iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** XI **Month of publication:** November 2026

DOI: https://doi.org/10.22214/ijraset.2026.85085

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue XI Nov 2026- Available at www.ijraset.com

Beyond the Racetrack: Evaluating the Transferability of Formula 1 AI and Telemetry Technologies to Passenger Vehicles

Prathamesh Bhamare[1], Prof. Sheetal Shevkari[2], Mahesh Rawate[3], Prajwal Mane[4], Akshata Powar[5], Aishwarya Navagune[6], Aarti Tonpe[7]
[1, 2, 4, 5, 6, 7]*Student,* [2]*Assistant Professor, MIT Arts, Commerce, and Science, Alandi, Pune*

Abstract: ***This research investigates the level of transferability of selected Artificial Intelligence (AI) and Telemetry (TM) techniques from Formula 1 to passenger cars which are equipped with lower number of sensors and computing power. The research is not to copy a full fledged Formula 1 telemetry system, but to take over its analytical method and to use consumer grade OBD-II telemetry to deliver useful predictions. After data quality checks and duplicate identification, three OBD-II driving datasets, with a total number of 555,000 raw observations were analysed. Binary driver-state classification was conducted with Logistic Regression (as a baseline model) and with the nonlinear ensemble models Random Forest, Extra Trees, LightGBM and XGBoost. Due to the imbalance between the classes, the performance of the model was assessed using accuracy, macro-precision, macro-recall, Macro-F1 score and confusion-matrix analysis. The best accuracy on any one of the individual driver datasets was 99.9873%. The highest accuracy (99.9949%) and Macro-F1 (99.9924%) were obtained by the best models for the combined dataset. The driver datasets reported the greatest relationships with the classification target, with throttle position, accelerator-pedal position, engine load, and intake-related variables showing the strongest relationships. The results indicate that a rather modest sensor suite can be used to produce useful vehicle intelligence. The results should not be viewed as an immediate demonstration of how ready the technology is to be used in the field, however. Telemetry for the vehicles is sequential, meaning that the observations in the randomly split training and test sets can have a high degree of correlation. Other tests such as driver-wise, trip-wise, time-based and cross-vehicle tests are thus required for additional validation. Both tire and brake wear prediction are mentioned as future research directions, rather than results of experiments, since there do not exist any validated physical measurements of degradation from the available data sets. This paper focuses on the application of machine learning techniques in driver behavior classification, predictive maintenance, vehicle telemetry, OBD-II and CAN bus, and tire/brake degradation for application in Formula 1, and introduces the concepts of Random Forest and XGBoost models.***

Keywords: Formula 1, vehicle telemetry, OBD-II, CAN bus, machine learning, driver behavior classification, predictive maintenance, tire degradation, brake degradation, XGBoost, Random Forest, passenger vehicles.

I. INTRODUCTION

Formula 1 is a very data intensive racing environment where teams are constantly analysing vehicle performance in dynamic and challenging environments. These technologies include telemetry, simulations, statistical analysis, and machine learning to gain insight on tire characteristics, braking capabilities, fuel efficiency, vehicle balance and Race Strategy. Data can be utilized only after a race, but they can also be utilized while a race is in progress as well. Being able to process vast amounts of car data and present it speedily and accurately in a format that's useful to the driver, crew, and NASCAR race officials is thus a key aspect of today's Formula 1 engineering.

But Formula 1's tech world is quite unlike the typical roadgoing car. With specialized sensors, dedicated data engineers, advanced computing infrastructure, and well-controlled analytical processes, F1 teams can leverage their expertise to gain a more comprehensive understanding of their vehicles. Specialized sensors, dedicated data engineers, advanced computing infrastructure, and well-controlled analytical processes can help F1 teams gain a more comprehensive understanding of their vehicles. Mass-market vehicles, on the other hand, have a smaller number of sensors and computing capabilities. They also face differences in the way the cars are maintained, driven, on the road, and used. Therefore, it is not feasible to directly convert an entire Formula 1 telemetry system to a standard passenger car, making it neither technically feasible nor cost-effective.

©IJRASET: All Rights are Reserved

This research project thus aims at the identification of those parts of Formula 1 vehicle intelligence that can be practically transferred to passenger vehicles by means of easily available OBD-II telemetry. The experimental aspect is related to two-class driver states (driver versus non-driver). The available OBD-II datasets do not provide validated measurements of tire wear, brake-disc temperature or brake-pad thickness which makes the degradation of tires and brakes a transferability challenge and a modeling direction for the future.

Three open-source OBD-II driving datasets are employed and Python-based machine-learning techniques are used. Logistic Regression is compared to Random Forest, Extra Trees, XGBoost and LightGBM, to this problem, which is a nonlinear structured-data classification problem. Apart from the assessment of the predictive performance, the research suggests the three-stage system of Basic, Intelligent and Advanced vehicle intelligence. This framework takes into account model accuracy as well as availability of the sensors, computational load, cost, validation requirements and practical deployment considerations.

Everything in Formula 1 is not able to be transferred, though what is is important. Rather, the analytical process: data collection, data cleaning of signals, extraction of meaningful features, implementation of predictive models, results interpretation, and decision support based on the obtained results. The results of such a study should not, therefore, be interpreted as proof that an analytical method can be transferred from a Formula 1 car to a passenger vehicle and can convert all of the telemetry data.

II. LITERATURE REVIEW

A. *Formula 1 and Data Analytics*

Data is a major factor for improving Formula 1 cars' efficiency and racing strategy.Teams analyze telemetry data, simulation results; tire wear; braking performance; fuel consumption; aerodynamics. Machine learning models are useful as they have many of these variables that interact nonlinearly. Random Forest and XGBoost are thus suitable models for predicting structured vehicle and in, performance data as demonstrated by [12]-[16]. Some of the previous F1 studies have achieved satisfactory results in applications such as tire degradation and pit window prediction [11].

Nevertheless, these findings cannot be directly compared to this research because the data sources, as well as variables, targets and validation methods are not similar enough. The results of the present study cannot be directly compared with those of this study, however, due to variations in data source, variables, target and validation methods.

B. *OBD-II and Consumer Vehicle Data*

OBD-II provides valuable vehicle data like speed, engine RPM, load, throttle position, temperature and diagnostics. Consumer vehicle telemetry, however, isn't anywhere near as detailed as Formula 1 telemetry.

There are some high-tech f1 data which cannot be acquired from conventional vehicles. Hence, some programs require proxy variables that are inspired from f1. For instance, brake stress can be estimated by changes or through speed variations, repeated braking events, road conditions and vehicle characteristics instead of directly measuring brake temperature/wear.

