Revving up insights: machine learning-based classification of OBD II data and driving behavior analysis using g-force metrics

Siddhanta Kumar Singh[1], Anand Sharma[2]
[1]Department of Computer and Communication Engineering, School of Computer Science and Engineering, Manipal University Jaipur, Jaipur, India
[2]Department of Computer Science and Engineering, School of Engineering and Technology, Mody University of Science and Technology, Lakshmangarh, India

Article Info

Article history:

Received Oct 3, 2024
Revised Nov 30, 2024
Accepted Dec 25, 2024

Keywords:

Controller area network
G-force
On-board diagnostics II
Sideslipping
Swerving
Weaving

ABSTRACT

This research work uses machine learning (ML) approaches to classify on-board diagnostics II (OBD II) data and g-force measures to provide a thorough analysis of driving behavior. The research paper effectively demonstrates the classification of driving behaviours using OBD II and g-force data. Driving behaviours are analyzed by using ML algorithms such as random forest (RF), AdaBoost, and K-nearest neighbors (KNN). The analysis goes beyond a summary by discussing how OBD II data, g-force metrics, and the algorithms interrelate to classify ten distinct driving behaviors (e.g., weaving, swerving, and sideslipping). The RF classifier achieved the highest accuracy, which reinforces the strength of the chosen models. The inclusion of comparisons with other techniques supports arguments about the model's performance. The related works section connects the references to the central topic by highlighting prior approaches and research studies related to OBD II and driver behaviour analysis. The goals of this study are improving the accuracy of driving behaviour classification, with implications for traffic safety, driver education, and insurance sectors.

Corresponding Author:

Siddhanta Kumar Singh
Department of Computer and Communication Engineering, School of Computer Science and Engineering
Manipal University Jaipur
Jaipur, India
Email: siddhanta.singh@jaipur.manipal.edu

1. INTRODUCTION

The number of vehicles grows annually due to the rapid rising economy and government liberalization policy for foreign automakers. The number of non-professional drivers is also rising quickly at the same time. The primary causes of traffic accidents are now the individual drivers, as most inexperienced drivers lack driving experience, are ignorant of the state of the vehicles, and have low knowledge of traffic safety. Therefore, it is utmost important to determine the driving behavior so that the local authority can retrieve the vehicle information and analyze and then take proper action.

On-board diagnostics (OBD) is a standard protocol [1] for vehicles that monitors various aspects of vehicle performance and health. It emerged in the mid-1990s as a significant advancement over the original OBD I system, which was introduced in the 1980s. OBD II's primary purpose is to check the engine's major components and alert the driver and concerned person if there is any malfunction, thus aiding in the maintenance and repair of vehicles. OBD II systems consist of a standardized digital communications port, known as the data link connector (DLC), which is typically located under the driver side dashboard. This port allows external devices, such as scan tools and diagnostic software, to interface with the vehicle's computer system. OBD-II has basically four communication protocols for interfacing the OBD-II port. But the preference of protocol is dependent on the vehicle manufacturer. The key work here lies in a comprehensive model development that categorizes behavior of driving with high accuracy using OBD II and G-force data. The research work introduces a method that contrasts with previous approaches by incorporating multiple weak classifiers into a robust classification system via AdaBoost. Additionally, the paper's application of random forest (RF) and its comparison with other algorithms like Naive Bayes and logistic regression demonstrate new findings in the classification of driving behaviours.

2. RELATED WORKS

Various diagnostic methods including Autel Maxidiag (elite series) and Launch X 431 were used [2] to enhance the problems and malfunction identification. Research by Ramai *et al.* [3] offers the foundation for an inexpensive method of online EV monitoring. Two Hyundai Ioniq EV had a Raspberry Pi ZeroW and additional parts fitted in order to connect to them via the OBD-II connector. Wen *et al.* [4] conducted the security analysis of wireless OBD-II scanners in this study. They designed and built DONGLESCOPE, an automated program that tests these dongles on a real car in real time, covering all possible assault phases. Options for monitoring important vehicle performance were described in this study by Yadav and Pathak [5], along with a synopsis of the sensors used to retrieve these parameter values.

Shaikh *et al.* [6] developed an Android application that monitors driving behavior and notifies the user of any discrepancy in driving habits in an effort to avert a catastrophic event. Vaiti *et al*. [7] research proposed a data driven method for cluster emission calculation based on vehicle parameters related to emissions. Ameen *et al.* [8] propose a way to categorize four driving behaviors: dangerous, aggressive, secure, and typical behavior, with the goal of reducing the chance of accidents. Three light-duty passenger cars (LDPVs) were tested by Zheng *et al.* [9] utilizing a laboratory dynamometer and the NEDC as a type-approval cycle.

