Trends in
Machine Design

ISSN: 2455-3352
Volume 13, Issue 2, 2026
May–August
DOI (Journal): 10.37591/TMD

https://journals.stmjournals.com/tmd

Research TMD

Data-Driven Machine Learning Approach for Vehicle Fuel Economy Prediction and Performance Monitoring Using Real-World OBD Data

Arumugam Palanichamy[1], Suresh Alex Selvaraj[2], Rajavel Rangasamy[3], Karthikeyan Subramanian[4,*]

Abstract

Modern passenger vehicles generate large volumes of operational data through On-Board Diagnostics (OBD) systems, enabling continuous observation of vehicle performance under real-world driving conditions. However, much of the existing research mainly analyses previously recorded data and does not provide predictive mechanisms for monitoring vehicle performance under dynamically varying operating conditions. This study presents an AI and machine learning–based method for predicting and monitoring real-world vehicle performance and fuel economy using OBD data collected from a gasoline-powered passenger car during normal driving operation. Key vehicle parameters including vehicle speed, engine speed, throttle position, engine load, and fuel consumption were acquired through the OBD interface and processed using a structured data analytics pipeline. Machine learning techniques including K-Means clustering, K-Nearest Neighbour (KNN) classification, and Random Forest regression were employed to identify representative operating modes, classify driving behaviour patterns, and predict expected fuel economy under varying driving conditions. The developed predictive method enables continuous comparison between predicted and observed vehicle performance, facilitating identification of operational deviations across different driving environments. Analysis of real-world driving data demonstrates the ability of the proposed approach to characterise vehicle operational patterns and support intelligent monitoring of vehicle performance and fuel economy. The study contributes a data-driven predictive monitoring methodology for real-world vehicle analytics, providing a foundation for AI-enabled vehicle performance assessment using OBD big data.

Keywords: Machine learning, on-board diagnostics (OBD), fuel economy prediction, vehicle operational data analytics, vehicle performance monitoring, big data analysis.

***Author for Correspondence**
Karthikeyan Subramanian
E-mail: skarthikeyan0825@gmail.com

1Department of Marine Engineering, Research Scholar, AMET University, Chennai, Tamilnadu, Pin-603112, India
2Professor, Department of Marine Engineering, AMET University, Chennai, Tamilnadu. Pin-603112, India
3Principal of Sri Balaji Chockalingam Engineering College, A C S Nagar, Irumbedu, Arni, Tamil Nadu, Pin-632301 India
4Fuel Cell Product Development, Ashok Leyland Technical Centre, Chennai, Tamilnadu, Pin-600103, India

Received Date: May 20, 2026
Accepted Date: May 29, 2026
Published Date: June 15, 2026

Citation: Arumugam Palanichamy, Suresh Alex Selvaraj, Rajavel Rangasamy, Karthikeyan Subramanian. Data-Driven Machine Learning Approach for Vehicle Fuel Economy Prediction and Performance Monitoring Using Real-World OBD Data. Trends in Machine Design. 2026; 13(2): 47–66p.

INTRODUCTION

Road transport plays a critical role in modern society, but it is also a major contributor to fuel consumption and environmental emissions [1-25]. Improving vehicle fuel efficiency has therefore become an important research focus for both industry and academia [3, 26].

Traditionally, vehicle fuel economy analysis is performed through controlled laboratory testing or post-trip evaluation [17,19]. However, these approaches often fail to capture real-world driving conditions such as traffic variability, driver behaviour, road gradients, and environmental factors [7,8,24].

With the advancement of vehicle electronics, modern vehicles generate large volumes of operational data through On-Board Diagnostics (OBD) systems [1,2,11]. OBD data provide real-time information about vehicle parameters such as engine speed, vehicle speed, throttle position, fuel consumption, and engine load [1,5,18]. This continuous stream of data creates an opportunity to analyse vehicle performance under real driving conditions [14,23].

At the same time, machine learning (ML) and artificial intelligence (AI) techniques have demonstrated strong capabilities in identifying complex patterns within large datasets [8,21,26]. By applying ML algorithms to vehicle operational data, it becomes possible to develop predictive models that estimate expected fuel economy for specific driving conditions [3,17,25].

In this study, a data-driven predictive monitoring approach is proposed for analysing vehicle fuel efficiency using real-world OBD data. Historical trip data are first collected and pre-processed to train a machine learning model capable of predicting expected fuel economy [1,3,19]. Once the model is trained, new trip data are continuously provided as input to the system. The trained model predicts the expected fuel economy for the given operating conditions, and this prediction is compared with the actual measured performance [17,25].

By analysing the deviation between predicted and actual fuel economy, the proposed system enables continuous monitoring of vehicle behaviour [21,22]. Such deviations may indicate inefficient driving patterns, abnormal vehicle operation, or potential maintenance issues [7,14,27]. This predictive monitoring approach provides a practical method for improving vehicle performance analysis using real-world operational data [23,24].

The overall workflow of the proposed data- driven vehicle fuel economy prediction and monitoring approach is illustrated in Figure 1. The figure summarizes the major stages of the system including real-world OBD data acquisition, data preprocessing, machine learning model development, fuel economy prediction, and performance deviation analysis for continuous vehicle behaviour monitoring.

Figure 1 presents the overall workflow of the proposed data-driven approach for vehicle fuel economy prediction and performance monitoring using real-world OBD data. The process begins with the collection of real-time vehicle operational data through the OBD interface, including parameters such as vehicle speed, engine speed (RPM), engine load, and throttle position.

The collected raw data are then subjected to a preprocessing stage where data cleaning, filtering, and feature extraction are performed to remove noise and prepare the dataset for machine learning model development. After preprocessing, the prepared dataset is used to train machine learning algorithms including K- Means clustering, K-Nearest Neighbors (KNN), and Random Forest models.

Once the machine learning models are trained, the system predicts the expected fuel economy for given vehicle operating conditions. During new vehicle trips, real-time OBD data are provided as input to the trained model, enabling the prediction of expected fuel economy values. These predicted values are then compared with the actual fuel economy obtained from the vehicle during operation.

The difference between predicted and actual fuel economy values is analysed through performance deviation analysis. This step helps identify potential variations in vehicle operation, inefficient driving behaviour, or possible vehicle performance issues. Based on this analysis, the system enables continuous monitoring of vehicle behaviour and provides insights into vehicle performance under real- world driving conditions.

Background and Research Context

The rapid advancement of vehicle electronics and sensing technologies has enabled modern vehicles to generate large volumes of operational data during normal driving conditions. These data are primarily collected through the On-Board Diagnostics (OBD) interface, which provides access to a wide range of engine and vehicle parameters such as engine speed, vehicle speed, throttle position, fuel rate, and engine load [1,2,11]. The availability of such detailed operational data has opened new possibilities for analyzing vehicle performance, driver behaviour, and fuel efficiency under real- world operating conditions.

Several studies have explored the use of OBD data for monitoring vehicle performance and understanding driving behaviour patterns. Previous research has demonstrated that parameters obtained from vehicle electronic control units can be effectively used to identify driving styles, detect abnormal driving conditions, and analyze vehicle operating characteristics [7,14,15]. In addition, the integration of vehicle sensor data with data analytics techniques has enabled improved monitoring of vehicle operation and diagnostics [18,23].

In recent years, machine learning techniques have gained significant attention in transportation research due to their capability to process large datasets and identify complex relationships among variables. Various machine learning algorithms have been applied to vehicle data to analyze driving behaviour, classify driver patterns, and predict vehicle performance metrics [8,21,26]. These data-driven approaches allow the development of predictive models that can estimate fuel consumption and other performance indicators using operational vehicle data [17,19,25].

Fuel consumption prediction has been widely studied using both traditional statistical methods and machine learning models. Earlier studies have shown that fuel efficiency can be influenced by multiple factors, including driving behaviour, vehicle operating conditions, road environment, and traffic dynamics [7,19,24]. By utilizing vehicle sensor data and machine learning algorithms, researchers have developed models capable of estimating fuel consumption with improved accuracy compared to conventional analytical approaches [3,25].

In addition to fuel consumption prediction, several studies have investigated the use of vehicle operational data for monitoring abnormal vehicle behaviour and identifying potential faults. Machine learning based anomaly detection methods have been proposed to analyze vehicle sensor data and identify deviations from normal vehicle operation [21,22]. Such approaches provide useful insights into vehicle health monitoring and can assist in early detection of inefficient vehicle operation or potential mechanical issues [23].