C. *Driver Behaviour and Predictive Maintenance*

Public data sets such as UAH-DriveSet can be examined to investigate driving state classification such as normal driving, aggressive driving and sleepy driving.

Likewise, OBD-II and CAN-bus data have been extensively utilized for driving style analysis, fuel prediction, fault detection and predictive maintenance [17]-[31]. The results of these studies demonstrate the feasibility of using standard vehicle telemetry for practical applications of machine learning.

However, one should always be aware of sensors' availability, computational cost, inference time, reliability, and interpretability when implementing. Likewise, OBD-II and CAN-bus data have been extensively used for driving style analysis, fuel prediction, fault detection, and predictive maintenance [17]-[31].

D. *Research Gap*

These days, the focus is on racing strategy and telemetry for f1 cars, and diagnosis, driving behaviour, fuel consumption and maintenance requirements for consumer vehicles. Real world deployment, however, will need to take into account sensor availability, computational needs, inference speed, reliability, and interpretability.

This paper aims to fill this gap by. Defining f1 analytical concepts that will be practically possible to be implemented in consumer vehicles.Evaluating OBD ii based driver state classification based on test results. Connect by linking models' effectiveness to real world needs, not just precision.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

This study fills this void by

1) Identification of realistic F1 analytical concepts that can be applied to consumer vehicles.
2) Reviewing experimental results on driver-state classification with OBD-II.
3) Linking models to the challenges of implementation rather than focusing on accuracy.

III. RESEARCH OBJECTIVES

1) To determine the most pertinent Formula 1 AI and telemetry features for mass-market passenger cars.
2) Create and evaluate Logistic Regression, Random Forest, Extra Trees, XGBoost and LightGBM models for the binary classification of driver state from public open source OBD-II data.
3) To identify the strongest relationship of OBD-II telemetry variables with the classification target.
4) To perform comparisons based on the F1-oriented benchmarks used in the literature, but while explicitly taking into account differences in the tasks and the validation protocol.
5) To create a 3-tiered classification of passenger-vehicle intelligence based on sensing, computation, cost, analytical capability and validation requirements.
6) To determine what extra data and validation is needed before tire and/or brake degradation proxies can be reliably used in passenger-vehicle deployment.

IV. RESEARCH METHODOLOGY

A. Research Design

The research was conducted using a quantitative ml approach, which included a feasibility study of a qualitative approach. There are three sets of OBD-II data for the purpose of classification of binary states of the driver. Also, the outcome has been assessed in terms of availability of sensors, cost, computational requirements, explainability, safety, and feasibility of implementation. This research is not concerned with implementing a Formula 1 telemetry system.

B. Data Sources and Preprocessing

The data used in this research is that of Driver 1, Driver 2 and Driver 3. There are 33 telemetry features for each dataset, and the target label and driver identifier. Throttle, Position of the Pedal (Accelerator), Engine Load, Intake Pressure and Speed. Data is checked to determine if any data are null, duplicate, constant variables and classes imbalances. There were no constant features or copies of data removed; there was no driver id for learning purposes.

All the other numerical attributes were split into a 80/20 train test split. The features that were not constant were eliminated, and duplicate records were eliminated, and the drivers were not used for driver-based learning because they were not included.

C. Models and Evaluation

The models used in the test were Logistic Regression, Random Forest, Extra Trees, XGBoost, LightGBM. Accuracy, macro-precision, macro-recall, Macro-F1, and confusion matrix was used to evaluate performance. Class imbalance was a factor that led to giving more attention to macro f1.

D. Proxy Scope of Degradation

This research is not intended to measure real-life tire and brake wear as there is no data available to validate these parameters such as mileage, temperature, etc. for degradation modeling to be performed in the future with the clear objective and physical validation.

E. Three-Tier Framework

Level 1: Basic Intelligence – detects faults, alerts maintenance and provides a driving summary.

Level 2: Intelligent Vehicle Systems, applies ml to classify driver's state, detect anomalies and give maintenance recommendations based on stress levels.

Level 3: At the advanced intelligence level, high-frequency sensing, sensor-fusion, edge computing and vehicle-specific predictive models are added.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

F. Validation and Feasibility

The feasibility is evaluated based on the availability and quality of information, computational requirements, costs, explanation demands, safety gains and validation challenges. Current results are as a benchmark only. Independent test trips, drivers and vehicles are required to validate in the real world as the near-by telemetry data may be highly correlated.

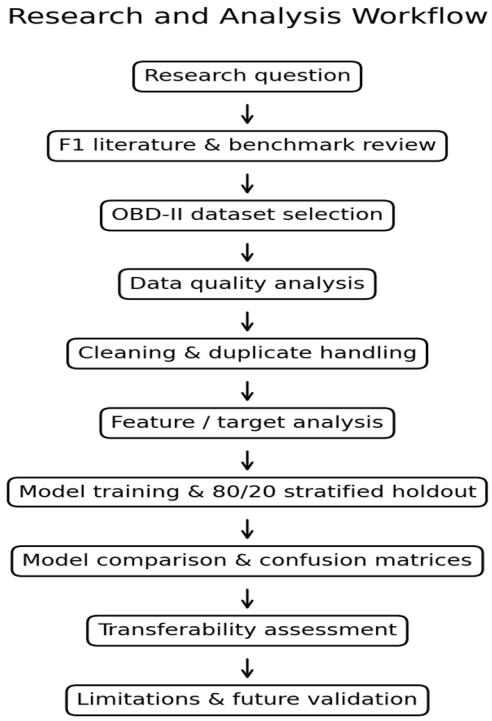


Fig. 1. Research and analysis workflow used to evaluate transferability from Formula 1 telemetry principles to passenger vehicles.

V. F1-INSPIRED TECHNOLOGY TRANSFER TO PASSENGER VEHICLES

A. Transfer Objective and Design Principle

The research does not attempt to create an actual Formula 1 telemetry system. Rather it will change its core analytical approach of car models to collecting data, forming clean signals, creating features, and then generating decisions using ML models. This is probably a more inexpensive solution that can take advantage of OBD-II/CAN data and some other sensors instead of the large and costly network of proprietary sensors that Formula 1 teams require. Available OBD-II/CAN data and selected additional sensors can be used in a lower-cost system as opposed to the extensive and proprietary sensor infrastructure that is utilized in Formula 1.

