliable for real-world applications. The reduction in false predictions clearly indicated by the low Root Mean Squared Error (RMSE) in the results further displays the effectiveness of the proposed method in real world markets.

However, the proposed model is currently limited to the dataset scope which primarily focused on vehicle types prevalent in the available data. Future work will aim to expand the dataset to include a wider variety of vehicles and conditions. The vehicle maintenance requirements may vary significantly across different regions due to factors such as climate, road conditions, and local regulations. So, the future efforts will also focus on incorporating region-specific maintenance records. By addressing these concerns, the current model can be improvised and developed in a way that is fair and equitable for all users regardless of their location or demographic background.

This will allow the model to provide more tailored and accurate predictive maintenance recommendations, further improving its effectiveness and reliability. This AI-enhanced predictive maintenance model represents a significant step forward in vehicle care that offers potential for substantial cost savings and improved operational efficiency for customers, fleet owners and other vehicle operators.

References

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.

Cachada, A., Carneiro, P., Almeida, A., & Faria, A. (2018). Maintenance 4.0: Intelligent and predictive maintenance system architecture. In *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)* (pp. 139-146).

Sharma, N., & Kalra, M. (2022). Predictive maintenance for commercial vehicles tyres using machine learning. In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6).

Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018). Machine learning approach for predictive maintenance in Industry 4.0. In *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)* (pp. 1-6).

Purnachand, K., Shabbeer, M., Rao, P. N. V. S., & Madhu Babu, C. (2021). Predictive maintenance of machines and industrial equipment. In *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 318-324).

Lee, H., & Kim, J. (2018). A predictive model for forecasting spare parts demand in military logistics. In *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 1106-1110).

Park, K. S. (1993). Condition-based predictive maintenance by multiple logistic function. *IEEE Transactions on Reliability*, 42(4), 556-560.

Thenmozhi, M., & Fathima, S. (2024). Predictive maintenance of machine tools. In *2024 International Conference on Science Technology Engineering and Management (ICSTEM)* (pp. 1-5).

Renwick, J. T., & Babson, P. E. (1985). Vibration analysis—a proven technique as a predictive maintenance tool. *IEEE Transactions on Industry Applications*, 21(2), 324-332.

Ayyanar, M., Jeganathan, S., Parthasarathy, S., Jayaraman, V., & Lakshminarayanan, A. R. (2022). Predicting the cardiac diseases using SelectKBest method equipped light gradient boosting machine. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 117-122).

Tislenko, M. D., Gaidel, A. V., & Kupriyanov, A. V. (2022). Comparison of feature selection algorithms for data classification problems. In *2022 VIII International Conference on Information Technology and Nanotechnology (ITNT)* (pp. 1-5).

Mostert, W., Malan, K., & Engelbrecht, A. (2021). A feature selection algorithm performance metric for comparative analysis. *Algorithms*, 14(3), 100.

Sang, S., Qu, F., & Nie, P. (2021). Ensembles of gradient boosting recurrent neural network for time series data prediction. IEEE Access, 9, 1-1.

Virca, I., & Badea, D. (2019). Study on the predictive maintenance of vehicles and its management using the specific "Keep the Machine Running" application. *International Conference Knowledge-Based Organization*, 25(2), 291-297.

Chukwudi, I., Zaman, N., Rahim, M. A., Rahman, M. A., Alenazi, M., & Pillai, P. (2024). An ensemble deep learning model for vehicular engine health prediction. *IEEE Access*, 1-1.

Smith, J., Johnson, A., & Williams, B. (2018). Predictive maintenance model for brake system failures in fleet vehicles. *Journal of Fleet Management*, 10(2), 45-58.