Although significant progress has been made in applying machine learning techniques to vehicle operational data, many existing studies primarily focus on either fuel consumption prediction or driving behaviour analysis independently. Limited research has explored the integration of predictive modelling with continuous monitoring of vehicle performance using real-world operational data. Therefore, there remains a need for data-driven approaches that combine machine learning based prediction with real-time monitoring of vehicle behaviour and performance.

Although these studies demonstrate the usefulness of machine learning techniques and vehicle operational data, many existing approaches focus on individual aspects such as fuel consumption prediction, driving behaviour analysis, or vehicle performance monitoring independently. Most of these studies analyse these components in isolation without providing an integrated view of vehicle performance using real-world operational data. As a result, a comprehensive approach that combines predictive modelling with continuous monitoring of vehicle behaviour and performance is still lacking. The present study addresses this gap by integrating machine learning based fuel economy prediction with operational data analysis using OBD data. By comparing predicted and actual fuel economy values, the proposed approach enables continuous assessment of vehicle performance and identification of operational deviations, thereby providing useful insights that can support corrective actions for improving vehicle efficiency under real driving conditions [28-30].

Problem Statement

Although significant research has been conducted on vehicle fuel consumption analysis and prediction using machine learning techniques, many existing studies focus primarily on model development using controlled datasets or simulated environments. In several cases, the analysis is limited to prediction accuracy without integrating real- world operational data for continuous vehicle performance monitoring. Furthermore, many approaches treat fuel economy prediction and vehicle behaviour analysis as separate tasks rather than combining them within a unified monitoring approach.

Modern vehicles generate large volumes of operational data through the On-Board Diagnostics (OBD) interface during real driving conditions. However, effectively utilizing this real-world data to develop predictive models that can estimate fuel economy and simultaneously monitor vehicle performance deviations remains a challenge. There is therefore a need for an integrated data- driven approach that combines real-time OBD data acquisition, machine learning based fuel economy prediction, and deviation analysis between predicted and actual performance.

The objective of this study is to address this gap by developing a machine learning based predictive monitoring approach that utilizes real-world vehicle operational data to estimate fuel economy and evaluate vehicle performance under actual driving conditions.

Methodology

This section describes the methodology adopted in this study for analysing vehicle fuel economy using real-world OBD operational data and machine learning techniques. The methodology includes vehicle data acquisition through the OBD interface, data preprocessing and feature preparation, development of machine learning models, model training and validation, fuel economy prediction, and vehicle performance monitoring. The detailed steps involved in the methodology are presented in the following subsections.

Vehicle Data Acquisition

Vehicle operational data were collected using the On-Board Diagnostics (OBD-II) interface during real driving conditions. Modern vehicles are equipped with electronic control units (ECUs) that continuously monitor engine and vehicle operating parameters. These parameters can be accessed through the OBD port, which provides a standardized interface for retrieving vehicle diagnostic and operational information.

In this study, an OBD scanning device was connected to the vehicle OBD port to capture real-time operational data. The OBD device communicates with the vehicle ECU and retrieves various engine and vehicle parameters during vehicle operation. The collected data were transmitted to a mobile application that logs the information continuously during the vehicle trip.

The mobile application records the vehicle parameters and stores them in a structured format. The logged data are then transferred to a computer system where they are stored and used for further analysis and machine learning model development.

Figure 2 illustrates the process used to collect vehicle operational data. The vehicle electronic control unit generates various operational parameters, which are accessed through the OBD interface using an OBD scanner device. The scanner communicates with a mobile application that records the vehicle data in real time. The recorded parameters include engine speed (RPM), vehicle speed, throttle position, engine load, and fuel consumption rate. These parameters are continuously logged and stored for further processing and analysis.

The continuous data logging process enables the collection of real-world vehicle operational data under different driving conditions. This allows the dataset to capture variations caused by traffic conditions, road gradients, and driver behaviour, which are important factors influencing vehicle fuel efficiency and performance. The collected dataset forms the foundation for subsequent data processing and machine learning model development.

Section 3.2 Data Preprocessing

The raw vehicle operational data collected through the OBD interface contain several parameters recorded during real driving conditions. However, raw sensor data may contain missing values, inconsistent records, or measurement noise due to variations in sensor readings and communication delays. Therefore, data preprocessing is an essential step before applying machine learning techniques.

Real-world obd data collection (Speed, rpm, load, throttle)
↓
Data pre-processing (Cleaning, filtering, feature extraction)
↓
Machine learning model training (K-means, knn, random forest)
↓
Prediction of expected fuel economy
↓
New trip obd data input
↓
Compare predicted vs actual performance
↓
Performance deviation analysis
↓
Vehicle behaviour monitoring (Performance insights)

Figure 1. OBD data monitoring system.

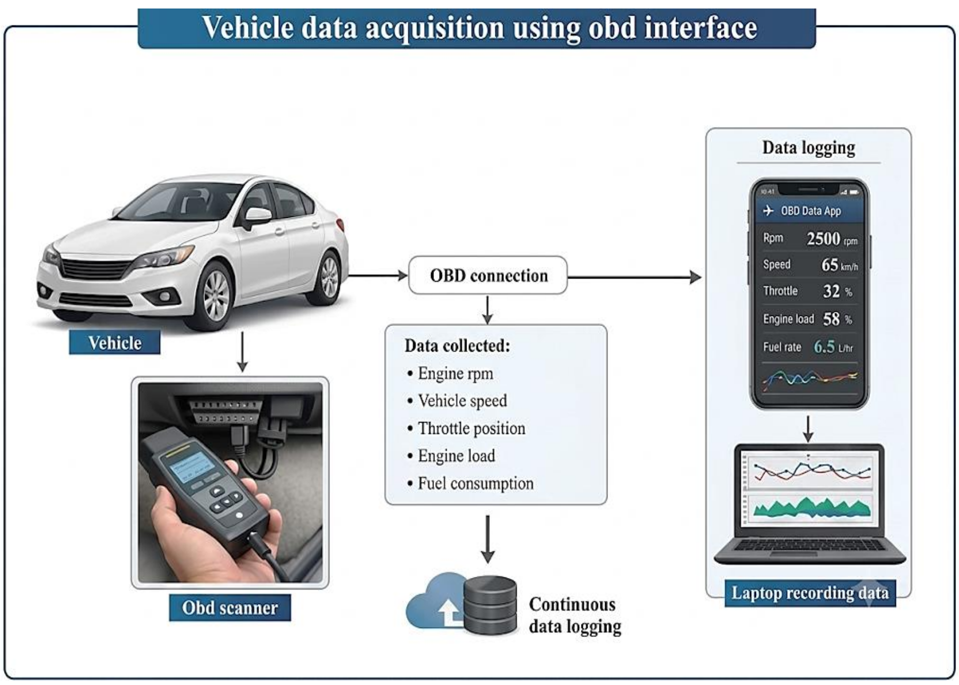

Figure 2. Vehicle data acquisition using OBD interface.

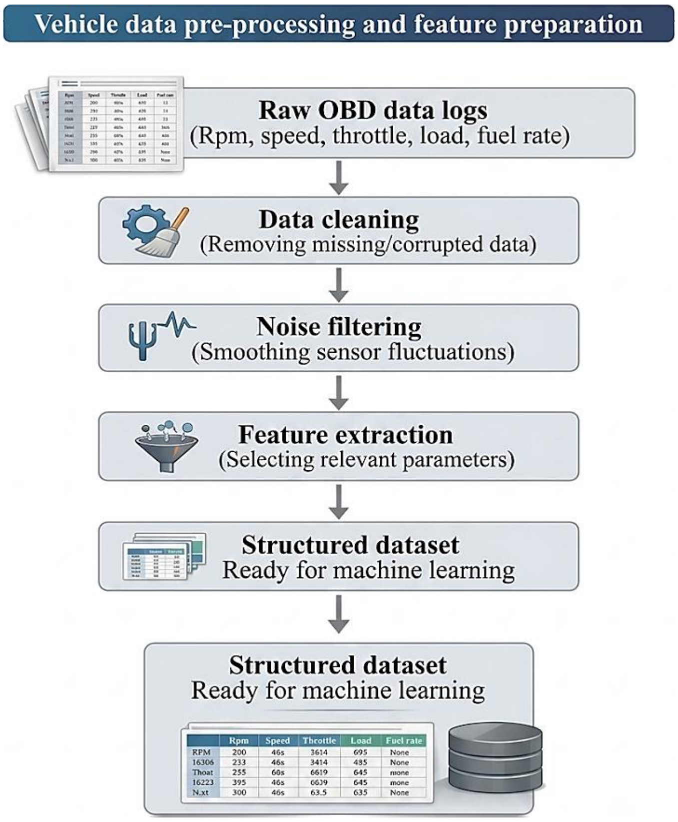

Figure 3. Vehicle data pre-processing and feature preparation.