B. F1 Capability-to-Passenger-Vehicle Mapping

F1 capability	F1-style function	Passenger-vehicle implementation	OBD-II / added signal	ML / analytics method
High-density telemetry	Continuous multi-sensor vehicle-state capture	Reduced telemetry stream	RPM, speed, load, throttle, temperatures, pressure	Data-quality + time-window pipeline
Telemetry cleaning	Remove/correct anomalies before modeling	Outlier checks, duplicate removal, plausibility rules	All available OBD-II fields	Preprocessing / validation
Vehicle-state modeling	Infer car performance from telemetry	Operating-state classification	Throttle, accelerator, load, manifold pressure	Extra Trees / XGBoost / LightGBM

Driver behavior analysis	Assess driver inputs and vehicle response	Binary/multiclass driver-state classification	Throttle, speed, acceleration, steering where available	Classification models
Braking performance	Analyze braking intensity and corner phase	Braking-demand proxy	Speed, deceleration, brake switch/pressure if available	Window features + classifier/regressor
Tire performance	Estimate degradation/performance loss	Tire-stress / degradation proxy	Speed, load, temperature, wheel-speed/TPMS if available	Regression / time-series model
Power-unit monitoring	Track system state and anomalies	Engine anomaly / health score	RPM, load, coolant, intake, fuel trims, voltage	Anomaly detection / classification
Historical + live comparison	Compare current state with historical patterns	Baseline profile and trend detection	Current telemetry + stored trips	Gradient boosting / anomaly scoring
Real-time decision support	Convert model output to action	Maintenance / eco-driving / safety alert	Edge device or smartphone	Low-latency inference + thresholds

C. Proposed Passenger-Vehicle Telemetry Architecture

It is a six stage process.

Data collection → Data preprocessing → Feature extraction → Machine learning prediction → Decision support Historical data storage.

All obd2 / CAN signals are collected and checked. After a short time meaningful measures such as throttle variation, speed changes, acceleration strength, load levels are calculated. The features are then fed into the ml models and historical trip data is used as a reference to compare with.

D. Machine-Learning Layer for the Transfer

The experiment is completed and confirmed both binary driver-state classification and performance of the Logistic Regression, Random Forest, Extra Trees, XGBoost and LightGBM classifiers. Overall this tree-based model performed better.

Development should be future-oriented, and based on time-based telemetry analysis. Instead of analyzing each individual OBD-II data point it could analyze 5-10 seconds of data for changes in speed, throttle, acceleration, load and temperature. These features can be used to help train gradient boosting models, or temporal models like LSTM or Temporal CNN for larger datasets.

E. Proposed ML Tasks for F1-to-Passenger Transfer

Transfer task	Target	Useful passenger-vehicle signals	Candidate model	Deployment output
Driver-state classification	Class 0 / Class 1 or validated behavior classes	Throttle, pedal, speed, load, acceleration	Extra Trees / XGBoost / LightGBM	Behavior score
Operating-state anomaly detection	Normal vs abnormal window	RPM, load, throttle, temperature, voltage	Isolation Forest / One-Class model	Anomaly score
Braking-demand proxy	Low/medium/high stress class	Speed, deceleration, brake	XGBoost / Random Forest	Braking-stress index

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

		signal, load		
Component-stress proxy	Stress or deterioration category	Load, speed, temperature, repeated events	Gradient boosting regression/classification	Stress trend
Predictive maintenance	Future fault/service event	Telemetry + DTC + service history	XGBoost / LightGBM	Failure/service probability
Driver-aware eco/safety feedback	Risk or efficiency score	Throttle, speed variation, acceleration	Gradient boosting	Real-time advisory

F. Implementation Prototype and Reproducible Code

Github Repository Link : https://github.com/prats010/codes

G. From Static Classification to F1-Style Telemetry Intelligence

It is a binary classification model as a preliminary step of an intelligent system for passenger vehicles based on f1 (fuzzy logic). A better architecture would combine live state classification, anomaly detection and predictive modelling in a single pipeline. A sensor could process OBD-II data to generate metrics across trends, and recognise irregularities in the way the cars are being driven and alert to potential repair issues. Previous trip data may prove helpful in establishing baseline data and changes over time.

The fundamental principle is that the data from a telemetry system is not intelligence. Value is derived from ongoing information analysis, comparison with trends from the past and the decision-making based on forecasting. This is an example of an F1-inspired approach that has a smaller sensor set.

H. Experimental Roadmap for the Research

Stage	Experiment	Deliverable	Status
1	OBD-II data-quality + class analysis	Clean dataset, class distribution, feature correlations	Completed
2	Binary driver-state classification	Accuracy, Precision, Recall, Macro-F1, confusion matrices	Completed
3	Time-window feature engineering	Temporal behavior features	Next
4	F1-inspired stress proxies	Braking/load/thermal operating-stress indices	Next
5	Predictive maintenance model	Future service/fault probability	Future
6	Driver-wise / trip-wise validation	Generalization estimate	Required
7	Edge deployment	Latency, memory, power, offline inference	Future
8	Cross-vehicle validation	Transferability across vehicles	Future

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue XI Nov 2026- Available at www.ijraset.com

VI. RESULTS AND DATA ANALYSIS

A. Dataset Composition and Data Quality Analysis

In the experiment, three OBD-II driving dataset comprising 555,000 observations were used. Driver 1 had 174,600 records, Driver 2 had 183,600, and Driver 3 had 196,800. For each data set, there were 33 variables measured at each time point, a target variable, and an identifier of the driver.No missing value. Two constant variables were removed as they did not provide any useful information, and the driver identifier was removed to prevent the models from learning driver identity.

Prior to analysis, duplicate entries have been removed. Driver 1 had 18,278 duplicates, Driver 2 had 23,414, and Driver 3 had 25,975, resulting in 67,667 duplicates overall. Their removal decreased the risk to overly optimistic models' performance.

Dataset	Raw observations	Duplicates	Class 0	Class 1
Driver 1	174,600	18,278	42,272	132,328
Driver 2	183,600	23,414	35,065	148,535
Driver 3	196,800	25,975	41,061	155,739
Combined	555,000	67,667	118,398	436,602

