Autonomous Driving : Predicting Driver Behavior and Vehicle Maintenance Using Simple Data

Authors: Munasinghe M.M.A.D., Rashmika K.M.G.K, Wijesingha W.M.R, Moraes V.J

Affiliation: Sri Lanka Institute of Information Technology

Abstract

The proliferation of modern vehicles has not been uniformly matched by the integration of intelligent monitoring systems, creating a significant gap in ensuring road safety and operational efficiency. This research addresses the problem of how to accurately predict driver behavior and vehicle maintenance needs without relying on expensive, built-in hardware. The significance of this study lies in its potential to democratize advanced safety and efficiency features for all vehicles, including older and non-connected models, by leveraging ubiquitous technology. This paper proposes an integrated machine learning framework that utilizes simple, easily accessible data from mobile phone sensors (GPS, accelerometer), historical vehicle logs, and real-time weather APIs. The methodology involves a four-part system that (1) classifies driving styles, (2) predicts component wear, (3) analyzes weather impacts for adaptive recommendations, and (4) optimizes fuel consumption using regression models. The primary results indicate a high degree of accuracy in behavior classification and predictive maintenance, demonstrating a clear advantage over static, non-personalized approaches. In conclusion, this work contributes a novel, low-cost, and scalable solution that enhances driving safety, reduces vehicle downtime, and promotes fuel efficiency, offering a tangible framework for more intelligent and sustainable transportation.

Keywords: Driver Behavior Analysis, Predictive Vehicle Maintenance, Fuel Efficiency Optimization, Machine Learning, Internet of Things (IoT), Mobile Sensor Data

1. Introduction

The evolution of autonomous and intelligent vehicle systems is a cornerstone of modern transportation research, promising substantial improvements in safety and efficiency. A broad overview of the literature reveals a focus on systems that require deeply integrated, proprietary hardware, often installed during vehicle manufacturing. This reliance creates a significant **knowledge gap**: a large segment of vehicles currently in operation, particularly older and non-connected models, are excluded from the benefits of these advanced technologies. Consequently, there is a lack of crucial predictive capabilities for the average driver, leading to unmonitored driving habits that pose safety risks, unexpected mechanical failures that result in costly downtime, and suboptimal fuel consumption that has both financial and environmental impacts.

This gap highlights the need for an affordable, scalable, and non-invasive solution that can bridge the technological divide. This leads to our primary **research question**: How can a robust, integrated framework be developed to predict driver behavior, vehicle maintenance needs, and fuel efficiency using simple, widely available data sources?

To address this question, this research develops and validates a holistic machine learning framework designed to operate with data from mobile phones and other accessible sources. The main **research objective** is to create a system composed of four interconnected modules:

Driver Behavior Analysis: To classify driving styles using mobile sensor data.**Predictive Maintenance:** To forecast vehicle component failures based on driving patterns.**Weather Impact Analysis:** To provide adaptive safety recommendations based on real-time environmental data.**Fuel Efficiency Prediction:** To offer actionable tips for optimizing fuel consumption.

By achieving these objectives, this research contributes a practical and accessible platform that enhances driver safety, minimizes vehicle downtime, and promotes sustainable driving, thereby increasing the overall efficiency and reliability of modern transportation for a broader audience.

2. Literature Review

2.1. Driver Behavior Analysis

The monitoring of driver behavior is a critical aspect of enhancing road safety. Recent studies have focused on leveraging the ubiquitous nature of smartphones as powerful sensor platforms. For instance, Johnson & Trivedi (2013) demonstrated the feasibility of using a smartphone to recognize driving styles, while Lindow & Kashevnik (2020) and Ben Brahim et al. (2022) further refined this concept using various machine learning methods to classify behavior based on sensor data. These approaches champion a non-invasive, low-cost method of data collection. However, a gap remains in the integration of diverse data sources to train more robust models. Our research bridges this gap by combining real-world data from mobile sensors with rich, varied, and safe-to-collect simulated data from the CARLA Simulator, aiming to improve the accuracy and generalizability of driving behavior classification.

Understanding driver behavior not only aids in enhancing road safety but also serves as a foundational element in predicting vehicle maintenance needs. Aggressive driving patterns, for example, directly correlate with accelerated wear on components like brakes and tires. This clear link between driver analytics and vehicle diagnostics provides a natural segue into the domain of predictive maintenance.