The platform presented in this work Peppes *et al*. [10] combines open-source technology with machine and deep learning techniques to collect, store, process, analyze, and correlate data coming from cars. The results of a driving behavior literature review are discussed by Hermawan and Husni [11]. This study covers methods to collect OBD II data, and analyze, model, and assess it. Gharbins [12] assessed the degree of proficiency among technicians in utilizing the OBD II instrument for standard maintenance on automobiles containing electronic components, in addition to the degree of diagnostic equipment and reference materials available in nearby repair shops.

Big data analysis requires the use of several languages and technologies, including Hadoop, Python, Spark, R, and MATLAB, which are all covered by Meenakshi *et al.* [13]. The real-world statistics of a mild hybrid car might differ depending on a number of factors, such as the vehicle, engine cycle, and powertrain, as examined by Barbier *et al.* [14]. By examining the signals from the electronic control unit's PIDs, Campoverde *et al.* [15] created an algorithm that can identify two typical driving behaviors, such as braking to slow down and disengaging to shift gears. Subscription-based car maintenance options were recommended by Maalik and Ponnampalam for people who don't have the time for repairs and upkeep [16]. Hamed *et al.* [17] employ machine learning (ML) to improve the fuel consumption forecast accuracy model to decrease consumption of fuel. OBD data for vehicle dynamics analysis and forecasting is examined by Navali *et al*. [18]. According to the results, the OBD may provide data for a range of real-time and offline applications. Using a transformer neural network (TNN) ML technique, Fernández *et al*. [19] established a way for creating accurate speed correction data from OBD II data. Using OBD, Song and Kim [20] propose a method for determining CAN specifications linked to important vehicle metrics.

Research by Kim and Baek [21] present a method that automatically extracts private in-vehicle data by correlating sensor data with the sought information. A portable system to monitor mobile use while driving and, if required, take control of a driver's phone whenever the car attains a speed limit (>10 km/h) was proposed by Khandakar *et al.* [22]. A mathematical, graphic, and analytical approach for examining customer driving behavior is provided by Navneeth *et al.* [23]. Through the OBD interface, Kumar and Jain [24] suggested method that gathers vital performance data of vehicle, such as RPM, speed, position of accelerator paddle, determined motor load, and other characteristics. The suggested approach categories driver behavior using ML algorithms including AdaBoost, support vector machine (SVM) and RF. This paper’s main goal is on DB analysis methods, and it is rendered in an elaborated manner [25].

3. METHOD

All the real time OBD data with the attributes like device time, absolute throttle position (ATP) (%), accelerator pedal position (APP) (%), air fuel ratio, average trip speed (whilst stopped or moving)(km/h), engine load (%), revolutions per minute, fuel flow (FF) rate (gal/min), intake manifold pressure (psi), fuel trim bank (FTB) 1 short term (%), kilometers per litre (long term average) (kpl), City driving (%), idle driving (%), run time since engine start (s), speed (OBD II) (km/h), trip average KPL (kpl) were collected.

The w_{ATP}, w_{APP}, w_{FFR}, w_{FTB1L}, w_{IMP}, w_{TPM} are the normalized weights ATP, APP, FF rate, FTB 1 long, IMP and TP (manifold) respectively. So, the sum of the weights is given by (1):

$$\text{Sum} = w_{\text{ATP}} + w_{\text{APP}} + w_{\text{FFR}} + w_{\text{FTB1L}} + w_{\text{IMP}} + w_{\text{TPM}} \quad (1)$$

where the normalized weights given by (2) to (7):

$$w'_{\text{ATP}} = \frac{w_{\text{ATP}}}{\text{Sum}} \quad (2)$$

$$w'_{\text{APP}} = \frac{w_{\text{APP}}}{\text{Sum}} \quad (3)$$

$$w'_{\text{FFR}} = \frac{w_{\text{FFR}}}{\text{Sum}} \quad (4)$$

$$w'_{\text{FTB1L}} = \frac{w_{\text{FTB1L}}}{\text{Sum}} \quad (5)$$

$$w'_{\text{IMP}} = \frac{w_{\text{IMP}}}{\text{Sum}} \quad (6)$$

$$w'_{\text{TPM}} = \frac{w_{\text{TPM}}}{\text{Sum}} \quad (7)$$

the final model for engine load can be written as (8):

$$Engine\ Load = w'_{ATP}.ATP + w'_{APP}.APP + w'_{FFR} * FFR + w'_{FTB1L} * FTB1L + w'_{IMP} * IMP + w'_{TPM} * TPM \quad (8)$$