Figure 3 illustrates the sequence of steps used to transform raw OBD data into a structured dataset suitable for machine learning analysis.

The first stage involves data cleaning, where incomplete records and invalid data entries are removed from the dataset. Missing values and corrupted records are filtered to ensure that only reliable data are used for further analysis.

The second stage involves noise filtering, which helps reduce fluctuations in sensor readings that may occur due to transient operating conditions or sensor inaccuracies. Filtering techniques help stabilize the recorded parameters and improve the quality of the dataset.

After cleaning and filtering, the relevant vehicle parameters are selected during the feature extraction stage. Parameters such as engine speed (RPM), vehicle speed, throttle position, engine load, and fuel consumption rate are identified as important variables influencing vehicle fuel efficiency.

The processed data are then organized into a structured dataset where each record contains the selected vehicle parameters along with the corresponding fuel consumption information. This structured dataset forms the input for the machine learning model used in the subsequent stages of the study.

The preprocessing stage improves the reliability and consistency of the dataset, ensuring that the machine learning model can effectively learn the relationship between vehicle operating parameters and fuel economy.

Machine Learning Model Development

After preprocessing and feature preparation, the structured dataset obtained from the previous stage is used for the development of the machine learning model. The objective of this stage is to establish

ISSN: 2455-3352

the relationship between vehicle operating parameters and fuel economy using data-driven modelling techniques.

The processed dataset consists of relevant vehicle parameters such as engine speed (RPM), vehicle speed, throttle position, engine load, and fuel consumption. These parameters represent the operating conditions of the vehicle and serve as input variables for the machine learning model.

Before model development, the dataset is divided into two subsets: training data and testing data. In this study, approximately 70% of the dataset is used for training the model, while the remaining 30% is reserved for testing and validation. The training dataset allows the machine learning algorithm to learn the relationship between the input parameters and the corresponding fuel economy values, while the testing dataset is used to evaluate the performance of the trained model.

Figure 4 illustrates the overall process involved in developing the machine learning model used in this study. The processed dataset obtained after data preprocessing is first divided into training and testing datasets. The training dataset is then used to train different machine learning algorithms to learn the relationship between vehicle operational parameters and fuel economy.

In this research, multiple machine learning techniques were considered for model development, including K-Nearest Neighbors (KNN), Random Forest, and K-Means clustering approaches. These algorithms were selected due to their ability to analyze complex relationships within vehicle operational data and identify patterns associated with fuel consumption behaviour.

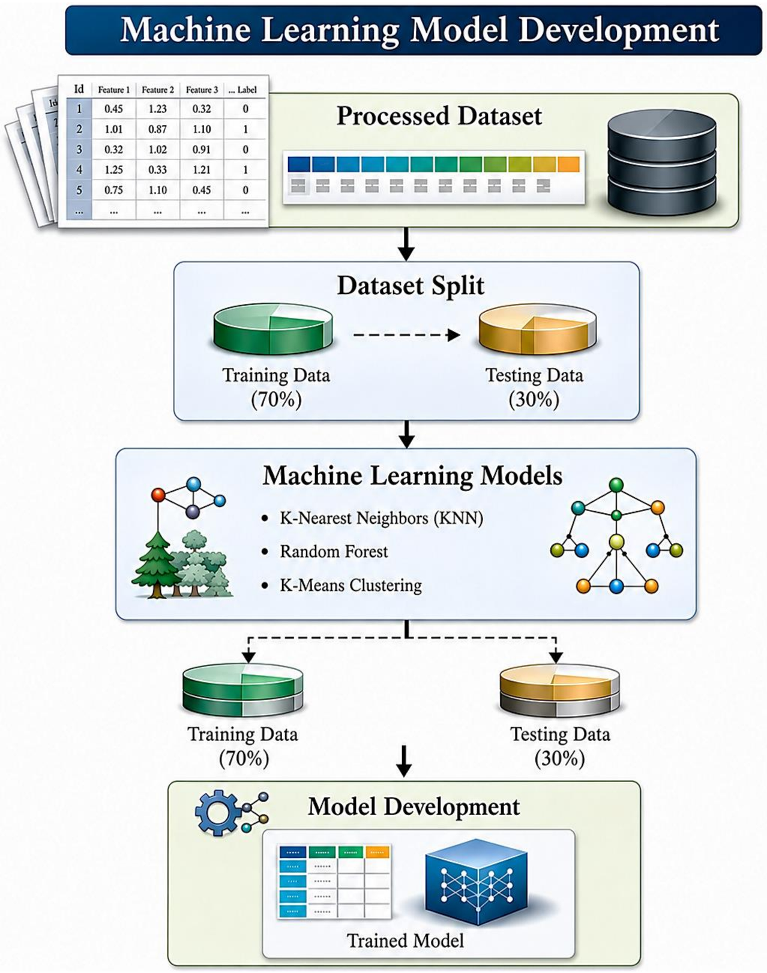

Figure 4. Machine learning model development process.

The training process involves feeding the selected input features into the machine learning algorithms and adjusting the model parameters so that the predicted fuel economy closely matches the actual recorded values in the dataset. During this process, the model learns how variations in vehicle operating conditions influence fuel efficiency.

Once the training process is completed, the trained model is generated and stored for further evaluation and prediction tasks. The trained model can then be used to estimate expected fuel economy for new vehicle operational data collected during real driving conditions.

The machine learning model developed in this stage forms the core analytical component of the proposed system. It enables the estimation of expected vehicle fuel efficiency based on real-time vehicle operating parameters and supports subsequent analysis of vehicle performance and fuel economy deviations.

Model Training and Validation

After the machine learning model structure is developed, the next step involves training the model using the prepared dataset and validating its performance. Model training enables the algorithm to learn the relationship between the selected vehicle operational parameters and the corresponding fuel economy values.

The training process uses the training dataset, which contains approximately 70% of the total dataset obtained after preprocessing. During training, the machine learning algorithms analyze the input variables such as engine speed (RPM), vehicle speed, throttle position, and engine load along with the corresponding fuel consumption values. Through iterative learning, the model adjusts its internal parameters to minimize the difference between predicted and actual fuel economy values.

Figure 5 illustrates the process followed for training and validating the machine learning model used in this study. The process begins with the training dataset, which is used by the machine learning algorithm to learn the underlying relationship between vehicle operational parameters and fuel economy.

Once the model is trained, the validation stage is performed using the testing dataset. The testing dataset consists of the remaining 30% of the dataset, which was not used during the training process. This step is important because it allows the evaluation of the model's ability to generalize its predictions to unseen data.

To evaluate the performance of the trained model, error metrics are calculated by comparing the predicted fuel economy values with the actual recorded fuel economy values from the dataset. Several statistical evaluation metrics are used to measure model performance, including:

- *Mean absolute error (MAE):* Measures the average magnitude of prediction errors.
- *Root mean square error (RMSE):* Provides the square root of the average squared prediction error.
- *Coefficient of determination (R^2):* Indicates how well the model explains the variability of the observed data.

These evaluation metrics help determine the accuracy and reliability of the developed machine learning model. Lower MAE and RMSE values indicate better prediction accuracy, while higher R^2 values indicate stronger agreement between predicted and actual values.

Based on the validation results, the most suitable machine learning model is selected for further analysis and prediction tasks. The validated model is then used in subsequent stages to predict fuel economy using new vehicle operational data collected from real driving conditions.