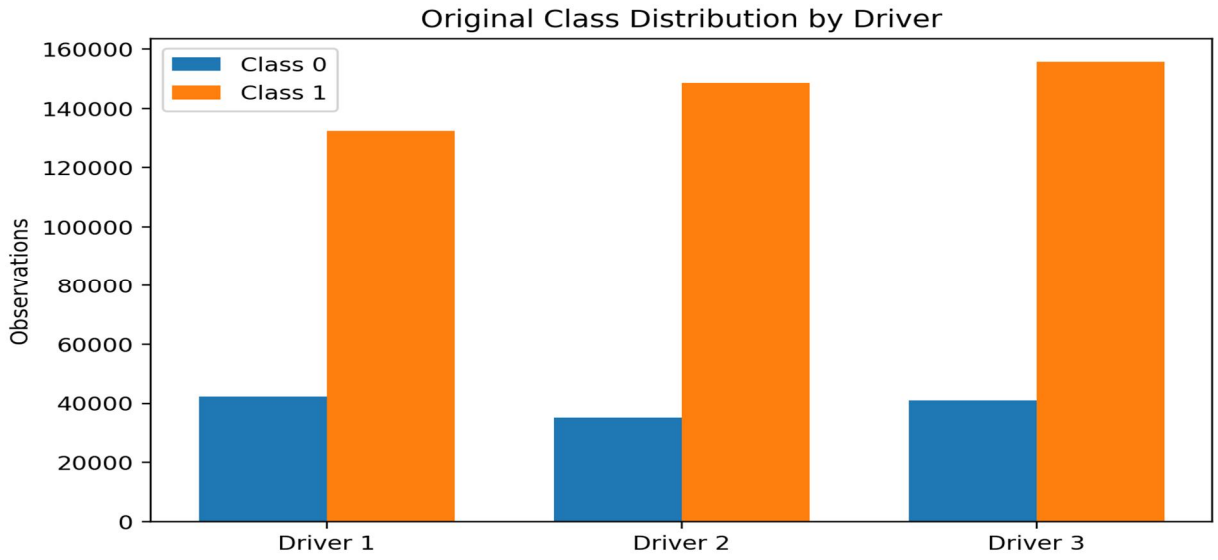


Fig. 2. The initial distribution of the target class, before duplicates are removed.

B. Class Distribution and Imbalance Analysis

The combined dataset was imbalanced, with 118,398 Class 0 observations (21.33%) and 436,602 Class 1 observations (78.67%). This was seen across all three drivers with Driver 1 having 24.21% Class 0, Driver 2 having 19.10% and Driver 3 having 20.86%.

So, precision was not sufficient to assess models. To make both classes equally important, Macro-precision, Macro-recall and Macro-F1 were also implemented. Labels are kept as Class 0 and Class 1 since there is no information about their behavior from this dataset.

C. Exploratory Feature Analysis

The highest correlations reported with the target were for throttle, accelerator pedals, engine loads and intakes, as determined by correlation analysis. It is in line with all three drivers' datasets.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Feature	Driver 1	Driver 2	Driver 3
Absolute Throttle Position B	-0.863	-0.854	-0.861
Commanded Throttle Actuator	-0.856	-0.847	-0.853
Relative Throttle Position	-0.843	-0.852	-0.847
Accelerator Pedal Position D	-0.803	—	-0.789
Accelerator Pedal Position E	-0.802	—	-0.789
Absolute Load	—	—	-0.775
Absolute Throttle Position	—	-0.806	—

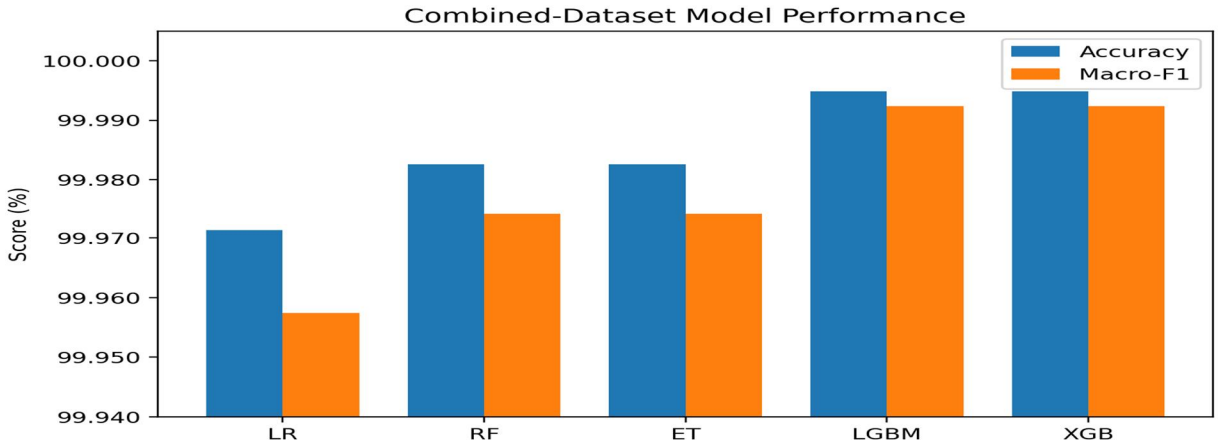


Fig. 3. Highest reported feature–target correlations. The values reported for the correlations are absolute and only those that did not miss are available in the analysis given.

The best correlations were observed between all three drivers, and primarily with the throttle position, throttle control, accelerator-pedal position, and the vehicle load. There were strong negative correlations between Absolute Throttle Position B, Commanded Throttle Actuator, and Relative Throttle Position in all data sets, with the correlations in general ranging from approximately -0.84 to -0.86. Driver 3 had a very good relationship with Absolute Load (-0.775).

The trend suggests that this is a measure of how a vehicle is being used generally, and not linked to one driver's 'style' and so could be usable across the board. However, correlation does not mean causation and it is not a measurement of the importance of features directly for ml models. More sophisticated analysis using methods such as SHAP value and permutation importance need to be conducted.

D. Driver-Specific Model Comparison

Dataset	Model	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro-F1 (%)
Driver 1	Logistic Regression	99.9084	99.8791	99.8711	99.8751
Driver 1	Random Forest	99.9685	99.9752	99.9390	99.9571
Driver 1	Extra Trees	99.9742	99.9790	99.9508	99.9649
Driver 1	LightGBM	99.9714	99.9771	99.9449	99.9610
Driver 1	XGBoost	99.9742	99.9790	99.9508	99.9649
Driver 2	Logistic	99.9537	99.9224	99.9278	99.9251

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

	Regression				
Driver 2	Random Forest	99.9782	99.9593	99.9702	99.9648
Driver 2	Extra Trees	99.9782	99.9702	99.9593	99.9647
Driver 2	LightGBM	99.9755	99.9522	99.9685	99.9603
Driver 2	XGBoost	99.9755	99.9468	99.9740	99.9604
Driver 3	Logistic Regression	99.9289	99.9057	99.8788	99.8923
Driver 3	Random Forest	99.9771	99.9721	99.9587	99.9654
Driver 3	Extra Trees	99.9873	99.9830	99.9785	99.9808
Driver 3	LightGBM	99.9822	99.9798	99.9663	99.9731
Driver 3	XGBoost	99.9822	99.9798	99.9663	99.9731