This equation involves fuel usages which contrast with the existing (13). It has been found that fuel consumption negatively affects driving scores and is closely tied to driver conduct. In (9) and (10) calculate average fuel consumption, which is obtained by averaging all instantaneous fuel consumption measurements:

$$Fuel\ Consumption\left(\frac{lit}{km}\right) = \frac{\text{Fuel Flow}\left(\frac{\text{lit}}{\text{hr}}\right)}{\text{Speed}\left(\frac{\text{km}}{\text{hr}}\right)} \quad (9)$$

$$Fuel\ Flow = \frac{\text{MAF}}{\lambda * \text{AFR} * \rho} \quad (10)$$

where λ is the OBD II parameter which denotes air and fuel ratio and has standard value 1, AFR denotes the stoichiometric ratio having standard value 14.7, MAF denotes mass air flow rate in gm/sec from MAF sensor, ρ is the petrol density which is typically 770 gm/liter. It is measured using (11):

$$Fuel\ Consumption = \frac{\text{Speed}}{\text{Mass Air Flow}} \quad (11)$$

IDL_ENG indicates that the engine is not running. This suggests that fuel is being wasted, which will eventually lower the driving score as indicated (12).

$$\text{Score} = \begin{cases} -1, 800 \le \text{rpm} \le 1000 \text{ and Gear} = \text{N and Speed} = 0 \\ 1, \text{Otherwise} \end{cases} \quad (12)$$

Engine load parameter reading at idle is around 20%, while a reading of 100% indicates that the engine is under full load. In general, a load parameter reading of 70% to 80% during normal driving mode is considered optimal for both performance and fuel efficiency. An engine load >80% or <70% negatively affects driving behavior negatively. Engine load as given in terms of air flow in (13):

$$Engine\ Load = \frac{Current\ Air\ Flow}{Max\ Air\ Flow(rpm).\frac{\text{Barometric Pressure}}{29.92}.\sqrt{\frac{298}{\text{Tamb}+273}}} \quad (13)$$

High speed braking is a type a braking when a vehicle applies brake suddenly as in Figure 1, and acceleration on y-axis abruptly decreases and keeps even negative for some time while acceleration on x-axis remains flat. Normal acceleration and normal deceleration are in the range of 0.1 m/sec^2<normal acceleration <2.74 m/sec^2 and -0.1 m/sec^2<normal deceleration<-2.74 m/sec^2 respectively.

Brakes applied abruptly negatively affects driving-score if a<2.74 m/sec^2 and σ<2.05 and brake was applied at v>55 km/hour. We can define the driving score S as given by (14):

$$S = f(a, \sigma, v) \tag{14}$$

To represent the negative effect on driving scores when the conditions are met, we can introduce a penalty function P as:

$$P = \begin{cases} k, if\ a < 2.74\ and\ \sigma < 2.05\ and\ v > 55 \\ 0, otherwise \end{cases} \tag{15}$$

where k is a positive constant that represents the penalty value. Then, the driving score S can be modeled as shown in (16):

$$S = S_0 - P(a, \sigma, v) \tag{16}$$

where, S_0 is the initial driving score before the penalty.

If in the first 1.5 seconds, a force of 56.69 kg is experienced, then hard braking will be experienced. Sudden unintended acceleration (SUA) is described as an unplanned, sudden, high-power acceleration from a standing start, or a very slow starting speed combined with what appears to be a loss of braking efficiency as shown in Figure 2.

When a vehicle speeds up suddenly, acceleration (a_x) on x-axis remains flat while acceleration (a_y) on y-axis sharply goes up. Thus, the standard deviation (σ_{a_x}) and value range of acceleration on x-axis are small. And a_x(t) is x-axis accel as a time-function, a_y(t) is y-axis accel as a time-function, σ_{a_x} is denotes standard deviation of a_x and R_{a_x} is value range of a_x. Given that a_x remains flat, we can represent as (17) a constant (C):

$$a_x(t) = C \tag{17}$$

given that a_y sharply increases, we can model it as a step function or an exponential function. A step function (S) is a simple way to represent a sudden increase:

$$S = \begin{cases} 0, if\ t < t_0 \\ A, if\ t \geq t_0 \end{cases} \tag{18}$$

where A is the sharp increase in acceleration at time t_0. Alternatively, an exponential function as given below can represent a sharp increase more smoothly:

$$a_y(t) = A(1 - e^{-\lambda(t-t_0)}) \tag{19}$$

where λ is the rate at which a_y increases, and t_0 is the time when the vehicle starts to speed up suddenly.
Since a_x remains constant, therefore,

The standard deviation $\sigma_{a_x} \approx 0\ and$ The value range $R_{a_x} \approx 0$ because a_x does not vary. Hard acceleration is considered if acceleration >2.74 m/sec^2. Some of the visual impressions of SUA could be car with blurred background, leaving tire marks, front end lifted and emitting smoke. when we rev up, fuel is squandered when the engine is revved up again without accomplishing any productive activity. This is seen as negative in our proposed approach for calculating driving score.
The negative driving score S can be:

$$S = -k.f(R).\delta(G).\delta(V) \tag{20}$$

where R, V, and G are the rpm, speed and gear respectively and k is a constant factor that determines the severity of the penalty.
The indicator functions are given in (21) and (22).