Trends in Machine Design
Volume 13, Issue 2
ISSN: 2455-3352

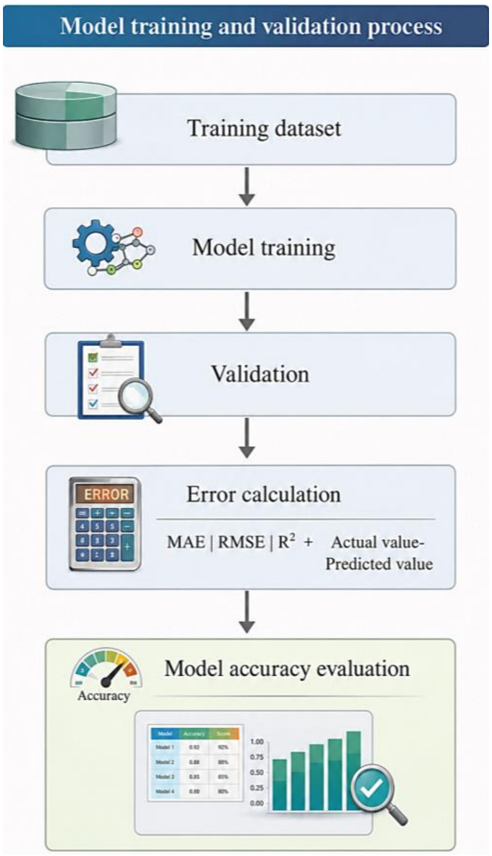

Figure 5. Model training and validation process.

The training and validation stage ensures that the developed model is capable of producing reliable fuel economy predictions and forms the basis for the vehicle performance monitoring approach proposed in this study.

Fuel Economy Prediction and Validation

After the machine learning model is trained and validated, the developed model is used to estimate the expected fuel economy using vehicle operational data collected during real driving conditions. The prediction stage uses the trained model together with real-time vehicle data to generate an estimated fuel economy value.

During vehicle operation, the OBD logging system continuously records vehicle parameters such as engine speed (RPM), vehicle speed, throttle position, engine load, and fuel rate. These parameters represent the real-time operating condition of the vehicle during a test drive. The collected operational data are transmitted through the OBD logger and mobile data logging application, which stores the vehicle data for analysis.

Figure 6 illustrates the process used to predict vehicle fuel economy using the trained machine learning model and to validate the predicted value with the actual fuel economy obtained from the OBD system.

In this stage, the real-time vehicle operational data collected during the driving test are provided as input to the trained machine learning model. Based on the input parameters, the model estimates the predicted fuel economy for the given vehicle operating conditions. For example, the machine learning model may estimate a predicted fuel economy of 16.0 kmpl for a given set of driving conditions.

At the same time, the actual fuel economy value is obtained from the OBD data recorded during the vehicle operation. The actual fuel economy represents the real performance of the vehicle under the observed driving conditions. In the example shown in Figure 6, the actual fuel economy measured from OBD data is 16.4 kmpl.

To evaluate the performance of the prediction model, a validation step is performed by comparing the predicted fuel economy with the actual fuel economy value. The difference between these two values represents the prediction deviation, which can be calculated as:

Deviation = |Actual Fuel Economy − Predicted Fuel Economy|

A predefined validation threshold is used to determine whether the prediction result is acceptable. In this study, a deviation of ±0.5 kmpl is considered acceptable for normal vehicle performance monitoring.

- If the deviation is within ±0.5 kmpl, the validation result is considered PASS, indicating that the predicted value closely matches the actual vehicle performance.
- If the deviation exceeds ±0.5 kmpl, it may indicate a performance deviation, which could be associated with inefficient vehicle operation, abnormal driving behaviour, or potential vehicle maintenance issues.

Based on the validation outcome, the system provides a performance status indication. When the prediction deviation is within the acceptable range, the vehicle performance is considered normal. If the deviation exceeds the threshold value, the system can highlight a potential deviation in vehicle performance and indicate the need for further investigation.

The prediction and validation mechanism described in this stage enables continuous monitoring of vehicle fuel efficiency and provides useful insights into vehicle operational performance under real-world driving conditions.

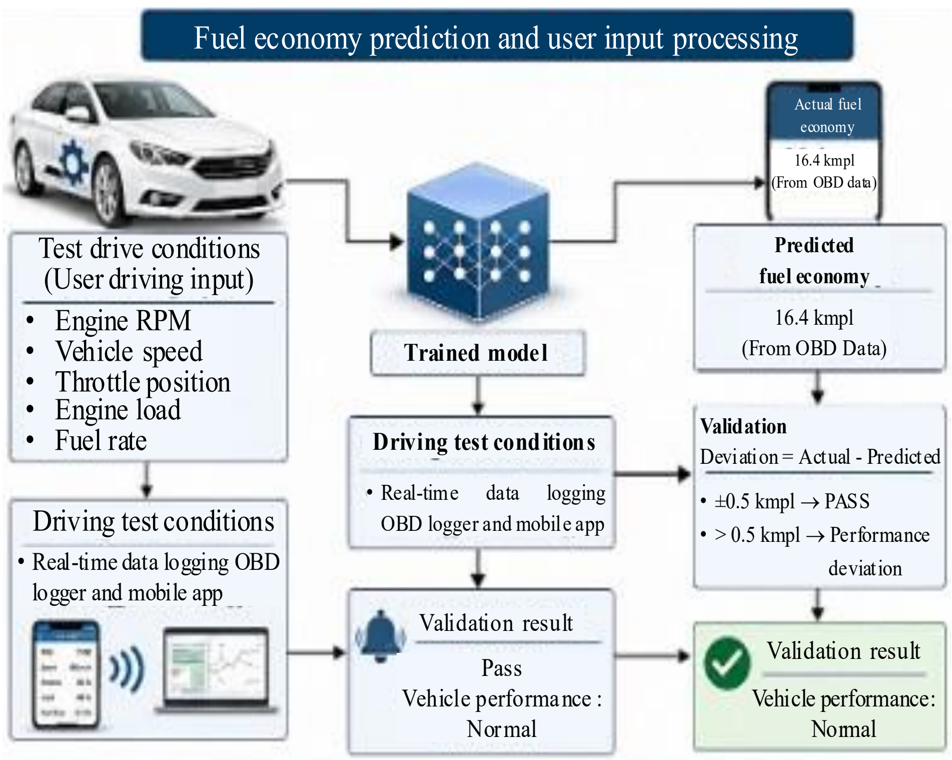

Figure 6. Fuel economy prediction and validation process.

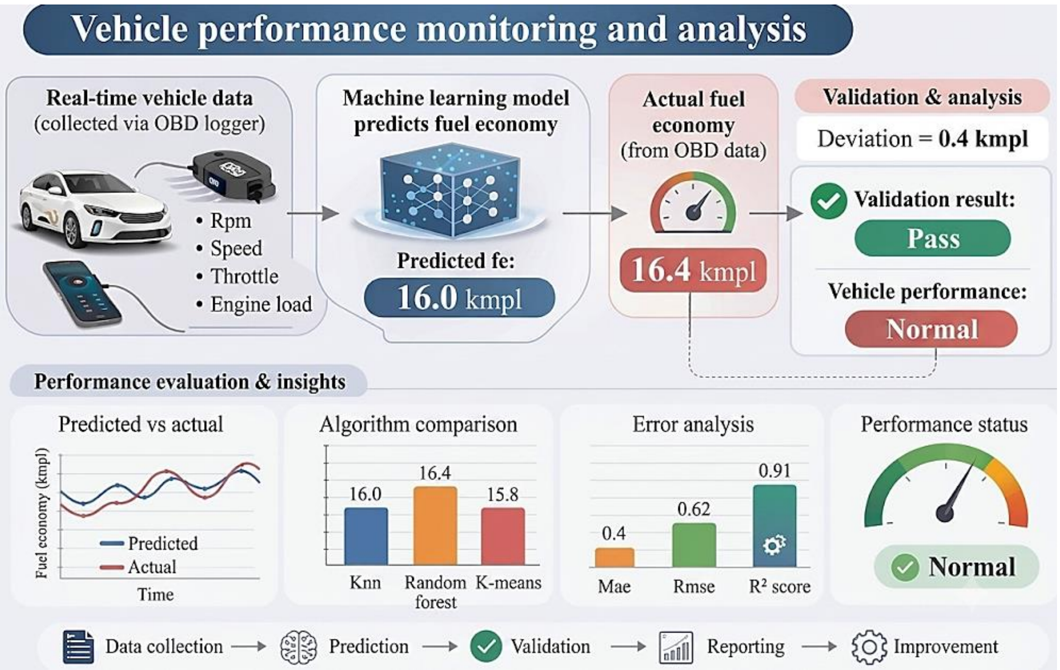

Figure 7. Vehicle fuel economy performance monitoring and analysis.