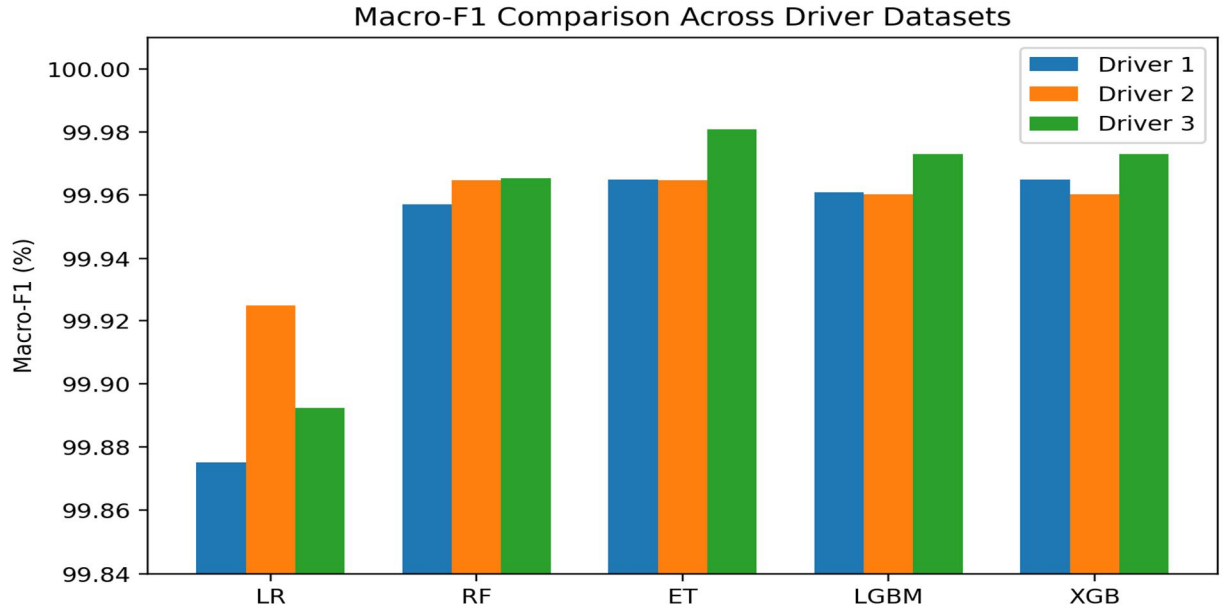


Fig. 4. Macro-F1 comparison between the three driver sets.

Ensemble models for trees performed considerably better than Logistic Regression baseline. Extra Trees outperformed all other models for Driver 3 and tied with XGBoost for Driver 1 in terms of the Macro-F1. Random Forest had the highest score and the margins between its score and the Extra Trees score for Driver 2 were small.

E. Combined-Dataset Model Analysis

Model	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro-F1 (%)
Logistic Regression	99.9713	99.9609	99.9540	99.9574
Random Forest	99.9826	99.9681	99.9802	99.9742
Extra Trees	99.9826	99.9698	99.9785	99.9742
LightGBM	99.9949	99.9950	99.9898	99.9924
XGBoost	99.9949	99.9950	99.9898	99.9924

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

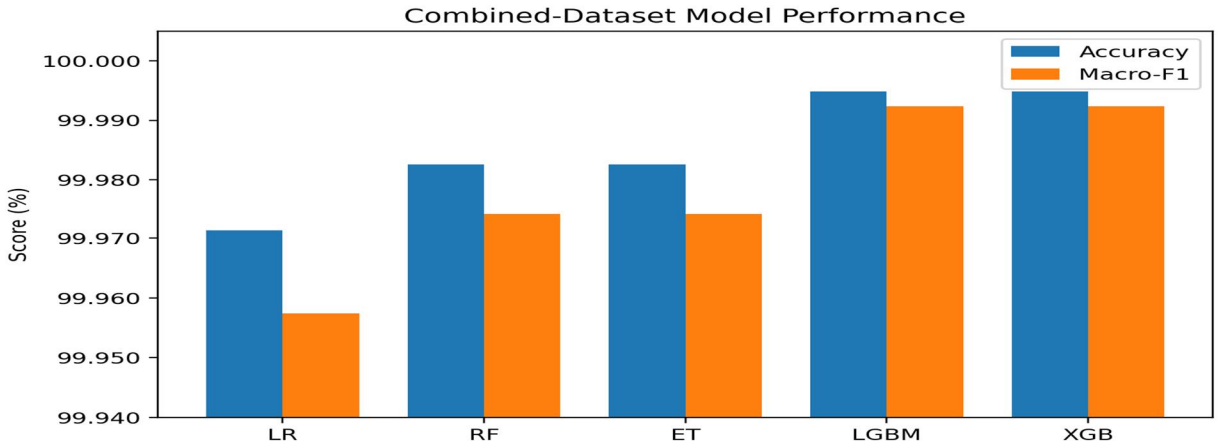


Fig. 5. Accuracy and Macro-F1 for the combined data set.

The combined-dataset performance of LightGBM and XGBoost was the best with 99.9949% accuracy and 99.9924% Macro-F1. Both gradient boosting models have similar performance so they are able to model a nonlinear relationship with obd-ii data effectively. Random Forest and Extra Trees also performed well with high Macro-F1 scores above 99.97%, and Logistic Regression with a slightly lower score performed very well.

F. Confusion Matrix Analysis

The confusion matrix will be used for there was not any indication that a majority class prediction was causing high scores in the confusion matrix. All the best models performed well with minimum error rate to classify both classes. Combined dataset had 20954 true negatives and 76508 true positives with just 4 false positives and 1 false negative, showing a very good accuracy.