$$\delta(G) = 1\ if\ G = 0\ otherwise\ \delta(G) = 0 \tag{21}$$

$$\delta(V) = 1\ if\ V = 0\ otherwise\ \delta(V) = 0 \tag{22}$$

Penalty function: $f(R)$=max (0, R−900). This function returns the amount by which RPM exceeds 900. If RPM is 900 or less, the function returns 0. Therefore, the final score is given by (23).

$$S = -k.\max(0, R - 900).\delta(G).\delta(V) \tag{23}$$

The weaving pattern, as depicted in Figure 3, exhibits a sharp fluctuation in acceleration along the x-axis, which persists for a certain amount of time. A negative score is awarded, if the SD (x-axis data) is large and the Range (x-axis data) is large and acceleration (y-axis data) is Smooth. Smoothness of acceleration (SmoothAy) can be evaluated as shown below using the mean absolute deviation (MAD) of the acceleration, a lower MAD indicates smoother acceleration.

$$\text{SmoothAy} = \frac{1}{n}\sum_{i=1}^{n}\left|A_{y,i} - \bar{A}_y\right| \tag{24}$$

where $A_{y,i}$ is the individual acceleration data point and $\bar{A}_y$ is the mean acceleration. For negative award function, the score should be high when SD_x and R_x are high and SmoothAy is low (indicating smooth acceleration) as shown in (25):

$$Score = w_1.\left(\frac{SD_x}{\max(SD_x)}\right) + w_2.\left(\frac{R_x}{\max(R_x)}\right) - w_3.\left(\frac{\overline{\text{SmoothAy}}}{\max(\text{SmoothAy})}\right) \tag{25}$$

where w_1, w_2, and w_3 are weights that determine the relative importance of each term, and the terms are normalized by their maximum values to ensure they are comparable. Therefore, we can rewrite the model as:

$$Score = w_1.\left(\frac{SD_x}{SD_{x,max}}\right) + w_2.\left(\frac{R_x}{R_{x,max}}\right) - w_3.\left(\frac{\overline{\text{SmoothAy}}}{\overline{\text{SmoothAy}}_{max}}\right) \tag{26}$$

where, SD_x is standard deviation of x-axis data; $SD_{x,max}$ is maximum standard deviation observed in the dataset; R_x is range of x-axis data; $R_{x,max}$ is maximum range observed in the dataset; $\overline{\text{SmoothAy}}$ is mean absolute deviation of acceleration along y-axis; $\overline{\text{SmoothAy}}_{\text{max}}$ is maximum mean absolute deviation observed in the dataset; and w_1, w_2, w_3 are weights to balance the components (can be tuned based on empirical data or specific requirements).

Swerving is an enormous peak in acceleration on the x-axis is observed when swerving takes place, as illustrated in Figure 4. The negative score S can be formulated as (27):

$$S = w_1.peak(a_x) + w_2.range(a_x) + w_{3.}\sigma(a_x) + w_4.|\mu(a_x)| - w_5.flatness(a_y) \tag{27}$$

where, w_1, w_2, w_3, w_4, and w_5 are weights that can be adjusted based on the importance of each factor, flatness(a_y) is inversely proportional to the variability of a_y. A potential measure could be the reciprocal of the standard deviation of a_y:

$$flatness(a_y) = \frac{1}{\sigma(a_y)+\varepsilon} \tag{28}$$

where ϵ is a small constant to avoid division by zero.

Sideslipping is shown in Figure 5, sideslipping causes a rapid decline in y-axis acceleration. The driving score S can be written as(29).

$$S = w_1.\max\left(\left|\frac{da_y}{dt}\right|\right) + w_2.\left(-min(a_y)\right) + w_3.(-\bar{a}_y) + w_4.\left(\max(a_y) - \min(a_y)\right) + w_5.|\bar{a}_x| \tag{29}$$

To balance the importance of each component, we might choose w_1, w_2, w_3, w_4, and w_5 as .1,.5,.5,.3 and .2 respectively. These weights can be adjusted based on empirical data or specific application needs. $\max\left(\left|\frac{da_y}{dt}\right|\right)$ denotes sharp fall in a_y can quantify the sharp fall by looking at the second derivative (jerk) or by defining a threshold for the rate of change.

$min(a_y)$denotes minimum value of a_y: min(a_y)<0
$\bar{a}_y$ denotes mean value of a_y: $\bar{a}_y$<0
$\max(a_y) - \min(a_y)$ denotes range of a_y and ensures this value is large.
$|\bar{a}_x|$denotes mean value of a_x not near zero, indicating significant sideways motion.