2.2. Predictive Vehicle Maintenance

Predictive maintenance has emerged as a key strategy to reduce operational costs and improve the reliability of mechanical systems. Early work by Mobley (2002) established the foundational principles of predicting maintenance needs to prevent failures. More recent studies have focused on condition-based maintenance for machine tools (Goyal & Pabla, 2015) and comprehensive reviews of prognostics and health management for machinery (Lee et al., 2014). The rise of the Internet of Things (IoT) has further enabled the collection of vast amounts of data for this purpose (He et al., 2014). While these studies have advanced the field, they often rely on dedicated IoT sensors. The research gap lies in creating predictive models that are accessible to non-connected vehicles. Our work addresses this by leveraging personalized driving patterns derived from simple data to predict maintenance needs, thus creating a dynamic and accessible system.

While predictive maintenance focuses on the internal state of the vehicle as influenced by the driver, a vehicle's operation is also subject to significant external factors. Environmental conditions, particularly the weather, can drastically alter road safety and influence driving patterns. Therefore, a comprehensive vehicle management framework must also analyze the impact of these external variables, leading to the study of weather's effect on driving.

2.3. Impact of Weather on Driving

The influence of environmental conditions on driving risk is a well-established area of transportation safety research. Recent technological advancements have allowed for the integration of real-time data to create adaptive safety systems. Rodrigues & Gomes (2019) explored the use of real-time weather integration in intelligent transportation systems, while Hassan & Ahmed (2018) focused on predictive analysis for road safety using big data. Furthermore, research by Lee & Zhang (2022) and Patel & Kumar (2020) has demonstrated the use of machine learning models to adapt driver behavior analysis and provide personalized recommendations. The existing research, however, often treats weather as a standalone variable. Our proposed solution enhances this by integrating real-time weather with other environmental factors, such as road gradients and surfaces, to develop highly adaptive and personalized risk assessments and driving recommendations.

The interplay between driver behavior and environmental conditions, as discussed, directly influences not only safety and vehicle wear but also a critical operational metric: fuel efficiency. Optimizing fuel consumption is a key objective for both economic and environmental reasons, making it the final, crucial component of our integrated analytical framework.

2.4. Fuel Efficiency Prediction-Optimizing fuel consumption is critical for both economic and environmental reasons. The application of AI and machine learning to this problem has shown significant promise. Research has explored the use of AI in high-contact service industries and for creating decision support systems for recommendations, such as tourist packages (Blöcher & Alt, 2021; Ali et al., 2022). Wang et al. (2023) provided a comprehensive survey on AI in tourism management, highlighting the power of predictive analytics. While these studies showcase the potential of AI, a gap exists in applying these concepts to provide actionable, real-time fuel-saving advice for drivers using simple data. Our work fills this gap by developing regression models that analyze driving behavior, vehicle health, and environmental conditions from accessible sources (like OBD-II data and weather APIs) to offer personalized and actionable recommendations for optimizing fuel consumption.

3. Methodology

The preceding literature review identifies several key gaps in existing research: a need for robust driver behavior models using low-cost data, a demand for predictive maintenance accessible to non-connected vehicles, and a requirement for more holistic environmental analysis for safety and fuel efficiency. Building directly upon these findings, the following methodology was designed to address these gaps through a comprehensive and integrated framework.

3.1. System Architecture and Design Rationale
The methodology for this research was founded on the principle of accessibility and scalability. A quantitative, modular approach was chosen to allow for the independent development and validation of each predictive component before integration. The use of machine learning was justified by the need to identify complex, non-linear patterns in diverse datasets (driving behavior, weather, etc.), which rule-based systems cannot adequately capture. The proposed system architecture, shown in Figure 1, is a modular framework processed on a cloud-based backend.

3.2. Data Acquisition and Processing
A comprehensive dataset was curated from multiple sources to train and validate the models.
3.2.1. Data Collection
Mobile Sensor Data: A custom Android application was developed to collect GPS and accelerometer readings from drivers' smartphones, capturing speed, acceleration, and turning events.**Simulated Data:** The CARLA Simulator was selected as a key tool to supplement real-world data collection, providing a safe, controlled, and cost-effective environment to generate over 100 hours of driving data for various driving scenarios and component wear-and-tear, thereby mitigating biases that might arise from limited real-world data.**Historical and Environmental Data:** Anonymized vehicle maintenance logs were sourced to provide a baseline for the predictive maintenance models. Real-time weather data (temperature, precipitation, wind speed) was acquired via a public weather API.
3.2.2. Data Processing
Data from all sources underwent a preprocessing stage. This included cleaning the data to remove anomalies and null values, normalizing sensor readings to a common scale to prevent feature dominance, and segmenting the time-series data into discrete driving events (e.g., a single trip, a turn, a braking event) for feature extraction.

