



Imbalanced Learning Analysis for Driving Behaviour Prediction Using Naturalistic Driving Data

Paper ID: 12

Antonis Kostopoulos, Thodoris Garefalakis, Eva Michelaraki*, George Yannis

National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou str., 15773, Athens, Greece

*Corresponding author: evamich@mail.ntua.gr



INTRODUCTION

Human behaviour significantly contributes to severe road injuries, underscoring a **critical road safety challenge**. Despite advancements in vehicle technology and infrastructure, risky driving performance such as speeding, distraction and impaired driving remain leading causes of crashes.

Beyond the roadway geometric design, traffic volume, and other risk indicators, **human behavior is the major factor** in provoking severe road injuries. Although there are statistically significant relationships between different features, the exact impact of every single traffic and human behavior characteristic on the volume and severity of road injuries has yet to be found. Thus, focusing on road safety enhancement and optimally detecting and quantifying the relations between different features, recent studies focused on driving behavior analysis by developing appropriate machine/deep learning models.

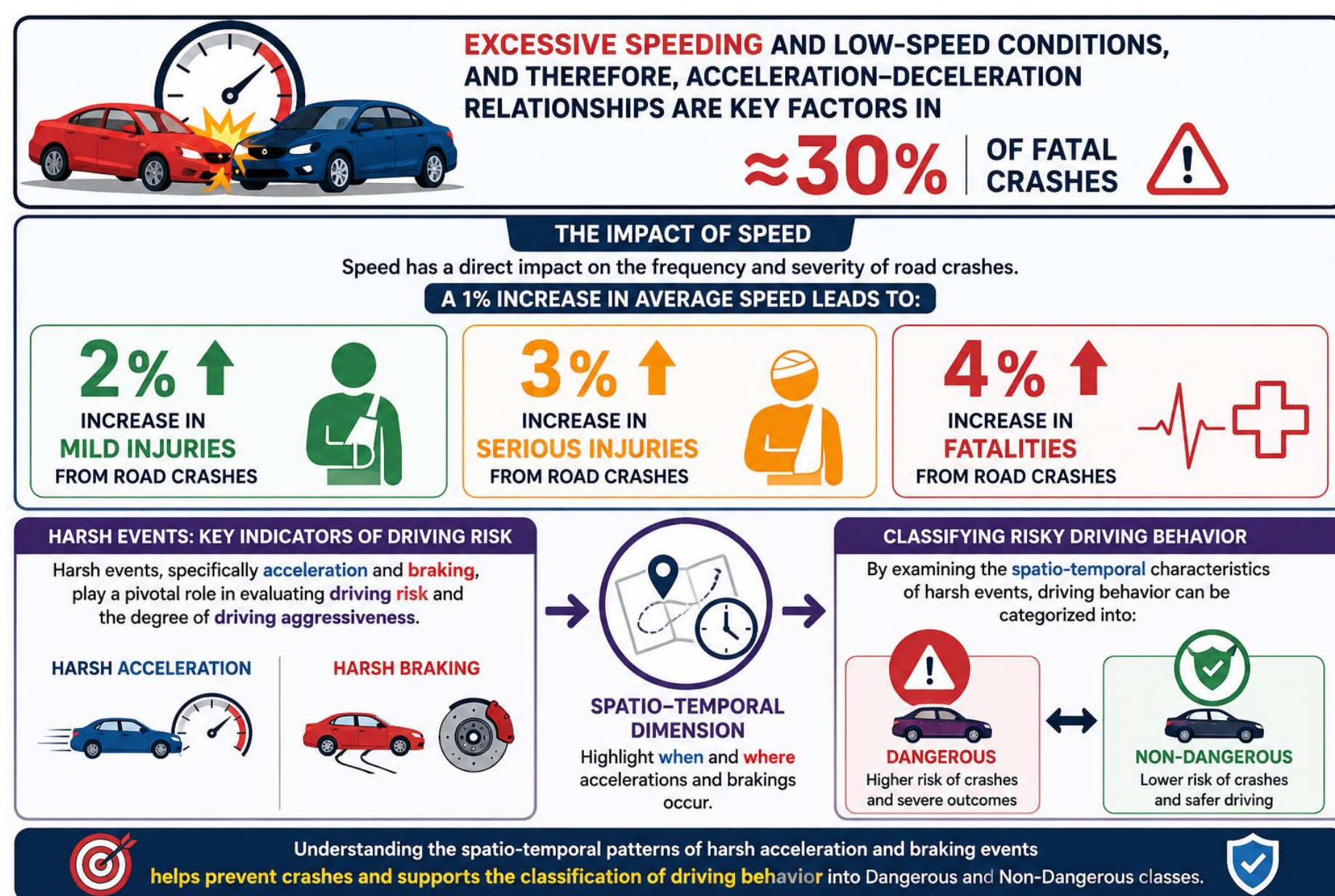