A fast U turn is the case when a driver makes a sudden U-turn shown Figure 6, to the right or left, x-axis acceleration increases rapidly to a very high value or decreases rapidly to a very low value, respectively. The driving score S can be modeled as a function of 5 tuples as given in (30):

$$S = f\left(\mu_{a_x}, \sigma_{a_x}, R_{a_x}, \mu_{a_y}, T\right) \tag{30}$$

where μ_{a_x} is the mean of a_x, σ_{a_x} is the standard deviation of a_x,R_{a_x} is the range of a_x, μ_{a_y} is the mean of a_y and T is the time duration of the maneuver. A form (31) for the driving score could be:

$$S = k_1.\left|\mu_{a_x}\right| + k_2.\sigma_{a_x} + k_3.R_{a_x} + k_4.\left(1 - \frac{|\mu_{a_y}|}{\max(|a_y|)}\right) + k_5.T \tag{31}$$

where k_1, k_2, k_3, k_4, and k_5 are weighting coefficients that determine the importance of each term. These coefficients can be adjusted based on empirical data or specific requirements for the driving score.

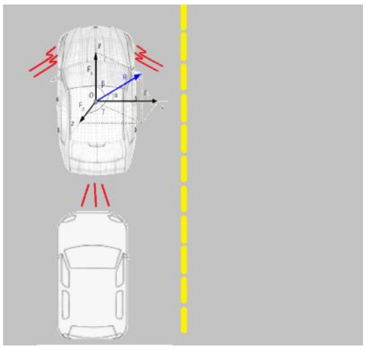

Figure 1. Sudden braking

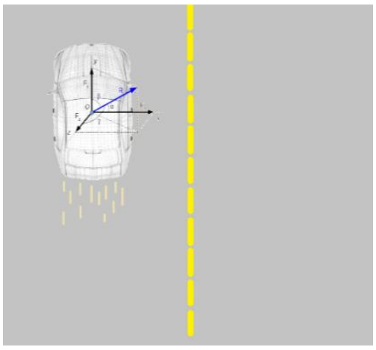

Figure 2. Sudden unintended acceleration

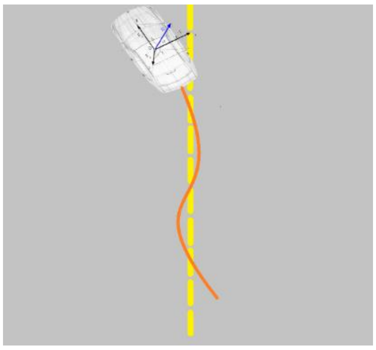

Figure 3. Weaving movement

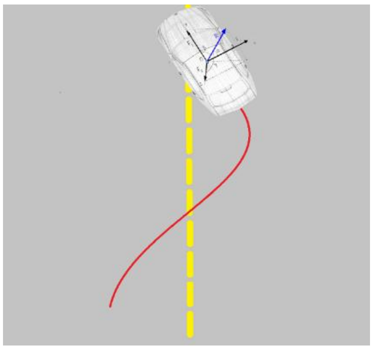

Figure 4. Swerving movement

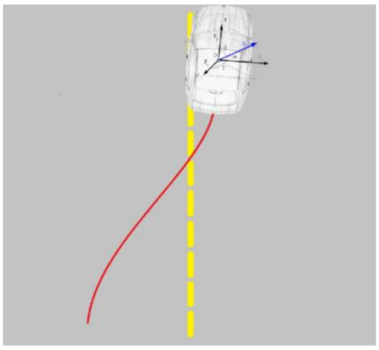

Figure 5. Sideslipping movement

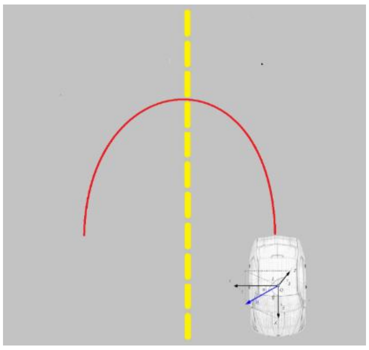

Figure 6. Fast U turn movement

Various mathematical models were developed, and the dataset was analyzed using classification algorithms in Python ML. The supervised algorithms applied include AdaBoost, RF, K-nearest neighbor (KNN), Naive Bayes, and logistic regression. The AdaBoost algorithm, an instance of adaptive boosting methodology, serves as a ML algorithm employed for classification, by amalgamating numerous weak classifiers to form a robust classifier while an ensemble learning technique called RF constructs several decision trees and aggregates their predictions to increase accuracy. KNN uses the majority label of their closest neighbours to categorise data points. Based on Bayes' Theorem, a probabilistic classifier assumes that characteristics are conditionally independent. For each training sample (x_i, y_i) is assigned a weight w_i. Initially, all weights are set equally.