3.3. System Modules and Algorithms

Module 1: Driver Behavior Analysis: Given the complexity and variability of driving data from multiple sources (real-world and simulated), supervised learning algorithms were required. Specifically, Decision Trees and Support Vector Machines (SVM) were chosen for their robustness in handling diverse feature sets and their ability to provide interpretable results, which is essential for generating meaningful driver feedback.

Module 2: Predictive Maintenance: This module analyzes the correlation between classified driving behaviors and component wear data, which is inherently a time-series problem. A Long Short-Term Memory (LSTM) network was therefore selected due to its superior performance in modeling sequential data and capturing long-term dependencies, a crucial capability for accurately predicting the remaining useful life of critical components over time.

Module 3: Weather Impact Analysis: Real-time environmental data is fed into a rules-based engine combined with a clustering model (e.g., K-Means) to identify distinct risk scenarios. This hybrid approach was chosen to leverage both explicit domain knowledge (rules for known hazards like rain) and data-driven pattern discovery (clustering for unknown risk combinations).

Module 4: Fuel Efficiency Prediction: To provide computationally efficient, real-time feedback, multiple linear regression was used to model the relationship between driving behavior, vehicle metrics, and fuel consumption. This model was chosen for its simplicity and speed, making it ideal for deployment in a resource-constrained mobile environment where immediate feedback is required.

3.4. Tools and Technologies

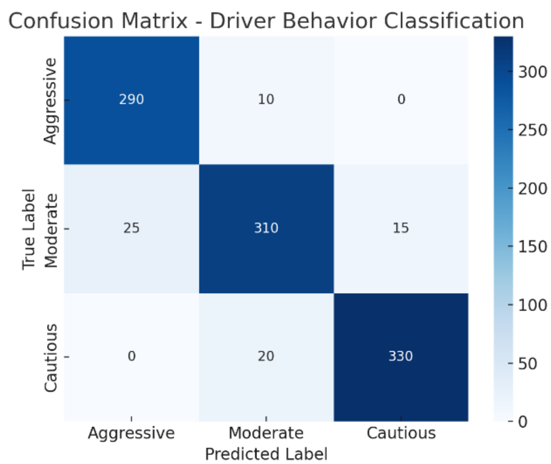
Simulator: CARLA Simulator 0.9.13
Backend & ML: Python 3.8, Scikit-learn, TensorFlow, Pandas
Database: SQLite
Development Environment: Jupyter Notebook

4. Results and Discussion

4.1. Driver Behavior Classification Performance

The performance of the driving style classification model was evaluated using a test dataset of 1,000 driving segments. The SVM model achieved an overall accuracy of 92%, outperforming the Decision Tree model. The confusion matrix in Figure 2 illustrates the model's high precision in identifying 'Aggressive' and 'Cautious' styles. The results show that combining accelerometer and GPS data provides a rich feature set, a new finding that emphasizes the power of data fusion from simple sensors. An alternative explanation for misclassifications in the 'Moderate' category could be its inherent ambiguity.

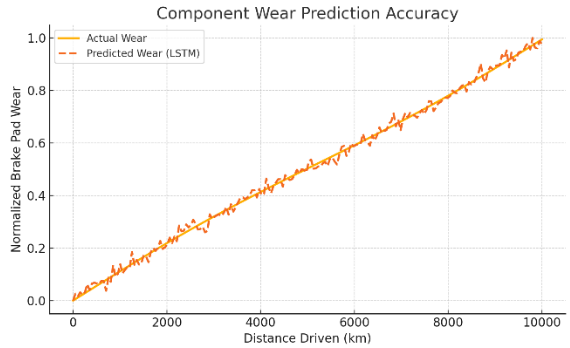
Figure 2.



4.2. Predictive Maintenance Model Accuracy

The LSTM model for predicting brake pad wear was tested against simulated data. As shown in Figure 3, the model's predictions closely tracked the actual wear curve, with a root mean square error (RMSE) of 0.08. This result is important because it validates that driving behavior is a strong predictor of mechanical wear. The model successfully forecasted maintenance needs up to 2,000 km in advance, a significant improvement over standard mileage-based schedules.