Fig. 1. Conceptual framework and classification of driving behaviour into dangerous and non-dangerous driving profiles

OBJECTIVES

This study tackles the complex challenge of predicting risky driving behaviour through an extensive analysis of more than 356,000 trips, focusing on patterns of harsh acceleration and braking events to highlight their predictive value for road safety. A comprehensive framework for **assessing and classifying driving behaviour as either dangerous or non-dangerous** was proposed. The framework incorporates defining safety levels using clustering algorithms, selecting the most relevant features and addressing issues related to imbalanced datasets.

DATA MANAGEMENT OVERVIEW

Figure 2 illustrates the process that is carried out every time the application records a new trip.

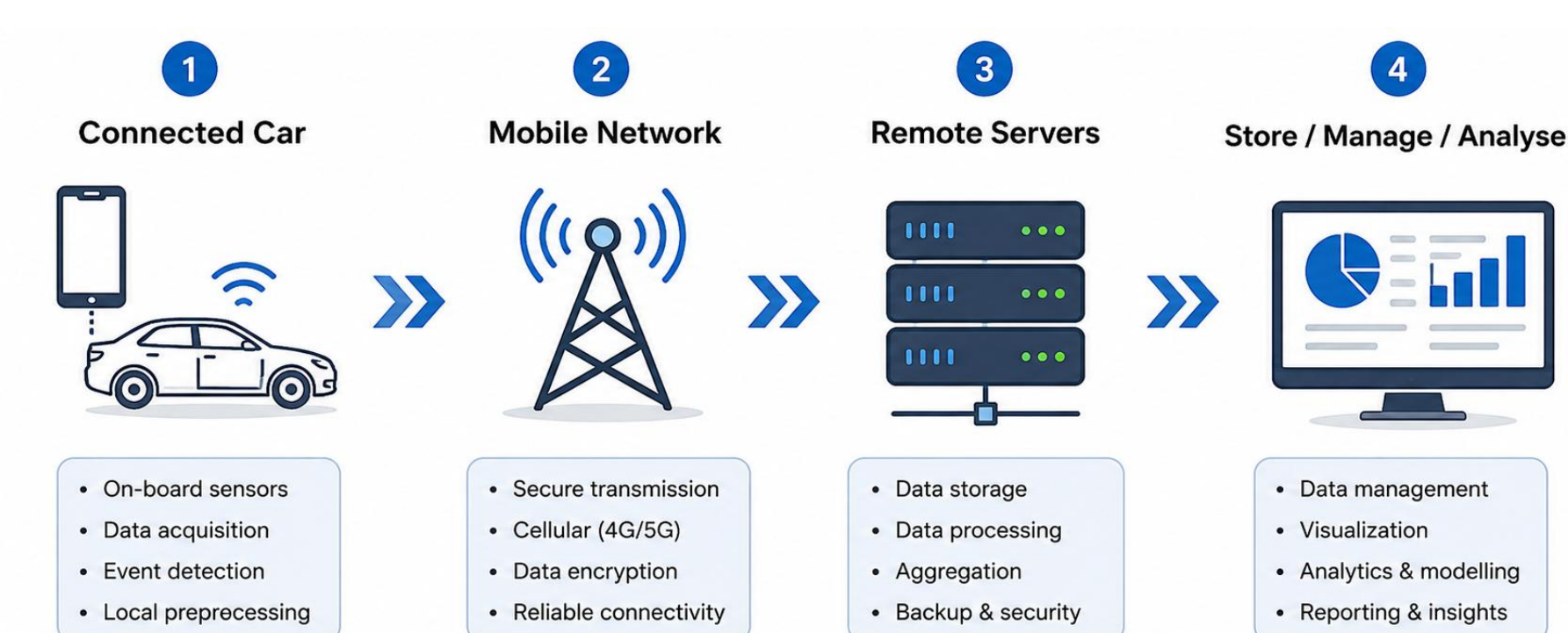


Fig. 2. Overview of the data communication & management system

METHODOLOGY

A **naturalistic driving experiment was conducted** using data provided by OSeven Telematics through a dedicated smartphone application that continuously records driving information without affecting the driving process. The analysis was performed separately for two types of harsh events: harsh acceleration and harsh braking.