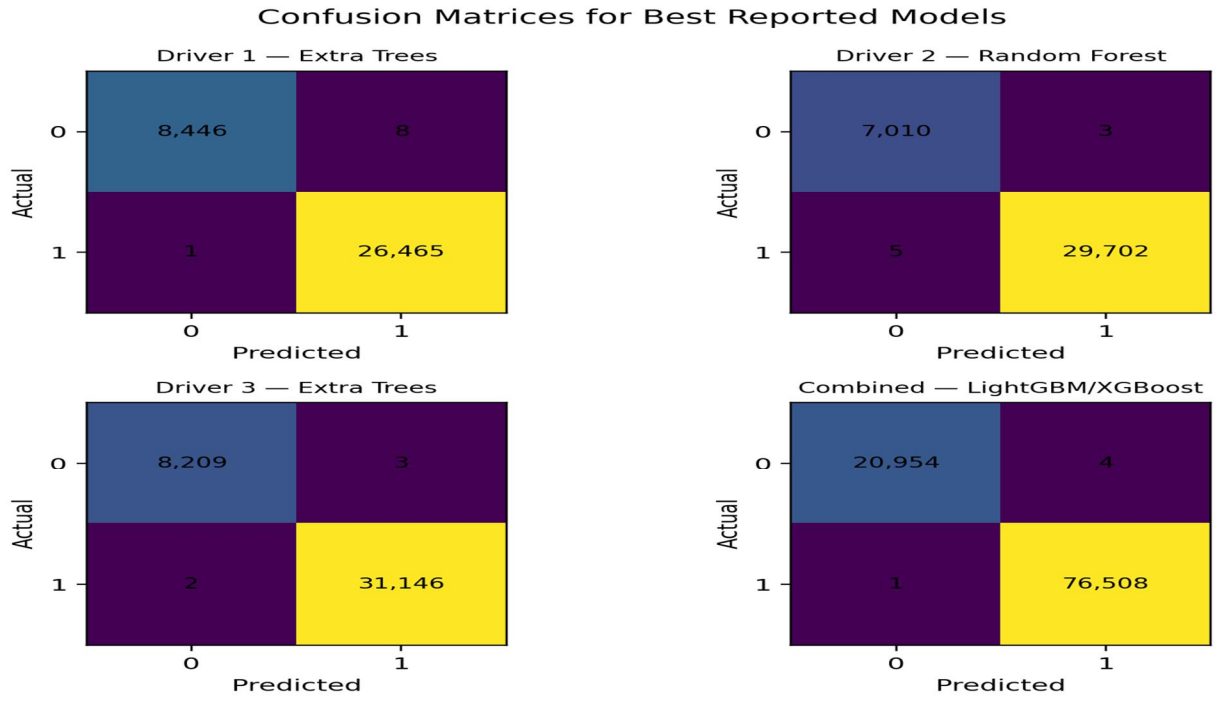


Fig. 6. Confusion matrices of the top reported models with the tested holdout method.

High Macro-F1 values justify it as low number of misclassification errors i.e. false positive and false negative. Moreover, the matrix demonstrates that these models did not have the reported accuracy predicting majority classes. However, these extremely high scores should be taken as a validation of this fact.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

G. Model Selection and Practical Interpretation

Individual drivers performed best with Extra Trees with XGBoost being marginally better for Driver 2 and Extra Trees for Driver 1. The best for a combined dataset was LightGBM and XGBoost. However, the selection of a model should also be based on speed, memory usage, explainability and performance on unseen drivers and vehicles.

H. Validation and Generalization Considerations

They are sampled in a stratified way from a random train-test split without duplicates. If the telemetry data for the vehicles is sequential, then it was possible to get similar results on a training set and a test set.

Therefore, the 99.9%+ results should be considered as benchmarks and not proofs of real-world generalization. Future studies may be based on trips, should be time based, leave one driver out, and be cross vehicle validated.

I. Using Formula 1 Based Benchmarks

The literature-based studies are not F1 studies, but those that focus on other topics such as tire wear and pit-stop prediction. These cannot be directly compared to the F1 study in this OBD-II binary classification task. This high accuracy is just an indicator that one is able to derive useful predictive patterns from the telemetry data of a vehicle, but it does not imply that this is better than Formula 1 analytics.

J. Transferability and Three-Tier Vehicle Intelligence Analysis

The F1 studies in the literature can be used for a range of problems including tyre degradation and pit-stop prediction, but generally require several OBD-II signals and machine learning. Level 1 involves basic diagnostics, and Level 3 includes advanced tire and brake prediction, which needs more sensors and validated and verified physical data.The one thing that can be moved from F1 to another is analytical not expensive sensors. In this session, attendees will learn how to utilize transferability and Three-Tier Vehicle Intelligence Analysis.

Considerable information is available in the three OBD-II datasets to enable high accuracy binary classification over the random holdout protocol evaluated. The basic diagnostics are still considered Level 1; advanced tire and brake diagnostics predictions are considered Level 3, and they need more sensors and validated physical data. The nonlinear tree-based models outperform the Logistic Regression baseline, and it shows that there are important interactions among the features in vehicle telemetry data.

K. Overall Findings

1) The three OBD-II datasets are rich enough to facilitate highly accurate binary classification with the protocol tested, random holdout.
2) The most frequent target associations reported are related to throttle position, accelerator-pedal position, engine load, and operating-state variables.
3) The 3rd results indicate that the use of feature interactions in structured vehicle telemetry is crucial and the nonlinear tree-based models outperform the Logistic Regression baseline.
4) When it comes to overall individual-driver performance, Extra Trees shows the most consistent performance, whereas LightGBM and XGBoost show best performance in the combined-dataset approach.
5) It is concluded that the F1 derived analytical methods can be adopted selectively to passenger vehicles while direct prediction of the degradation of the tires and brakes is still an issue for the future to be addressed by additional sensors and validated ground truth.

VII. DISCUSSION

A. Interpretation of Findings

The results are supportive for the transfer of the Formula 1 analytical techniques to a passenger vehicle. Data from OBD-II can be employed in telemetry analysis, feature feature engineering, machine learning prediction and decision support. But these results aren't representative of their performance on any of the other types of vehicles or roads in different conditions.

B. Degradation Modeling as a Separate Transfer Challenge

Most cars don't have direct measurements for wear, tyre and brake wear is more difficult to transfer. An adequate degradation proxy also needs to be well defined and tested through the examination of the physical parts and service history reports before it could be

used as an actual wear of a part. The transfer of tire and brake degradation is more challenging as the majority of consumer vehicles do not have direct measurements for the degradation.

C. Technology Transfer Implications

1) Don't give away the whole F1 and all the inclusive f1 infrastructure.
2) Be prepared to give standardized signals and describe any limitations on the vehicle.
3) Use available direct measures where they are not available, use explicitly defined and validated proxies.
4) Choose models for computational, latency, latency, explainability and validation constraints.
5) Provide confidence/uncertaintypos information on user facing outputs.
6) Verify models using other drivers of interest, such as the numbers of trips and vehicles, before putting them into production, due to safety concerns. Avoid using too many different types of signal and record any limitations for a particular vehicle.

D. Cost, Safety, and Explainability

A passenger-vehicle AI system should make sure to correctly predict events while remaining cost effective, efficient and trusted by users. More complex models might need more computing resources, storage, maintenance and security. For safety related use, it is important to include clear reasons (and not just a risk score).

The consequences of false positives are different than false negatives. If there are many false alerts, users might disrespect the warning and if the warning is not received, it could be a safety hazard. The thresholds, calibration, costs of errors and performance of the models should thus be assessed in future studies under realistic driving conditions.

VIII. CONCLUSION

The study examined the possibility of applying some of those techniques used by Formula 1 data analysis to an OBD-II equipped car. Three datasets were used for the evaluation of binary classification by Logistic Regression and some tree based models.