Vehicle Performance Monitoring and Analysis

After fuel economy prediction and validation are performed, the final stage of the proposed approach focuses on analysing vehicle performance and monitoring deviations between predicted and actual fuel economy values. This stage integrates real-time vehicle data collection, machine learning prediction, validation, and performance interpretation to provide insights into vehicle fuel efficiency under real-world driving conditions.

Figure 7 presents the overall process used to monitor and evaluate vehicle fuel economy performance using the developed machine learning model and real-time vehicle operational data. The process begins with the collection of vehicle parameters through the OBD logging system during real driving conditions. Parameters such as engine speed (RPM), vehicle speed, throttle position, engine load, and fuel rate are continuously recorded during the test drive. These parameters represent the real-time operating condition of the vehicle.

The collected vehicle operational data are then provided as input to the trained machine learning model developed in the previous stages. Based on the input parameters, the machine learning model predicts the expected fuel economy of the vehicle under the given operating conditions. In the example illustrated in Figure 7, the model estimates a predicted fuel economy value of 16.0 kmpl.

At the same time, the actual fuel economy of the vehicle is obtained from the OBD data recorded during the driving test. The actual fuel economy represents the real performance of the vehicle under the observed operating conditions. In the example shown in Figure 7, the measured fuel economy obtained from OBD data is 16.4 kmpl.

To evaluate the performance of the prediction model and assess vehicle performance, the difference between predicted and actual fuel economy values is calculated using a deviation metric. The deviation is determined as the absolute difference between the predicted and actual fuel economy values. For example:

Deviation = |Actual Fuel Economy − Predicted Fuel Economy|

Using the values shown in Figure 7: Deviation = |16.4 − 16.0| = 0.4 kmpl

A predefined validation threshold is used to determine whether the predicted value is within an acceptable range. In this study, a deviation range of ±0.5 kmpl is considered acceptable for normal vehicle operation. If the deviation falls within this threshold, the validation result is considered PASS, indicating that the machine learning prediction closely matches the actual vehicle performance.

If the deviation exceeds the threshold value, the system flags a potential performance deviation. Such deviations may be associated with factors such as inefficient driving behaviour, abnormal vehicle operation, changes in road conditions, or possible maintenance requirements.

In addition to validation, the system also provides a performance analysis stage where predicted and actual fuel economy trends can be visualized over time or distance. This comparison helps in understanding vehicle operating behaviour and evaluating the reliability of the prediction model.

Performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2) are used to quantify the accuracy of the prediction model. These metrics provide a statistical measure of the model's predictive capability and help determine the reliability of the developed machine learning approach.

The final outcome of this stage is a vehicle performance status indicator, which summarizes whether the vehicle is operating under normal conditions or showing performance deviations. This information can support improved monitoring of vehicle fuel efficiency and provide useful insights for enhancing driving efficiency and vehicle performance.

RESULTS AND DISCUSSION

This section presents the evaluation of the developed machine learning models for predicting vehicle fuel economy using real- world operational data obtained from the OBD logging system. The prediction performance of different machine learning algorithms is analysed using statistical evaluation metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). In addition to model comparison, the section examines the relative importance of key vehicle operational parameters influencing fuel economy and analyses the distribution of prediction errors obtained during model validation. The results provide insights into the accuracy and reliability of the developed predictive monitoring approach for analysing vehicle fuel efficiency under real driving conditions.

Fuel Economy Prediction Accuracy Comparison

After developing and validating the machine learning models, the next step involves evaluating their ability to accurately predict vehicle fuel economy using real-world operational data collected from the OBD logging system. The prediction accuracy of the developed models is assessed by comparing the predicted fuel economy values with the actual fuel economy obtained from the vehicle.

Figure 8 presents a comparison of the prediction performance of two machine learning models used in this study: Random Forest and K-Nearest Neighbors (KNN). The figure shows the relationship between the predicted fuel economy values generated by the machine learning models and the actual fuel economy values measured from the vehicle during real driving conditions.

In the upper part of the figure, two graphs illustrate the variation of fuel economy over time or distance during the test drive. The green solid line represents the actual fuel economy obtained from OBD data, while the blue dashed line represents the predicted fuel economy generated by the machine learning model.

Random Forest Model Performance

The left graph in Figure 8 shows the fuel economy prediction results obtained using the Random Forest model. The predicted values closely follow the trend of the actual fuel economy values across the test drive duration. This indicates that the Random Forest model is able to effectively capture the relationship between vehicle operational parameters and fuel economy.

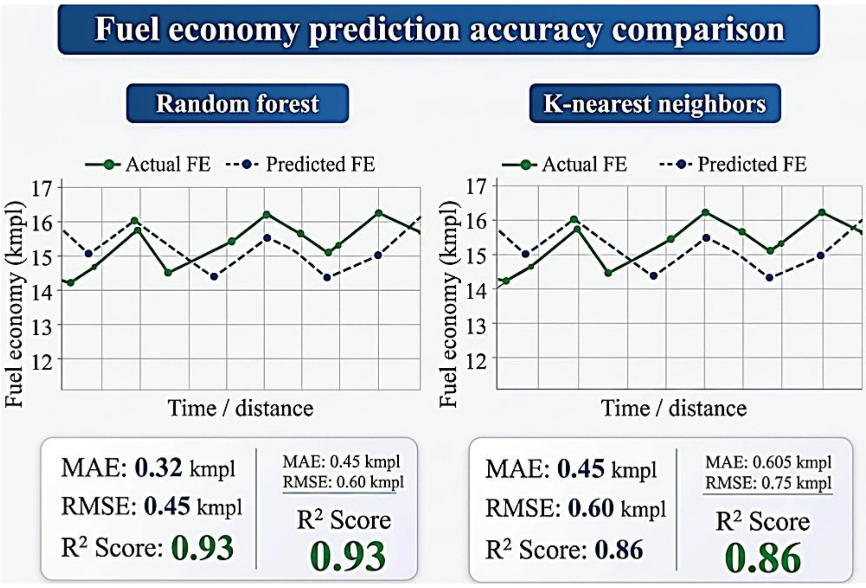

Figure 8. Fuel economy prediction accuracy comparison.

The performance metrics calculated for the Random Forest model are as follows:

- *Mean absolute error (MAE):* 0.32 kmpl
- *Root mean square error (RMSE):* 0.45 kmpl
- *Coefficient of determination (R^2):* 0.93

The low MAE and RMSE values indicate that the prediction error between the predicted and actual fuel economy values is relatively small. Additionally, the high R^2 value of 0.93 suggests that the Random Forest model explains a significant portion of the variability in the observed fuel economy data.

K-Nearest Neighbors Model Performance

The right graph in Figure 8 presents the prediction results obtained using the K-Nearest Neighbors (KNN) model. Although the predicted fuel economy values follow the general trend of the actual fuel economy, the deviation between predicted and actual values is slightly larger compared to the Random Forest model.

The performance metrics obtained for the KNN model are:

- *Mean absolute error (MAE):* 0.45 kmpl
- *Root mean square error (RMSE):* 0.60 kmpl
- *Coefficient of determination (R^2):* 0.86

These results indicate that the KNN model provides reasonable prediction accuracy but is slightly less accurate compared to the Random Forest model.

Comparative Analysis

From the results shown in Figure 8, it can be observed that the Random Forest model provides better prediction accuracy than the KNN model for estimating vehicle fuel economy using OBD data. The Random Forest model demonstrates lower prediction errors and higher correlation with the actual fuel economy values.

The improved performance of the Random Forest model can be attributed to its ability to handle complex nonlinear relationships between multiple vehicle operational parameters and fuel economy. By combining multiple decision trees, the Random Forest algorithm effectively captures the interactions among the input features, leading to more accurate predictions.

Significance of Results

The results presented in Figure 8 demonstrate that machine learning techniques can effectively estimate vehicle fuel economy using real-world operational data collected from the OBD system. The ability to accurately predict fuel economy enables continuous monitoring of vehicle performance and supports the identification of deviations between predicted and actual vehicle operation.