Five machine learning classifiers were developed, namely Random Forest, Gradient Boosting, XGBoost, Multilayer Perceptron and K-Nearest Neighbors.

The findings showed that Gradient Boosting and Multilayer Perceptron delivered the **best predictive results**, achieving recall values of about 67% for harsh acceleration and 68% for harsh braking events. The study also identified key thresholds for risky driving behaviour, namely 48.82 harsh accelerations and 45.40 harsh brakings per 100 km. These values offer useful reference points for assessing driver risk. Overall, combining machine learning models with feature selection and k-means clustering proved to be a promising method for improving road safety and reducing related socio-economic impacts.

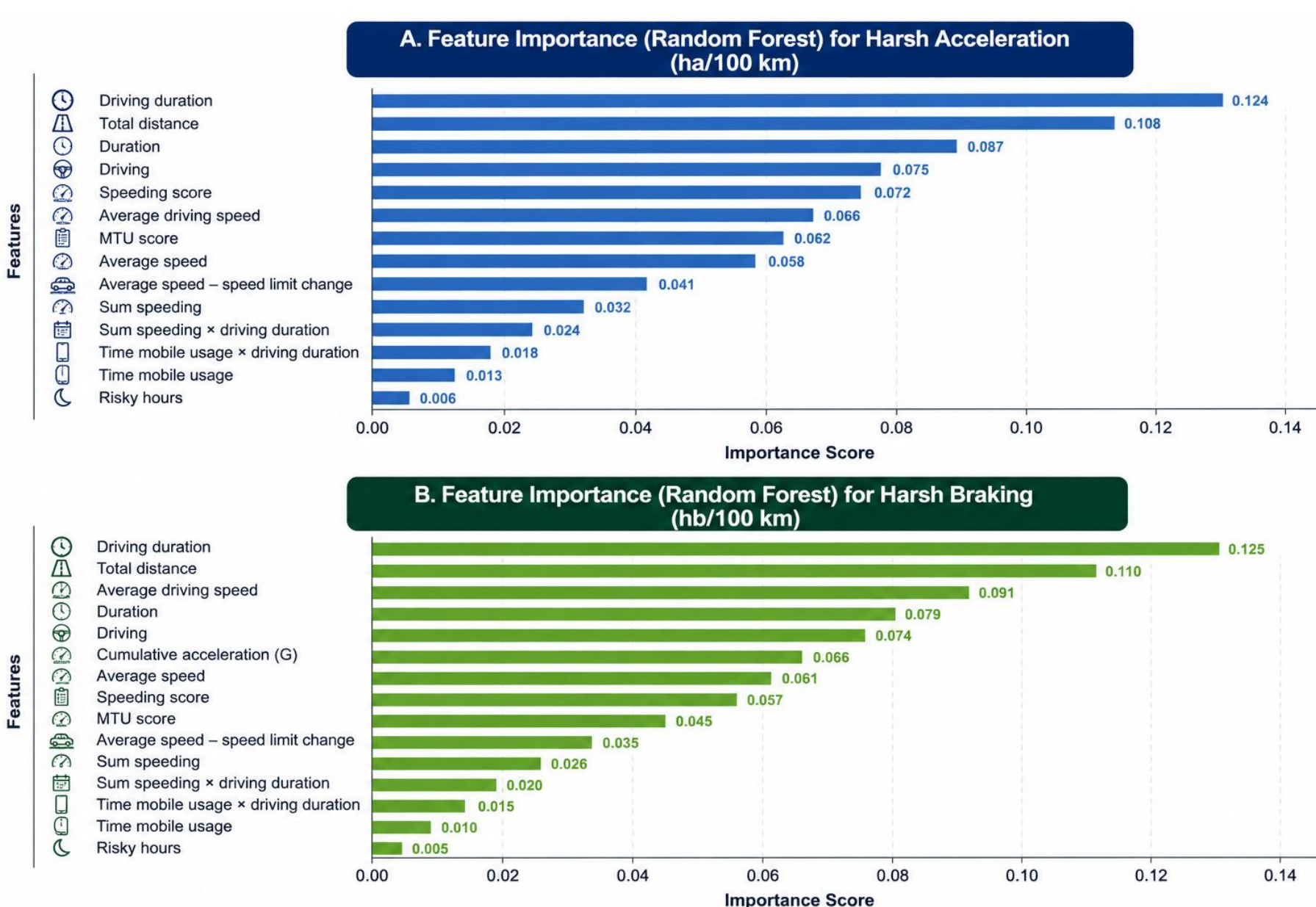


Fig. 3. Permutation feature importance for harsh acceleration/braking events per 100 km

Feature selection was applied to limit the number of input variables, reduce computational requirements and enhance the predictive ability of the classification models. The selected variables were determined according to their contribution to the classification task. As shown in Figure 3, total travel distance and driving duration had the strongest influence on the regression process. In contrast, mobile phone use duration and distance travelled during high-risk night-time hours (00:00-05:00) showed the lowest direct impact.

According to the feature importance, the **input variables for the classification process** for both harsh acceleration and braking events, are total distance, driving duration, average driving speed, speeding score, and mobile use score.

Table 1 displays some **descriptive statistics** for the input variables employed in the classification procedure, including mean, sd, max, min and mean values.

Table 1. Descriptive statistics for input variables

Variables	Description (units)	Mean	St. Dev.	Min	Max
Total distance	Total trip distance (km)	11.60	22.31	0.50	648.68
Driving duration	Total duration of driving, i.e., duration of stops has been excluded (sec)	769.97	967.15	61.00	23900.00
Average speed	Average speed while the vehicle is in motion (km/h)	42.57	17.58	5.57	183.91
Speeding score	Excessive speed score (%)	76.52%	32.92	10.00%	100.00%
Mobile use score	Hand-held Mobile Usage Score while Driving (%)	80.53%	34.62	10.00%	100.00%