The results indicated excellent predictive performance, where the highest accuracy of individual-driver was 99.9873% and the best for the combined dataset was 99.9949% accuracy and 99.9924% Macro-F1.

The variables relating to the throttle position, accelerator pedal position, and engine load were found to be more strongly related to the target.

The results indicate that it is possible analytically, but not for production. It is possible that some of the sequences are found in both of the data sets (training and test); so more testing is required. We will need to predict tire and brake wear in the future, because we don't have any validation of actual wear information.

All in all, this research provides guidelines on how such a data-driven approach to f1-sensing, processing, prediction, and decision support could be implemented for passenger vehicles, which would have simpler and cheaper systems

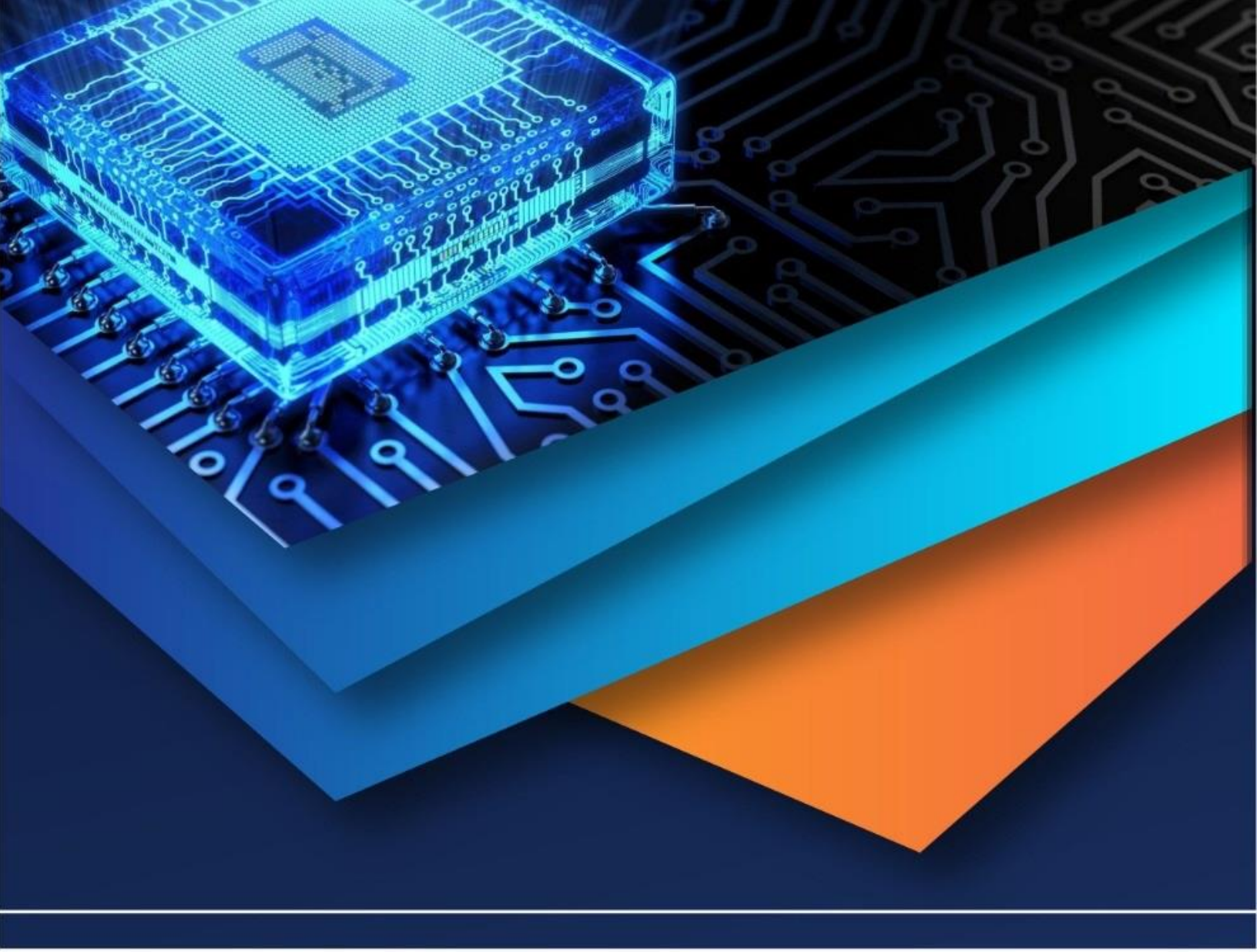
VIII. FUTURE RESEARCH DIRECTIONS

1) Cross test of cars on various brands, engine types, car models.
2) The generalization without temporal leakage is measured by 3 validation methods: 2. Trip based, block time based, and leave one driver out.
3) Direct component ground truth through tire pressure readings, tire temperature readings, brake inspection, pad thickness and service history.
4) Domain adaptation for transferring models between vehicles/datasets where there is very limited new data.
5) Explainable Vehicle AI: SHAP, Permutation importance, Counterfactual reasoning, Calibrated confidence estimations.
6) Use the metrics for deployment to measure latency, memory usage, energy usage and reliability of inference on cheap, embedded hardware.
7) Learning in the presence of privacy, such as federated learning with driver data stored on vehicle or trusted fleet setting.
8) Extension of electric vehicle 8: Battery state, re-generation braking, motor temperature, charging behavior.
9) Correlating proxy variables from telemetry with component service events (longitudinal maintenance investigations). Drivers' reaction to machine learning based recommendations.
10) Human factors testing to understand whether drivers interpret machine learning based recommendations correctly.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] E. T. Michailidis, A. Panagiotopoulou, and A. Papadakis, "A review of OBD-II-based machine learning applications for sustainable, efficient, secure, and safe vehicle driving," Sensors, vol. 25, no. 13, Art. no. 4057, Jun. 2025, doi: 10.3390/s25134057.
[2] E. Romera, L. M. Bergasa, and R. Arroyo, "Need data for driver behaviour analysis? Presenting the public UAH-DriveSet," in Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC), 2016, pp. 387–392, doi: 10.1109/ITSC.2016.7795584.
[3] A. Mohite, S. S. Patil, and A. S. Mali, "Driving style classification using deep learning techniques," in Intelligent Sustainable Systems, J. S. Raj, I. Perikos, and V. E. Balas, Eds. Singapore: Springer, 2023, vol. 665, pp. [verify page range], doi: 10.1007/978-981-99-1726-6_50.
[4] A. Mohite, S. S. Patil, and A. S. Mali, "Enhancing machine learning algorithm performance through feature selection for driver behavior classification," International Journal of Electrical and Computer Engineering Systems, vol. 35, no. 1, pp. 354–365, 2024, doi: 10.11591/ijeecs.v35.i1.pp354-365.
[5] A. A. Moukafih, H. Hafidi, and M. Ghogho, "Aggressive driving detection using deep learning-based time series classification," in Proc. IEEE Int. Conf. Innovations in Intelligent Systems and Applications (INISTA), 2019.
[6] Formula 1, "Rob Smedley on Car Performance Scores and the data behind the latest F1 Insights TV graphics," Formula1.com. Available:
[7] https://www.formula1.com/en/latest/article/rob-smedley-on-car-performance-scores-and-the-data-behind-the-latest-f1.5QoyFi4KL63EULKM4u5mGL. Accessed: Aug. 16, 2026.
[8] Amazon Web Services, "F1 Insights powered by AWS," AWS for Sports. Available: https://aws.amazon.com/sports/f1/. Accessed: Aug. 16, 2026.
[9] Formula One World Championship Limited, "Formula 1 selects AWS as its Official Cloud and Machine Learning Provider," Jun. 29, 2018. Available: https://corp.formula1.com/formula-1-selects-aws-as-its-official-cloud-and-machine-learning-provider/. Accessed: Aug. 16, 2026.
[10] P. Bhamare, S. Shevkari, and P. Mane, "How machine learning with Python and artificial intelligence benefits Formula 1 teams," International Journal of Innovative Research in Technology, vol. 12, no. 6, 2025.