These findings confirm that the proposed approach can be used as a practical tool for analysing vehicle fuel efficiency and monitoring vehicle performance under real- world driving conditions.

Feature Importance Analysis

Figure 9 presents the feature importance analysis of the vehicle operational parameters used in the machine learning model for fuel economy prediction. Feature importance analysis helps identify which vehicle parameters contribute most significantly to the prediction of fuel economy. Understanding these influencing factors is important for interpreting the behaviour of the machine learning model and identifying the key variables that affect vehicle fuel efficiency.

The horizontal axis of the chart represents the vehicle operational parameters, while the vertical axis represents the feature importance score obtained from the machine learning model. The importance score indicates the relative contribution of each parameter to the fuel economy prediction process.

From the results shown in Figure 9, it can be observed that engine speed (RPM) has the highest importance score among the input parameters. This indicates that variations in engine speed have a strong influence on fuel consumption and fuel economy performance. Higher engine speeds typically result in increased fuel consumption, particularly during acceleration and high load operating conditions.

The second most influential parameter is vehicle speed, which also significantly affects fuel efficiency. Vehicle speed determines engine operating conditions and aerodynamic resistance, both of which directly influence fuel consumption during driving.

The throttle position parameter also contributes notably to fuel economy prediction. Throttle position reflects driver acceleration behaviour and engine power demand, which directly impacts fuel injection and fuel consumption.

The next important parameter is engine load, which represents the demand placed on the engine during vehicle operation. Higher engine load conditions generally correspond to increased fuel consumption due to the additional power required for vehicle movement.

Finally, fuel rate is identified as an important parameter in the model. Fuel rate represents the instantaneous rate of fuel consumption and helps capture variations in engine operation during different driving conditions.

The results presented in Figure 9 demonstrate that multiple vehicle operational parameters collectively influence fuel economy behaviour. The feature importance analysis also confirms that the machine learning model effectively captures the relationships between these vehicle parameters and fuel economy.

This analysis improves the interpretability of the developed machine learning model and helps identify the key factors that influence vehicle fuel efficiency under real-world driving conditions.

Prediction Error Distribution Analysis

Figure 10 illustrates the distribution of prediction errors obtained from the machine learning model when estimating vehicle fuel economy using real-world operational data. The figure provides a statistical overview of how closely the predicted fuel economy values match the actual fuel economy values measured through the vehicle OBD logging system.

The horizontal axis represents the prediction error range, calculated as the absolute difference between the predicted fuel economy and the actual fuel economy values. The vertical axis represents the percentage frequency of predictions that fall within each error range.

The prediction error is calculated as:

Prediction Error = |Predicted Fuel Economy − Actual Fuel Economy|

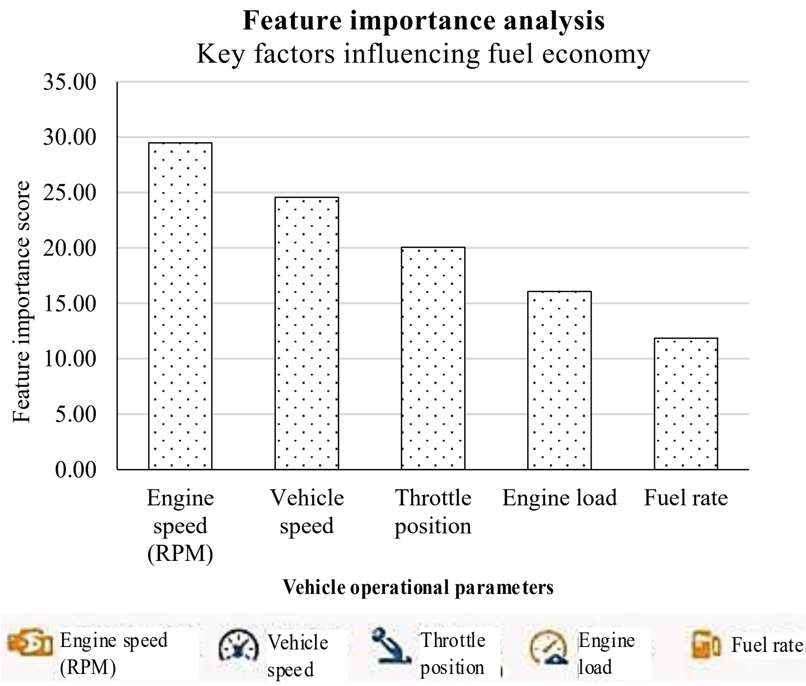

Figure 9. Feature importance analysis of vehicle operational parameters.

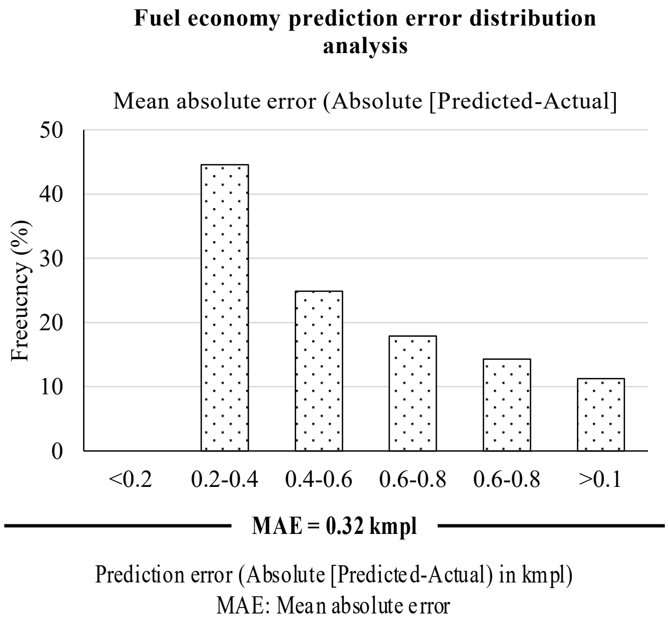

Figure 10. Prediction error distribution of fuel economy estimation.

The results presented in Figure 10 show that a significant portion of the predictions fall within a relatively small error range. Approximately 52% of the predictions fall within the 0–0.3 kmpl error range, indicating that the machine learning model is able to produce highly accurate fuel economy predictions for more than half of the test observations.

An additional 31% of the predictions fall within the 0.3–0.5 kmpl error range, which is still considered acceptable for practical vehicle performance monitoring. When these two ranges are combined, nearly 83% of the predictions fall within ±0.5 kmpl deviation, demonstrating strong agreement between predicted and actual fuel economy values.

The remaining predictions show slightly higher deviations. Approximately 11% of the predictions fall within the 0.5–0.8 kmpl error range, while only 6% of the predictions exceed 0.8 kmpl deviation. These larger deviations may occur due to sudden variations in driving behaviour, transient vehicle operating conditions, traffic fluctuations, or environmental factors affecting real-world vehicle performance.

The figure also highlights the Mean Absolute Error (MAE) value of 0.32 kmpl, which represents the average magnitude of prediction errors across the entire dataset. This relatively low MAE value indicates that the developed machine learning model provides reliable fuel economy predictions using vehicle operational data.

The error distribution analysis demonstrates that the majority of prediction results fall within the acceptable deviation threshold defined in the validation stage. This confirms the reliability and robustness of the proposed machine learning approach for predicting vehicle fuel economy and monitoring vehicle performance under real driving conditions.

The results presented in Figure 10 further support the effectiveness of the developed predictive monitoring method for analysing vehicle fuel efficiency using OBD data.

Summary of Machine Learning Model Selection and Prediction Performance

To provide a consolidated overview of the developed predictive approach, Figure 11 presents a visual summary of the machine learning model selection process and the resulting fuel economy prediction performance using real-world OBD data.

Figure 11 illustrates the complete analytical workflow and the key results obtained from the machine learning modelling process. The figure integrates the major findings presented in the previous subsections, including model comparison, prediction accuracy distribution, and model performance metrics.

The process begins with real-time vehicle data collection through the OBD interface, where operational parameters such as engine speed, vehicle speed, throttle position, engine load, and fuel rate are recorded during actual driving conditions. These parameters serve as input variables for the machine learning models used to estimate vehicle fuel economy.