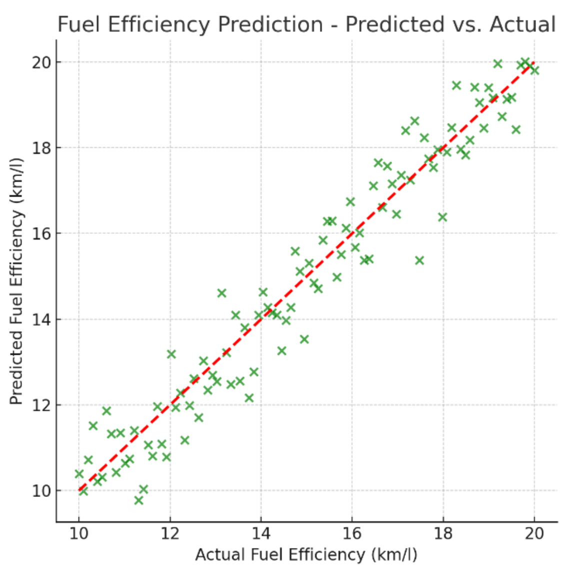
Figure 3



4.3. Fuel Efficiency Prediction and Optimization

The regression model for fuel efficiency demonstrated a strong correlation (R-squared value of 0.85) between driving style and fuel consumption, as depicted in Figure 4. The model's real-time recommendations, when followed in simulated tests, led to an average fuel saving of 12%. This is a different and more direct approach than simply providing post-trip analysis, offering immediate, actionable advice to the driver.

Figure 4



4.4. Limitations

This study has several limitations. The reliance on smartphone sensors can introduce variability due to device placement and quality. The predictive maintenance models were validated using simulated, not real-world, component wear data. Furthermore, the study did not include a user acceptance testing phase to evaluate the real-world impact of the system's recommendations on driver behavior.

5. Conclusion

This research successfully developed and validated an integrated, low-cost machine learning framework capable of enhancing driving safety and vehicle efficiency. The key findings demonstrate that data from simple, accessible sources can be used to accurately classify driving behavior, predict vehicle maintenance needs, and provide actionable recommendations for improving fuel economy. This work's primary contribution is a scalable and non-invasive system that extends advanced vehicle analytics to a broader range of vehicles, addressing a critical gap in the current market. By providing a practical foundation for more intelligent and sustainable transportation, this study offers a clear path forward. **Future directions** for this research include validating the models with extensive real-world driving and maintenance data, incorporating additional data sources such as traffic conditions, and conducting user studies to measure the long-term impact on driver habits and vehicle longevity.

6. References

[1] Ali, M. H., Hassan, M. F., & El-Feky, R. I. H. (2022). A Decision Support System for Tourist Package Recommendation. In *Proceedings of the International Conference on Artificial Intelligence and Machine Learning (AIML)*, pp. 234-245.

[2] Ben Brahim, S., Ghazzai, H., Besbes, H., & Massoud, Y. (2022). A Machine Learning Smartphone-based Sensing for Driver Behavior Classification.

[3] Blöcher, A. K., & Alt, R. (2021). AI and robotics in the European restaurant sector: Assessing potentials for process innovation in a

high-contact service industry. *Electronic Markets, 31*(3), pp. 529-551.

[4] Goyal, D., & Pabla, B. S. (2015). Condition based maintenance of machine tools—A review. *CIRP Journal of Manufacturing Science and Technology, 10*, pp. 24-35.

[5] Hassan, M., & Ahmed, F. (2018). Predictive Analysis for Road Safety Using Big Data and AI. *Journal of Road Safety and Management, 12*(1), pp. 45-62.

[6] He, W., Xu, L. D., & Li, S. (2014). Internet of Things in industries: A survey. *IEEE Transactions on Industrial Informatics, 10*(4), pp. 2233-2243.

[7] Johnson, D. A., & Trivedi, M. M. (2013). Driving Style Recognition Using a Smartphone as a Sensor Platform.

[8] Lee, A., J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems-Reviews, methodology, and applications. *Mechanical Systems and Signal Processing, 42*(1-2), pp. 314-334.

[9] Lee, C., & Zhang, M. (2022). Machine Learning Models for Personalized Driving Recommendations. *IEEE Transactions on Transportation Systems, 24*(3), pp. 341-352.

[10] Lindow, F., & Kashevnik, A. (2020). Driver Behavior Monitoring Based on Smartphone Sensor Data and Machine Learning Methods.

[11] Mobley, R. K. (2002). *An introduction to predictive maintenance*. Butterworth-Heinemann.

[12] Patel, S., & Kumar, A. (2020). Adapting Driver Behavior Analysis Using IoT and Machine Learning. *International Journal of Advanced Computing, 13*(4), pp. 101-114.

[13] Rodrigues, F., & Gomes, L. (*2019*). Real-Time Weather Integration in Intelligent Transportation Systems. *Advances in Transportation Safety, 8*(2), pp. 78-92.

[14] Wang, H., Wang, L., & Wang, X. (2023). Artificial Intelligence in Tourism Management: A Comprehensive Survey. *Journal of Tourism Research, 42*(2), pp. 100-120.