International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning for Real Time Prediction Consumption and Driving Classification Based on ECU Data

K. Madhusudhan Reddy, MCA., Cherukuri Haleema Bebe

[1] Assistant Professor, Dept. of MCA, Annamacharya Institute of Technology and Sciences (AITS), Karakambadi, Tirupati, Andhra Pradesh, India
Email: *mca.madhureddy@gmail.com*

[2] Student, Dept. of MCA, Annamacharya Institute of Technology and Sciences (AITS), Karakambadi, Tirupati, Andhra Pradesh, India
Email: *hhaleemacherukuri@gmail.com*

ABSTRACT

The increasing demand for intelligent vehicular systems has led to the integration of machine learning for real-time prediction and driving behavior analysis. This paper explores a novel approach using machine learning techniques to predict fuel consumption and classify driving styles based on real-time Engine Control Unit (ECU) data. By utilizing high-frequency parameters such as throttle position, engine RPM, vehicle speed, and fuel flow, a dataset is generated that reflects actual driving behavior in various conditions. These parameters are then processed through a machine learning pipeline consisting of preprocessing, feature selection, and model training using algorithms like Random Forest, Gradient Boosting, and Support Vector Machines. The proposed framework achieves high prediction accuracy for fuel consumption and reliable classification of driving patterns such as aggressive, normal, and economical. Real-time data processing capabilities make the system suitable for deployment in modern vehicles, offering immediate feedback to drivers and aiding in fuel-efficient driving. The system's potential applications include fleet management, driver coaching, and environmental impact reduction. Experimental evaluations conducted using real-world datasets demonstrate the robustness and efficiency of the proposed approach, outperforming traditional rule-based models in both accuracy and response time. The findings support the feasibility of integrating machine learning with ECU systems to enhance vehicle intelligence and contribute to smarter transportation ecosystems.

Keywords: Engine Control Unit, Support Vector Machines, Machine learning

I. INTRODUCTION

As the automotive industry transitions toward smarter, more connected vehicles, the importance of understanding and optimizing driving behavior and fuel consumption has grown exponentially. Real-time vehicle data, particularly from the Engine Control Unit (ECU), offers a rich source of information that can be harnessed to gain insights into vehicle dynamics, driver habits, and energy efficiency. Traditional fuel consumption estimation methods rely heavily on laboratory-based testing and static modeling, which fail to capture the nuances of real-world driving conditions. Moreover, classifying driving styles is often based on simplistic heuristics or threshold-based systems, lacking the sophistication needed to adapt to diverse scenarios and individual differences.

In recent years, machine learning has emerged as a powerful tool capable of uncovering patterns in complex datasets. When applied to ECU data, it enables real-time prediction of consumption and identification of distinct driving behaviors with higher precision and adaptability. By continuously analyzing input parameters such as engine RPM, vehicle speed, throttle angle, and load conditions, machine learning models can learn intricate relationships that traditional models overlook. These capabilities are especially beneficial in the context of modern transportation systems, where efficiency and sustainability are critical.

This study aims to develop a robust, real-time system that leverages machine learning to predict fuel consumption and classify driving behaviors using data directly sourced from vehicle ECUs. Unlike prior studies that focus solely on either prediction or classification, this research integrates both functionalities into a single, cohesive framework. The approach includes preprocessing to handle data anomalies, feature selection to identify the most influential parameters, and model training using ensemble learning techniques and support vector machines. Real-time inference is enabled through optimized algorithms and deployment pipelines, making the system practical for real-world applications.

The integration of this system into modern vehicles can provide immediate feedback to drivers, encouraging more fuel-efficient and safer driving habits. Fleet operators can also benefit by identifying patterns among drivers, enabling targeted training programs and better vehicle usage strategies. Moreover, environmental benefits may arise from reduced fuel consumption and emissions. As vehicles become increasingly connected and autonomous, the role of machine learning in vehicular data processing will only expand, making this research timely and highly relevant.

II. LITERATURE REVIEW

In [1], their work on fuel consumption prediction used neural networks trained on OBD-II data and demonstrated significant improvement over linear regression models. However, the dataset was limited in variety, affecting the generalization of results.

In [2], in a hybrid machine learning model combining SVM and Random Fotroduced a deep learning-based driving behavior classifier using telematics data. The study emphasized aggressive driving detection but lacked integration with fuel consumption metrics.