RESULTS AND DISCUSSION

Additional performance indicators, including precision, recall, false negative rate and F1-score, were also considered, as shown in Table 2. Overall, **all models showed acceptable predictive performance**, with relatively small differences among them. Gradient Boosting (GB) and Multilayer Perceptron (MLP) achieved the strongest results in terms of recall and false negative rate, with GB performing slightly better than MLP. Both models also produced satisfactory AUC values, reaching 75.1% for GB and 74.7% for MLP, indicating relatively good classification performance.

Table 2. Classification metrics for the developed classifiers for harsh acceleration events per 100 km

Classification Model	Accuracy	Precision	Recall	False Negative Rate	F-1 score
RF	70.83 %	55.16 %	66.39 %	33.61 %	52.64 %
GB	65.28 %	55.15 %	68.05 %	31.95 %	50.25 %
XGBoost	66.76 %	55.09 %	67.46 %	32.54 %	50.86 %
MLP	68.16 %	55.26 %	67.65 %	32.35 %	51.63 %
kNN	72.70 %	53.46 %	60.08 %	39.92 %	51.47 %

Similar to the harsh acceleration analysis, additional evaluation metrics were considered because of the dataset imbalance and the potential risk of the **“accuracy paradox”**. As presented in Table 3, all five models showed satisfactory ability to detect dangerous driving behaviour and achieved generally good performance results. However, GB and MLP performed better than the other models in terms of recall and false negative rate. As noted in Section 5.1, these indicators are particularly important for the objectives of this study. In addition, both GB and MLP obtained acceptable AUC values, with scores of 74.9% and 74.7%, respectively.

Table 3. Classification metrics for the developed classifiers for harsh braking events per 100 km

Classification Model	Accuracy	Precision	Recall	False Negative Rate	F-1 score
RF	67.78 %	57.20 %	66.48 %	33.52 %	55.09 %
GB	63.36 %	57.36 %	67.91 %	32.09 %	53.13 %
XGBoost	64.53 %	57.20 %	67.30 %	32.70 %	53.60 %
MLP	62.96 %	57.29 %	67.80 %	32.20 %	52.88 %
kNN	68.45 %	54.88 %	60.55 %	39.45 %	53.19 %

CONCLUSIONS

- ❖ This paper proposed a **comprehensive framework** for analysing and classifying driving behaviour as dangerous or non-dangerous.
- ❖ The framework involved **defining safety levels** using clustering methods, selecting the most relevant features and addressing class imbalance in the dataset.