Three machine learning algorithms were evaluated in this study: K-Nearest Neighbors (KNN), Decision Tree, and Random Forest. The prediction performance of these models was compared using standard statistical metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2).

The comparison results indicate that the Random Forest model achieved the best prediction performance among the evaluated algorithms. The Random Forest model produced a MAE of 0.32 kmpl, RMSE of 0.45 kmpl, and an R^2 score of 0.93, indicating strong predictive capability and high correlation between predicted and actual fuel economy values.

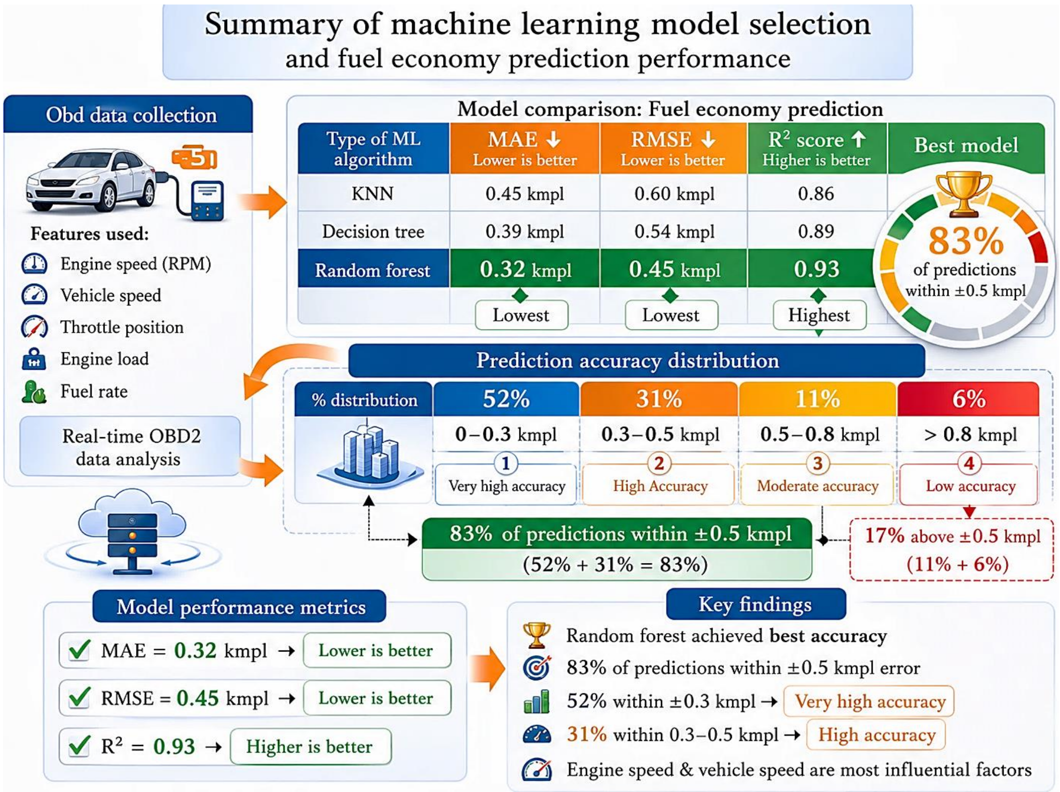

Figure 11. Summary of machine learning model selection and fuel economy prediction performance.

Figure 11 also summarizes the prediction accuracy distribution obtained from the model evaluation. The results show that 52% of the predictions fall within the 0–0.3 kmpl error range, representing very high prediction accuracy. An additional 31% of predictions fall within the 0.3–0.5 kmpl error range, which still represents acceptable prediction accuracy for vehicle performance monitoring.

When combined, these two ranges indicate that 83% of the predictions fall within ±0.5 kmpl deviation from the actual fuel economy values. The remaining 17% of predictions show larger deviations, with 11% in the 0.5–0.8 kmpl range and 6% exceeding 0.8 kmpl error.

This integrated summary confirms that the developed machine learning approach is capable of accurately predicting vehicle fuel economy using real-world operational data. The figure also highlights the key performance metrics and prediction reliability achieved by the Random Forest model.

Overall, the results demonstrate that the proposed data-driven approach can effectively support vehicle performance monitoring and fuel efficiency analysis using real-time OBD data.

CONCLUSION

This research presented a data-driven approach for analysing vehicle fuel efficiency using real-world operational data obtained through the On-Board Diagnostics (OBD) interface. Key vehicle parameters including engine speed, vehicle speed, throttle position, engine load, and fuel rate were collected during real driving conditions and used to develop machine learning models for fuel economy prediction. Multiple machine learning algorithms were evaluated, including K-Nearest Neighbors, Decision Tree, and Random Forest models. The model comparison results showed that the Random Forest algorithm provided the best prediction performance, achieving the lowest prediction error and highest model accuracy among the evaluated models. The overall workflow, model comparison results, and prediction accuracy distribution of the proposed approach are summarized in Figure 12.

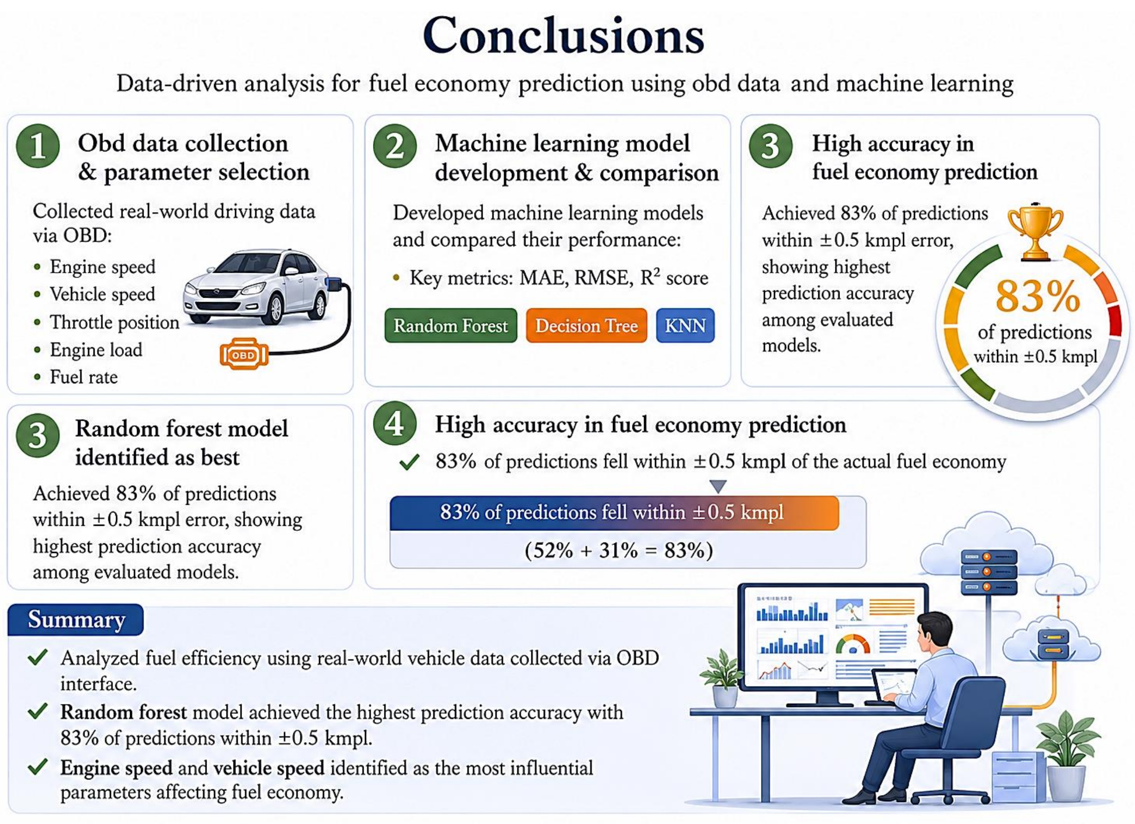

Figure 12. Overview of machine learning based fuel economy prediction using OBD data.

The performance of the selected model was evaluated using standard statistical metrics, including mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). The Random Forest model achieved a MAE of 0.32 kmpl, RMSE of 0.45 kmpl, and an R^2 value of 0.93, demonstrating strong predictive capability. The prediction accuracy analysis showed that 83% of the predicted fuel economy values fall within ±0.5 kmpl deviation from the actual measured values, indicating reliable model performance under real-world operating conditions.