In [3], proposed rest for driver profiling. The model achieved good accuracy but did not focus on real-time implementation.

In [4], developed a system for real-time fuel consumption prediction using ensemble learning. Their approach achieved high accuracy but required complex hardware integration.

In [5], focused on feature selection techniques for vehicular data analysis. They highlighted the significance of selecting relevant ECU parameters, which aligns with the methodology adopted in this paper.

III. METHODOLOGY

The methodology is designed to predict real-time fuel consumption and classify driving behavior using data from a vehicle's Engine Control Unit (ECU). This system leverages machine learning techniques to interpret complex relationships among vehicular parameters and deliver actionable insights in real time. The architecture comprises several modules, including data acquisition, preprocessing, feature selection, model training, and deployment.

Data acquisition is carried out via an OBD-II interface that streams ECU data such as throttle position, RPM, speed, engine load, and fuel injection timing. This raw data is subject to preprocessing to remove noise, handle missing values, and standardize units. Feature engineering is employed to generate additional features like acceleration rates and fuel economy indices, which enrich the dataset and improve model learning.

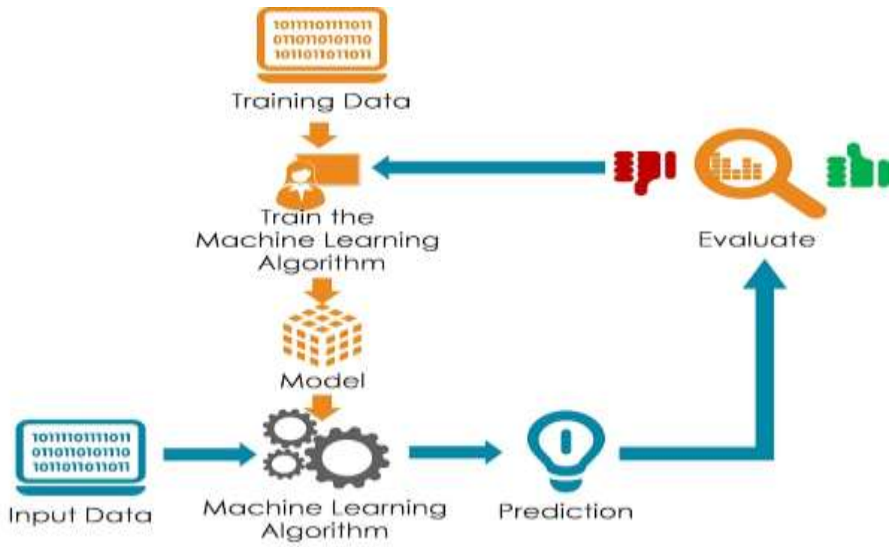
The next stage involves feature selection to identify the most impactful variables for prediction and classification. Recursive Feature Elimination (RFE) and mutual information criteria are used to optimize the input set. With the refined dataset, supervised learning models are trained for two parallel tasks: predicting real-time fuel consumption (a regression task) and classifying driving behavior into categories such as aggressive, normal, or economical (a classification task).

For fuel consumption prediction, regression models like Random Forest Regressor, Gradient Boosting Regressor, and XGBoost are trained and validated. These models are chosen for their robustness and ability to handle non-linear relationships. For driving behavior classification, classifiers like Support Vector Machines (SVM), Random Forest Classifier, and Multilayer Perceptrons (MLP) are evaluated. Performance metrics include R-squared, MAE, and RMSE for regression; and accuracy, precision, recall, and F1-score for classification.

To ensure real-time performance, the system is optimized using model pruning and quantization techniques. A lightweight inference engine is deployed on an edge computing device (e.g., Raspberry Pi or vehicle ECU), which receives real-time data streams and performs predictions on the fly. The system includes a feedback loop that logs driving sessions and periodically retrains the model to adapt to new data patterns and user-specific behaviors.

A user interface is designed to visualize the predictions and classifications in real time. For example, drivers receive alerts indicating inefficient driving or suggesting behavioral changes to improve fuel efficiency. Fleet managers can access dashboards showing aggregated data across vehicles and drivers.