$$w_i = \frac{1}{N}\ for\ all\ i \tag{32}$$

For each iteration t=1…up to T (T being the total number of iterations), a weak classifier ht(x) is trained using the weighted training-data and compute the classification error ϵ_t of ht as (33):

$$\epsilon_t = \sum_{i=1}^{N} w_i . \mathbf{1}(h_t(x_i) \neq y_i) \tag{33}$$

where, 1 is the indicator function.
Then calculate the weight αt for the weak classifier as (34):

$$\alpha_t = \frac{1}{2} ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{34}$$

$$w_i^{(t+1)} = w_i^t \exp\left(-\alpha_t y_i h_t(x_i)\right) \tag{35}$$

the weight is normalized using (36):

$$w_i^{(t+1)} = \frac{w_i^{(t+1)}}{\sum_{j=1}^{N} w_j^{(t+1)}} \tag{36}$$

the final strong classifier is (37).

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \tag{37}$$

The parameters α_t play a crucial role in determining the impact of each weak classifier on the ultimate decision, giving priority to classifiers that exhibit strong performance on the weighted training dataset. Drivers' ranks are awarded on the scale of 10 with 10 being excellent driving behavior. For every mistake committed, negative points will be rewarded in total score. Table 1 shows the score determinant.

Table 1. Driving score determinant

Driving parameters	Effect on driving score	Impact on driving score
Fuel_cons	-ve	HIGH
Idle_Eng	-ve	HIGH
Eng_Load	+ve	MODERATE
HIGH_SPEED_BRAKING	-ve	HIGH
SUA	-ve	HIGH
REV_ENGINE	-ve	HIGH
Weaving	-ve	HIGH
Swerving	-ve	HIGH
Sideslipping	-ve	HIGH
Fast U Turn	-ve	HIGH

4. RESULTS AND DISCUSSION

Feature classification offers a visual representation of derived parameters for various driver-classes from ten drivers from D1 to D10 on honda brio at different terrain for 10 kms. Figure 7 reveals that driver classes D1, D3, and D7 achieve the highest driving scores, each incurring a single negative penalty for idle engine, high-speed braking, and idle engine, respectively. Similarly, the remaining parameters provide a clear visualization of the data, facilitating the development of a model to classify drivers. The result is based on the those driving classes mentioned which is uniqueness of this work and the accuracy is also highest as mentioned below.

The driving behavior analysis [24] made and the accuracy was assessed using training and test dataset. The accuracy rates were as follows: AdaBoost at 77%, Naive Bayes at 88%, KNN at 98%, logistic regression at 99%, and RF at 100%. RF is substantially slower than all other classification techniques because it uses multiple decision trees for predictions and hence for speedy prediction we can assume logistic regression and KNN which also gives nice predictions.