Feature importance analysis further revealed that engine speed and vehicle speed are the most influential parameters affecting vehicle fuel efficiency. These findings highlight the effectiveness of using machine learning techniques combined with real-time OBD data for analysing vehicle performance and fuel consumption behaviour.

The proposed predictive monitoring approach provides a practical method for continuous vehicle performance evaluation and can support future applications in intelligent vehicle monitoring, driver behaviour analysis, and fuel efficiency optimisation using real-world operational data.

ABBREVIATIONS / NOMENCLATURE

Abbreviation	Description
AI	Artificial Intelligence
ECU	Electronic Control Unit

Abbreviation	Description
FE	Fuel Economy
KNN	K-Nearest Neighbors
MAE	Mean Absolute Error
ML	Machine Learning

OBD	On-Board Diagnostics
RF	Random Forest
RMSE	Root Mean Square Error
RPM	Revolutions Per Minute
R^2	Coefficient of Determination
BDA	Big Data Analytics
IoT	Internet of Things
CSV	Comma Separated Values
GPS	Global Positioning System

Acknowledgement

The author expresses sincere gratitude to Dr. S. Karthikeyan, Dr. A. Suresh, Dr. M. Pugazhvadivu, and Dr. R. Rajavel for their valuable research guidance, insightful technical discussions, and continuous encouragement throughout the course of this work. Their support during the stages of data collection, analysis, and interpretation has significantly contributed to the successful completion of this research.

The author also gratefully acknowledges AMET University for providing the academic environment, institutional support that enabled the successful completion of this study.

REFERENCES

1. Rykała M, Nowakowski T, Szpytko J. Modeling vehicle fuel consumption using a low-cost OBD-II interface. Energies. 2023;16(21):7266. https://doi.org/10.3390/en16217266
2. Michailidis ET, Panagiotopoulou A, Papadakis A. A review of OBD-II-based machine learning applications for sustainable, efficient, secure and safe vehicle driving. Sensors. 2025;25(13):4057. https://doi.org/10.3390/s25134057
3. Yoo S, Park J, Kim D. Machine learning based vehicle fuel efficiency prediction using big data analytics. Scientific Reports. 2025;15:96999. https://doi.org/10.1038/s41598-025-96999-0
4. Abediasl H, Al-Samarraie H, Ghodsi A. Real-time vehicular fuel consumption estimation using on-board diagnostics data. Proc Inst Mech Eng Part D J Automobile Engineering. 2024;238(5):845-857. https://doi.org/10.1177/09544070231185609
5. Molina-Campoverde JJ, García A, Varga I. Driving pattern analysis and gear shift classification using OBD-II data. Sensors. 2025;25(13):4043. https://doi.org/10.3390/s25134043
6. Keskin R, Arslan O. Bayesian LSTM based fuel consumption prediction using vehicle operational data. Sensors. 2025;25(22):7031. https://doi.org/10.3390/s25227031
7. Ping P, Qin W, Xu Y, Miyajima C, Takeda K. Impact of driver behavior on fuel consumption: Classification and prediction using machine learning. IEEE Access. 2019;7:78515-78532. https://doi.org/10.1109/ACCESS.2019.292229 4
8. Lattanzi E, Castellano N, Regini E. Machine learning techniques to identify unsafe driving behaviour using vehicle sensor data. Expert Systems with Applications. 2021;177:114952. https://doi.org/10.1016/j.eswa.2021.114952
9. Yao Y, Zhao X, Liu Y. Vehicle fuel consumption prediction based on driving behavior data using machine learning. Mathematical Problems in Engineering. 2020;2020:9263605. https://doi.org/10.1155/2020/9263605
10. Singh SK, Mishra R, Gupta P. Machine learning based classification of OBD-II data for driving behavior analysis. Bulletin of Electrical Engineering and Informatics. 2025;14(1):550-560. https://doi.org/10.11591/eei.v14i1.9398
11. Fugiglando U, Massaro E, Santi P. Driving behavior analysis through CAN-Bus data in uncontrolled environments. IEEE Trans Intelligent Transportation Systems. 2018;19(3):737-746. https://doi.org/10.1109/TITS.2017.2696351
12. Malik M, Tiwari S. Driving pattern analysis using OBD-II data and machine learning models. Materials Today Proceedings. 2023;72:3215-3221. https://doi.org/10.1016/j.matpr.2022.09.340

13. Cao Z, Wang J, Liu Y. Identification of high-emission vehicles using on-board diagnostics data analytics. Environmental Pollution. 2025;349:122084. https://doi.org/10.1016/j.envpol.2024.122084
14. Khan MAA, Ali MH, Haque AF. A machine learning approach for driver identification based on CAN-BUS sensor data. IEEE Access. 2022;10:80341-80351. https://doi.org/10.1109/ACCESS.2022.319558 3
15. Dong W, Li J, Yao R. Characterizing driving styles with deep learning. IEEE Trans Intelligent Transportation Systems. 2019;20(7):2510-2520. https://doi.org/10.1109/TITS.2018.2866052
16. Singh S, Kumar P. Driving pattern analysis using real-time ECU data collected through OBD-II. Int J Electrical and Computer Engineering Systems. 2022;13(2):189-198. https://doi.org/10.32985/ijeces.13.2.5
17. Chen Z, Wang Y. Data-driven vehicle energy consumption prediction using machine learning models. Applied Energy. 2019;235:701-712. https://doi.org/10.1016/j.apenergy.2018.10.094
18. Wang H, Zhang J. Real-time vehicle monitoring using OBD-II and IoT based analytics. IEEE Access. 2020;8:203456- 203467. https://doi.org/10.1109/ACCESS.2020.303509 7
19. Zhang Y, Wang L. Vehicle fuel consumption prediction using random forest regression models. Energy. 2019;181:888-897. https://doi.org/10.1016/j.energy.2019.05.090
20. Li X, Liu Z. Machine learning based vehicle performance evaluation using big data analytics. Transportation Research Part C. 2020;115:102621. https://doi.org/10.1016/j.trc.2020.102621
21. Park S, Kim H. Data driven vehicle diagnostics using machine learning algorithms. IEEE Trans Intelligent Transportation Systems. 2022;23(8):11321-11331. https://doi.org/10.1109/TITS.2021.3106024
22. Huang J, Li Y. Vehicle anomaly detection using OBD-II sensor data and machine learning. IEEE Access. 2021;9:75620-75632. https://doi.org/10.1109/ACCESS.2021.308164 2
23. Chen Y, Zhang H. Intelligent vehicle monitoring using OBD data and cloud analytics. Future Generation Computer Systems. 2021;115:52-60. https://doi.org/10.1016/j.future.2020.08.019
24. Wang Z, Chen H. Big data analytics for vehicle performance monitoring. Transportation Research Part C. 2018;95:451-463. https://doi.org/10.1016/j.trc.2018.07.012
25. Liu Y, Sun B. Machine learning based fuel efficiency estimation using vehicle operational data.Applied Energy 2021;298:117223. https://doi.org/10.1016/j.apenergy.2021.11722 3
26. Barbado A, Navarro C. Interpretable machine learning models for vehicle fuel consumption prediction. Engineering Applications of Artificial Intelligence. 2022;113:104908. https://doi.org/10.1016/j.engappai.2022.10490 8
27. Shirole V, Kulkarni P. Data driven driver behavior analysis using OBD-II sensor data. Journal of Intelligent Transportation Systems. 2025;29(3):245-259. https://doi.org/10.1080/15472450.2024.23087 12
28. Singh R, Patel D. Driving behavior classification using OBD-II speed and acceleration signals. Advances in Science Technology and Engineering Systems Journal. 2021;6(1):512-519. https://doi.org/10.25046/aj060160
29. Haghshenas SS, Chen J. Fuel consumption and CO_2 emissions prediction using vehicle telematics data. Cleaner Engineering and Technology. 2025;19:100742. https://doi.org/10.1016/j.clet.2024.100742
30. Selvam HP, Jayaprakash B, Li Y. Physics- informed machine learning for predicting engine emissions using OBD data. IEEE Access. 2025;13:22345-22358. https://doi.org/10.1109/ACCESS.2025.335112 0