[11] A. Sasikumar, A. A. Leema, and P. Balakrishnan, "Data-driven pit stop decision support for Formula 1 using deep learning models," Frontiers in Artificial Intelligence, vol. 8, Art. no. 1673148, 2025, doi: 10.3389/frai.2025.1673148.
[12] D. Thomas, J. Jiang, A. Kori, A. Russo, S. Winkler, S. Sale, and A. Rago, "Explainable reinforcement learning for Formula One race strategy," in Proc. 40th ACM/SIGAPP Symp. Applied Computing, Mar. 2025, pp. 1090–1097.
[13] S. Sitara, W. Fatima, and J. Johrendt, "Deep-racing: An embedded deep neural network (EDNN) model to predict the winning strategy in Formula One racing," 2023.
[14] A. Heilmeier, A. Thomaser, M. Graf, and J. Betz, "Virtual strategy engineer: Using artificial neural networks for making race strategy decisions in circuit motorsport," Applied Sciences, vol. 10, no. 21, Art. no. 7805, 2020.
[15] D. Piccinotti, A. Likmeta, N. Brunello, and M. Restelli, "Online planning for F1 race strategy identification," in Proc. AAAI Conf. Artificial Intelligence, 2021, pp. 12624–12632.
[16] J. J. Molina-Campoverde, J. Zurita-Jara, and P. Molina-Campoverde, "Driving pattern analysis, gear shift classification, and fuel efficiency in light-duty vehicles: A machine learning approach using GPS and OBD II PID signals," Sensors, vol. 25, no. 13, Art. no. 4043, 2025.
[17] R. Kumar and A. Jain, "Driving behavior analysis and classification by vehicle OBD data using machine learning," The Journal of Supercomputing, vol. 79, no. 16, pp. 18800–18819, 2023.
[18] A. E. Campos-Ferreira, J. de J. Lozoya-Santos, J. C. Tudon-Martinez, R. A. R. Mendoza, A. Vargas-Martínez, R. Morales-Menendez, and D. Lozano, "Vehicle and driver monitoring system using on-board and remote sensors," Sensors, vol. 23, no. 2, Art. no. 814, 2023.
[19] R. Canal, F. K. Riffel, and G. Gracioli, "Machine learning for real-time fuel consumption prediction and driving profile classification based on ECU data," IEEE Access, vol. 12, pp. 68586–68600, 2024.
[20] M. Rykała, M. Grzelak, Ł. Rykała, D. Voicu, and R. M. Stoica, "Modeling vehicle fuel consumption using a low-cost OBD-II interface," Energies, vol. 16, no. 21, Art. no. 7266, 2023.
[21] M. Malik and R. Nandal, "A framework on driving behavior and pattern using On-Board diagnostics (OBD-II) tool," Materials Today: Proceedings, vol. 80, pp. 3762–3768, 2023.
[22] S. K. Singh, V. Maan, and A. K. Singh, "OBD II-based performance analysis based on fuel usage of vehicles," in Proc. Int. Conf. Emerging Trends in Expert Applications & Security, Singapore: Springer, 2023, pp. 159–168.
[23] H. Abediasl, A. Ansari, V. Hosseini, C. R. Koch, and M. Shahbakhti, "Real-time vehicular fuel consumption estimation using machine learning and on-board diagnostics data," Proc. Inst. Mech. Eng., Part D: J. Automobile Engineering, vol. 238, no. 12, pp. 3779–3793, 2024.
[24] M. A. Uddin, N. Hossain, A. Ahamed, M. M. Islam, A. Khraisat, A. Alazab, and M. A. Talukder, "Abnormal driving behavior detection: A machine and deep learning based hybrid model," International Journal of Intelligent Transportation Systems Research, vol. 23, no. 1, pp. 568–591, 2025.
[25] V. Shirole, A. K. Shahade, and P. V. Deshmukh, "A comprehensive review on data-driven driver behaviour scoring in vehicles: technologies, challenges and future directions," Discover Artificial Intelligence, vol. 5, no. 1, Art. no. 26, 2025.
[26] S. Bouhsissin, N. Sael, and F. Benabbou, "Driver behavior classification: A systematic literature review," IEEE Access, vol. 11, pp. 14128–14153, 2023.
[27] S. Roussou, E. Michelaraki, C. Katrakazas, A. P. Afghari, C. Al Haddad, M. R. Alam, and G. Yannis, "Unfolding the dynamics of driving behavior: a machine learning analysis from Germany and Belgium," European Transport Research Review, vol. 16, no. 1, Art. no. 40, 2024.
[28] G. Al-refai, M. Al-refai, and A. Alzu'bi, "Driving style and traffic prediction with artificial neural networks using on-board diagnostics and smartphone sensors," Applied Sciences, vol. 14, no. 12, Art. no. 5008, 2024.
[29] Y. Mahale, S. Kolhar, and A. S. More, "Enhancing predictive maintenance in automotive industry: addressing class imbalance using advanced machine learning techniques," Discover Applied Sciences, vol. 7, no. 4, Art. no. 340, 2025.
[30] Y. Mahale, S. Kolhar, and A. S. More, "A comprehensive review on artificial intelligence driven predictive maintenance in vehicles: technologies, challenges and future research directions," Discover Applied Sciences, vol. 7, no. 4, Art. no. 243, 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)