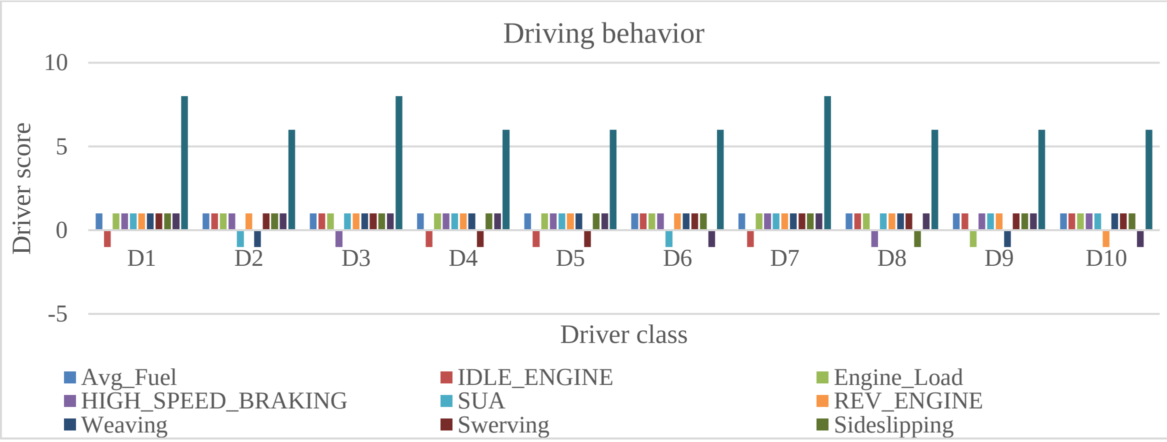

Figure 7. Driving score

5. CONCLUSION

ML techniques such as Ada Boost, RF, KNN, Naive Bayes, and logistic regression, were used to develop and validate a model for classifying driving behavior. The suggested techniques were simple to use and quite accurate, with random forest's maximum accuracy of 100%. This approach is still beneficial and can aid the traffic police, insurance company, local government, and claim processing. The findings of this research may also have a direct bearing on the development of driving assistance and classification systems. Although contemporary tools and algorithms are employed in this research project, there is still much room for the occasional implementation of further contemporary software tools and algorithms in accordance with future requirements. The method that is being given is not exclusive to cars with internal combustion engines; it may also be applied to contemporary hybrid and electric vehicles. The recommended course of action is doable and adaptable to new technology. Delays in data gathering and storage have an impact on the suggested method's results; as a result, classification of behavior is not possible for short road journeys, particularly when the route or vehicle are different.

ACKNOWLEDGEMENTS

We are thankful to the department and school for giving us opportunities and ample time and help whenever required.

FUNDING INFORMATION

No funding is involved for this paper.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Siddhanta Kumar Singh	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Anand Sharma			✓			✓		✓	✓	✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

All the authors do not have any conflict of interest.

INFORMED CONSENT

No personal or patient information is used for scientific reasons.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [SKSingh], upon reasonable request.

REFERENCES

[1] S. K. Singh and A. K. Singh, "Driving pattern analysis to determine driver behaviours for a local authority based on the cloud using OBD II," *International Journal of Electrical and Computer Engineering Systems*, vol. 13, no. 10, pp. 937–944, Dec. 2022, doi: 10.32985/ijeces.13.10.14.

[2] M. M. Oluwaseyi and A. M. Sunday, "Specifications and Analysis of Digitized Diagnostics of Automobiles: A Case Study of on Board Diagnostic (OBD II)," *International Journal of Engineering Research & Technology (IJERT)*, vol. V9, no. 01, Jan. 2020, doi: 10.17577/ijertv9is010045.

[3] C. Ramai, V. Ramnarine, S. Ramharack, S. Bahadoorsingh, and C. Sharma, "Framework for Building Low-Cost OBD-II Data-Logging Systems for Battery Electric Vehicles," *Vehicles*, vol. 4, no. 4, pp. 1209–1222, Oct. 2022, doi: 10.3390/vehicles4040064.

[4] H. Wen, Q. A. Chen, and Z. Lin, "Plug-N-Pwned: Comprehensive vulnerability analysis of OBD-II dongles as a new over-the-air attack surface in automotive IoT," *Proceedings of the 29th USENIX Security Symposium*, pp. 949–965, 2020.

[5] P. Yadav and P. K. Pathak, "On Board Diagnostics (OBD): A DTC (Diagnostic Trouble Code) Study," in *International Conference on Intelligent Technologies & Science - 2021 (ICITS-2021)*, 2021, pp. 1–4.

[6] M. K. Shaikh, S. Palaniappan, F. Ali, and M. Khurram, "Identifying Driver Behaviour Through Obd-Ii Using Android Application," *Palarch's Journal of Archaeology of Egypt/Egyptology*, vol. 17, no. 7, pp. 13636–13647, 2020.

[7] T. Vaiti, L. Tišljaric, T. Erdelic, and T. Caric, "Traffic Emissions Clustering Using OBD-II Dataset Based on Machine Learning Algorithms," *Transportation Research Procedia*, vol. 64, pp. 364–371, 2022, doi: 10.1016/j.trpro.2022.09.040.

[8] H. A. Ameen, A. K. Mahamad, S. Saon, M. A. Ahmadon, and S. Yamaguchi, "Driving behaviour identification based on OBD speed and GPS data analysis," *Advances in Science, Technology and Engineering Systems*, vol. 6, no. 1, pp. 550–569, Jan. 2021, doi: 10.25046/aj060160.

[9] X. Zheng *et al.*, "Real-world fuel consumption of light-duty passenger vehicles using on-board diagnostic (OBD) systems," *Frontiers of Environmental Science and Engineering*, vol. 14, no. 2, pp. 1–10, Apr. 2020, doi: 10.1007/s11783-019-1212-6.

[10] N. Peppes, T. Alexakis, E. Adamopoulou, and K. Demestichas, "Driving behaviour analysis using machine and deep learning methods for continuous streams of vehicular data," *Sensors*, vol. 21, no. 14, p. 4704, Jul. 2021, doi: 10.3390/s21144704.

[11] G. Hermawan and E. Husni, "Acquisition, Modeling, and Evaluating Method of Driving Behavior Based on OBD-II: A Literature Survey," *IOP Conference Series: Materials Science and Engineering*, vol. 879, no. 1, Jul. 2020, doi: 10.1088/1757-899X/879/1/012030.