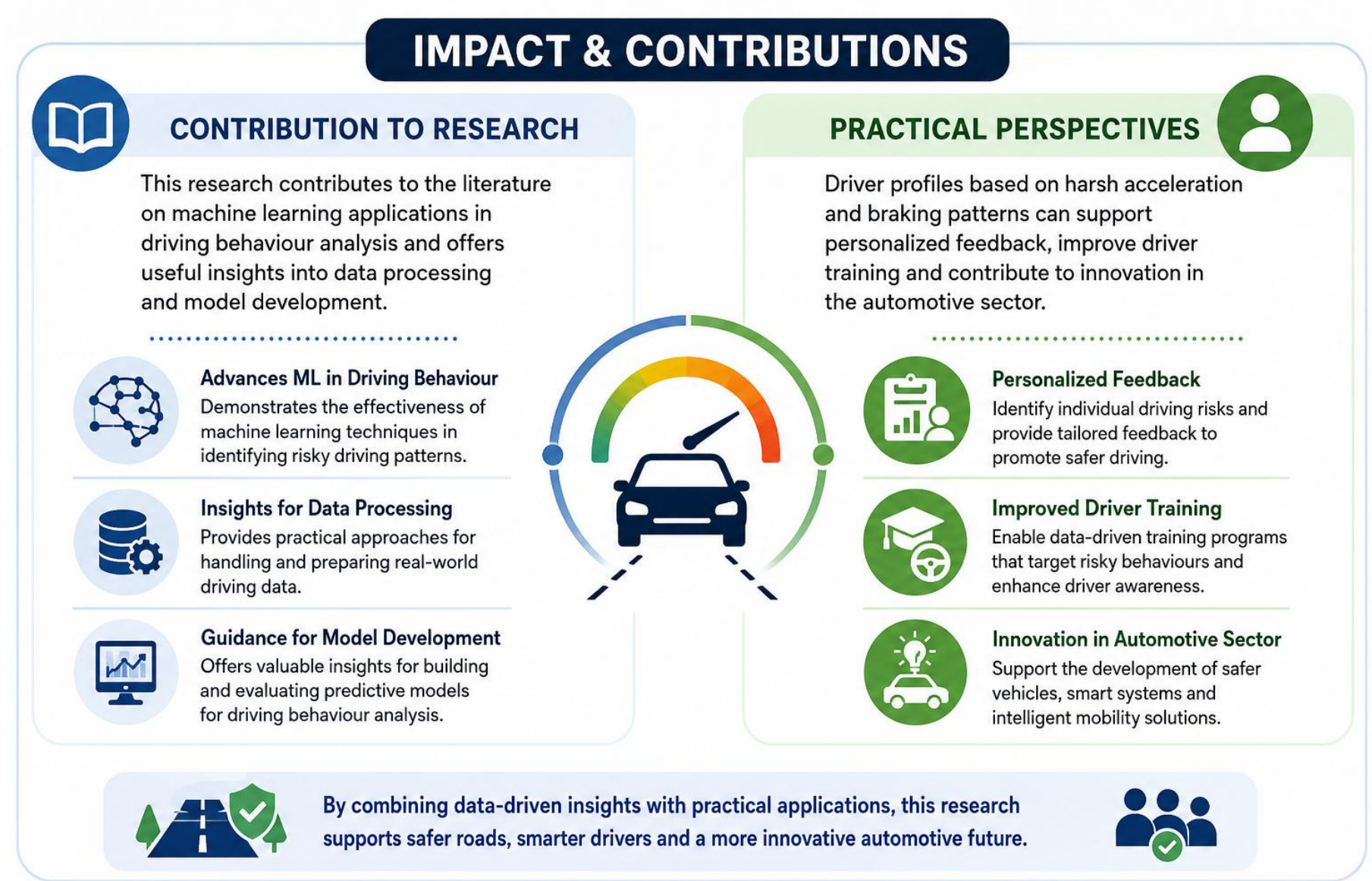


Fig. 4. Impact and contribution of this work