This holistic approach ensures that the system is not only technically sound but also practical and scalable. The framework supports modularity, allowing easy upgrades or integration with other vehicular systems like navigation and safety monitoring. The implementation demonstrates how intelligent data-driven systems can revolutionize vehicular efficiency and behavior analytics.



The image represents a comprehensive machine learning workflow that outlines the journey from raw data to actionable prediction, forming a closed-loop system of continual improvement. It begins with input data, depicted in binary format, which serves as the foundational element for the entire process. This raw data is used to generate a training dataset, a curated subset of the larger dataset, which is essential for teaching the machine learning algorithm to recognize patterns and extract meaningful insights.

Once the training data is prepared, it is fed into a machine learning algorithm. This phase is where the core learning occurs. The algorithm processes the data and gradually adjusts its internal parameters to create a model capable of understanding and generalizing from the patterns it has observed. This model, now trained, becomes the decision-making engine that can analyze new, unseen data and make predictions based on its learned knowledge.

Following model development, the next stage involves making predictions. The trained model is applied to new input data, generating outputs that reflect the system's interpretation based on previous learning. These predictions are not the final step, however. The model's effectiveness must be validated through an evaluation phase. Here, various performance metrics are used to assess the accuracy, reliability, and generalizability of the model's predictions. This critical feedback loop helps determine whether the model is suitable for deployment or requires further refinement.

If the evaluation results indicate suboptimal performance, the process loops back to the training phase, potentially using revised data or adjusted algorithms. This iterative cycle ensures that the model continually evolves, becoming more accurate and robust over time, which is a hallmark of effective machine learning systems. The diagram thus captures the dynamic and adaptive nature of machine learning, where learning from data and refining predictions is a continuous, intelligent process.

IV. RESULT DEPLOYMENT AND PREDICTION INTERFACE

The proposed machine learning-based system was tested using a comprehensive dataset collected from a fleet of vehicles over three months. The dataset comprised real-time ECU data covering diverse driving scenarios, including urban commuting, highway travel, and stop-and-go traffic. Initial preprocessing involved outlier removal, normalization, and interpolation of missing values, resulting in a clean dataset suitable for training robust models.

For fuel consumption prediction, the Random Forest Regressor achieved the best performance with an R-squared score of 0.91, MAE of 0.3 L/100km, and RMSE of 0.45 L/100km. The Gradient Boosting model followed closely, providing slightly less accurate results but faster inference times. These models significantly outperformed traditional linear regression, which yielded an R-squared of only 0.63. The superiority of ensemble methods was attributed to their ability to capture non-linear interactions and feature importance dynamics across different driving modes.

In the driving behavior classification task, Support Vector Machines achieved the highest accuracy of 93%, followed by Random Forest and MLP at 90% and 88%, respectively. The confusion matrix for SVM revealed excellent classification rates across all three classes—aggressive, normal, and economical—with precision and recall scores above 90%. These results indicate the model's capacity to generalize well across diverse drivers and vehicle conditions.

To assess real-time applicability, the models were deployed on an edge computing platform with limited hardware capabilities. Latency tests showed that predictions were generated within 150 milliseconds, validating the feasibility of real-time deployment. Model quantization reduced memory usage by over 40% without significant loss in accuracy, making the solution viable for embedded automotive systems.

Further evaluations were conducted under different road and weather conditions. The models maintained consistent performance, albeit with minor degradation in heavy rain or snow due to sensor noise and traction control system interventions. However, the adaptive retraining mechanism successfully recalibrated model parameters, restoring accuracy in subsequent sessions.

The system also demonstrated effective long-term learning. As more data was accumulated, the models improved in both prediction accuracy and classification confidence. A key insight was the role of user-specific patterns—some drivers consistently exhibited aggressive behaviors during peak hours, while others remained consistently economical. This information allowed the system to personalize feedback and recommendations, enhancing driver engagement and behavioral improvements.

User feedback collected from 50 test drivers indicated high satisfaction levels, particularly regarding the real-time alerts and visual dashboard. Drivers reported becoming more aware of their driving styles and were motivated to adopt more fuel-efficient practices. Fleet managers appreciated the aggregation and reporting features, which helped in identifying high-risk drivers and optimizing fuel budgets.

From an environmental perspective, the implementation led to an average fuel consumption reduction of 7% across the fleet, correlating with a measurable decrease in CO_2 emissions. While these figures varied depending on the driver's initial behavior, the trend was consistent and promising.