[12] J. Gharbins, *Research on assessing knowledge level of mechanics on OBD*, https://www.academia.edu/8632320/Research_on_assessing_knowledge_level_of_mechanics_on_OBD 2022. (Date Accessed Jul 2024).

[13] Meenakshi, R. Nandal, and N. Awasthi, "Obd-ii and big data: A powerful combination to solve the issues of automobile care," *Advances in Intelligent Systems and Computing*, 2021, pp. 177–189, doi: 10.1007/978-981-15-7907-3_14.

[14] A. Barbier, J. M. Salavert, C. E. Palau, and C. Guardiola, "Analysis of Real-Driving Data Variability for Connected Vehicle Diagnostics," *IFAC-PapersOnLine*, vol. 55, no. 24, pp. 45–50, 2022, doi: 10.1016/j.ifacol.2022.10.260.

[15] P. A. M. Campoverde, N. D. R. Campoverde, G. P. N. Quirola, and A. K. B. Naula, "Characterization of Braking and Clutching Events of a Vehicle Through OBD II Signals," in *Advances in Intelligent Systems and Computing*, 2021, pp. 134–143, doi: 10.1007/978-3-030-59194-6_12.

[16] U. Maalik, P. Ponnampalam, and P. Pirapuraj, "Intelligent Vehicle Diagnostic System for Service Center using OBD-II and IoT," *International Conference of Science and Technology*, pp. 209–214, 2021.

[17] M. A. Hamed, M. H. Khafagy, and R. M. Badry, "Fuel Consumption Prediction Model using Machine Learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 406–414, 2021, doi: 10.14569/IJACSA.2021.0121146.

[18] N. Navali, L. Vanajakshi, and D. M. Bullock, "Application of On-Board Diagnostics (OBD) Data for Vehicle Trajectory Prediction," in *Lecture Notes in Civil Engineering*, 2023, pp. 319–328, doi: 10.1007/978-981-19-4204-4_19.

[19] A. F. Fernández, E. S. Morales, M. Botsch, C. Facchi, and A. G. Higuera, "Generation of Correction Data for Autonomous Driving by Means of Machine Learning and On-Board Diagnostics," *Sensors*, vol. 23, no. 1, pp. 1–28, Dec. 2023, doi: 10.3390/s23010159.

[20] H. M. Song and H. K. Kim, "Discovering CAN Specification Using On-Board Diagnostics," *IEEE Design and Test*, vol. 38, no. 3, pp. 93–103, Jun. 2021, doi: 10.1109/MDAT.2020.3011036.

[21] B. Kim and Y. Baek, "Sensor-based extraction approaches of in-vehicle information for driver behavior analysis," *Sensors*, vol. 20, no. 18, pp. 1–20, Sep. 2020, doi: 10.3390/s20185197.

[22] A. Khandakar *et al.*, "Portable system for monitoring and controlling driver behavior and the use of a mobile phone while driving," *Sensors (Switzerland)*, vol. 19, no. 7, pp. 1–17, Mar. 2019, doi: 10.3390/s19071563.

[23] S. Navneeth, K. P. Prithvil, N. R. S. Hari, R. Thushar, and M. Rajeswari, "On-board diagnostics and driver profiling," in *Proceedings of the 2020 International Conference on Computing, Communication and Security, ICCCS 2020*, IEEE, Oct. 2020, pp. 1–6, doi: 10.1109/ICCCS49678.2020.9277449.

[24] R. Kumar and A. Jain, "Driving behavior analysis and classification by vehicle OBD data using machine learning," *Journal of*

Supercomputing, vol. 79, no. 16. pp. 18800–18819, Dec. 13, 2023, doi: 10.1007/s11227-023-05364-3.
[25] 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, doi: 10.1016/j.matpr.2021.07.376.

BIOGRAPHIES OF AUTHORS

Siddhanta Kumar Singh has 25 years of work experience in the fields of software training in engineering colleges, university and multinational companies in IT technologies in India and China. He is working as an Assistant Professor at Manipal University Jaipur. He did his B.E. (Computer Science and Engineering), M.Tech. (Computer Science and Engineering) and Ph.D. (Computer Science and Engineering). The areas of reserach interest are on board diagnostic II, driving behavioural analysis, ML, and IoT. He can be contacted at email: siddhanta.singh@jaipur.manipal.edu.

Anand Sharma is an Associate Professor in the Department of Computer Science and Engineering, Mody University of Science and Technology, Laksmangarh, India. His qualifications are BE (IT), MTech (IT) and Ph.D. He has more than 17 years of experience. He has more than 156 papers published, and 37 conference papers presented, 8 books written, and 7 patents published. He also received the best researcher award, best young scientist award, best academician award. His area of research is ML, network security, and OBD. He can be contacted at email: anand_glee@yahoo.co.in.