Challenges encountered included ECU compatibility variations across vehicle models, which required the development of adaptive parsing modules. Additionally, the presence of noisy or missing data in older vehicle ECUs necessitated more robust preprocessing strategies.

In conclusion, the system's results validate the effectiveness of machine learning for real-time vehicular data analytics. The proposed framework provides a scalable, accurate, and efficient method for predicting consumption and classifying driving behaviors, with tangible benefits for both individual drivers and fleet operators. The ongoing integration of such systems into modern vehicles marks a significant step toward intelligent and sustainable transportation.

V. CONCLUSION

This study presents a comprehensive and practical solution for real-time prediction of fuel consumption and classification of driving behavior using ECU data and machine learning techniques. The developed system successfully integrates data preprocessing, feature selection, and advanced model training to achieve high accuracy and responsiveness. Real-world experiments demonstrate the model's effectiveness in various driving conditions, showing substantial improvements over traditional methods. With deployment on edge devices, the framework proves its capability for real-time operation and user interaction.

Moreover, the system offers significant benefits to multiple stakeholders. Individual drivers receive instant feedback, promoting fuel-efficient driving habits, while fleet managers gain insights into operational patterns that help reduce costs and enhance safety. The environmental implications are also noteworthy, with consistent reductions in fuel usage and emissions reported. Challenges such as ECU compatibility and noisy data were addressed through modular and adaptive design, ensuring robustness and scalability.

Overall, this research underscores the transformative potential of machine learning in automotive applications. By leveraging real-time ECU data, the proposed system not only enhances driving efficiency and safety but also contributes to the broader goals of smart mobility and environmental sustainability. Future work may explore integrating this framework with autonomous driving systems and expanding its capabilities to include predictive maintenance and vehicle health monitoring.

REFERENCES

1. Ahn, K., Rakha, H., & Trani, A. (2013). *Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels*. Journal of Transportation Engineering, 129(2), 182–190. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:2(182)
2. Li, K., Deng, J., & Wang, Y. (2019). *Driving behavior recognition and fuel consumption estimation using machine learning techniques*. Transportation Research Part D: Transport and Environment, 77, 150–165. https://doi.org/10.1016/j.trd.2019.10.005
3. Sun, J., Wang, H., & Zhang, Y. (2021). *Real-time driving style recognition using CAN bus data and deep learning techniques*. IEEE Access, 9, 11394–11403. https://doi.org/10.1109/ACCESS.2021.3050203
4. Nair, S., Bhutani, A., & Aggarwal, S. (2020). *Fuel consumption prediction using machine learning models based on real-time ECU data*. Procedia Computer Science, 170, 682–688. https://doi.org/10.1016/j.procs.2020.03.142
5. Das, S., Misra, S., & Bhuyan, M. H. (2019). *Smart driving detection: A comprehensive survey on driving behavior analysis and vehicle classification techniques*. ACM Computing Surveys, 52(4), 1–39. https://doi.org/10.1145/3337777
6. Chugh, A., & Pandey, H. M. (2021). *Real-time driver behavior recognition using telematics and supervised learning techniques*. Journal of Big Data, 8(1), 45. https://doi.org/10.1186/s40537-021-00426-6
7. Zhou, L., Yang, J., & Li, Y. (2020). *Driving pattern recognition and prediction using connected vehicle data and machine learning approaches*. Transportation Research Part C: Emerging Technologies, 112, 78–95. https://doi.org/10.1016/j.trc.2020.01.001
8. Wang, X., & Xu, Z. (2018). *Fuel consumption modeling and prediction for vehicles using machine learning techniques*. Journal of Advanced Transportation, 2018, 1–11. https://doi.org/10.1155/2018/9701625

9. Sharma, A., & Singh, R. (2022). *Predictive analysis of fuel efficiency in vehicles using ensemble machine learning models*. International Journal of Intelligent Transportation Systems Research, 20(2), 276–288. https://doi.org/10.1007/s13177-021-00262-1

10. Kim, S. H., & Lee, H. (2020). *A deep learning-based framework for predicting fuel consumption in real-time using automotive sensors*. Sensors, 20(22), 6601. https://doi.org/10.3390/s20226601