Identifying Crucial Indicators of Task Complexity and Coping Capacity Associated with Crash Risk through Machine Learning Techniques: A Comparative Study using On-Road and Simulator Data

E. Michelaraki[1], T. Garefalakis[1], M. W. Khattak[2], M. Adnan[2], E. Papazikou[3], R. Talbot[3], C. Al Haddad[4], C. Antoniou[4], T. Brijs[2], G. Yannis[1]

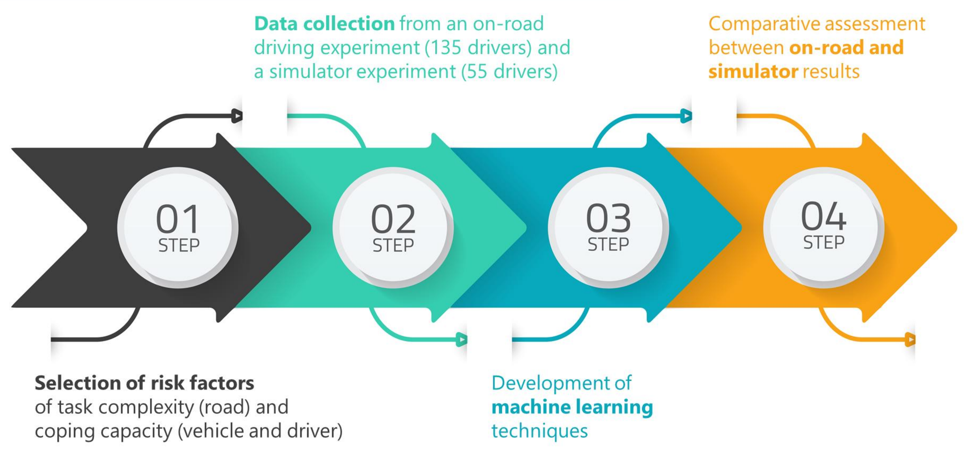
[1] National Technical University of Athens, Department of Transportation Planning and Engineering, Athens, Greece
[2] UHasselt, School for Transportation Sciences, Transportation Research Institute (IMOB), Hasselt, Belgium
[3] Loughborough University, Transport Safety Research Centre, School of Design and Creative Arts, UK
[4] Technical University of Munich, Chair of Transportation Systems Engineering, Munich, Germany

Introduction

The **driving task** can be characterized as the dynamic control task in which the driver has to select relevant information from a vast array of mainly visual inputs to make decisions and execute appropriate control responses.

Within this context, **task complexity** is related to the current status of the real-world context in which a vehicle is being operated. **Coping capacity** refers to the ability of drivers and road systems to manage and respond effectively to various challenges and stressful situations encountered while driving.

Objectives



The Experiments

For the purpose of this analysis, an **on-road driving experiment** was carried out involving 135 car drivers (with total duration of 4 months) and a large database of 31,954 trips was collected. Additionally, a **simulator experiment** was carried out involving 55 drivers (with total duration of 2 months) and a database consisting of 165 trips (55 drivers x 3 driving scenarios) was created. The most prominent driving behavior indicators, such as speeding, headway, duration, distance and harsh events were assessed. The field trials were structured into four phases, while the simulator trials consisted of three phases, as depicted in Figures 1 and 2.

Figure 1: Four phases of the on-road experiment

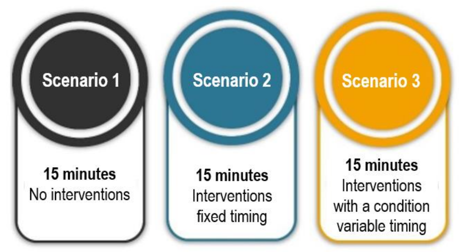


Figure 2: Three scenarios of the simulator experiment

A custom car simulator **developed by DriveSimSolutions** was designed (Figure 3), allowing for creation of custom scenarios and data collection at every simulation update frame. It is also visualized on a triple monitor setup consisting of three 49 inch 4K monitors, providing an 135° field of view (Figure 4).

Figure 3: Car simulator developed by DriveSimSolutions, using OEM Peugeot 206 parts

Figure 4: Example of an intersection in STISIM Drive 3

Methodology

- A feature importance analysis (i.e. **Extreme Gradient Boosting - XGBoost**) was implemented in order to evaluate the significance of various variables in forecasting STZ levels in terms of headway.
- Machine learning analysis (i.e. **Neural Networks**) was applied to make accurate and data-driven predictions by identifying complex patterns between task complexity and coping capacity on crash risk.
- A comprehensive assessment of the performance of three machine learning classifiers (i.e. **Decision Trees, Random Forests and k-Nearest Neighbors**) across different datasets (i.e. on-road and simulator experiment) was performed to predict STZ levels for headway.

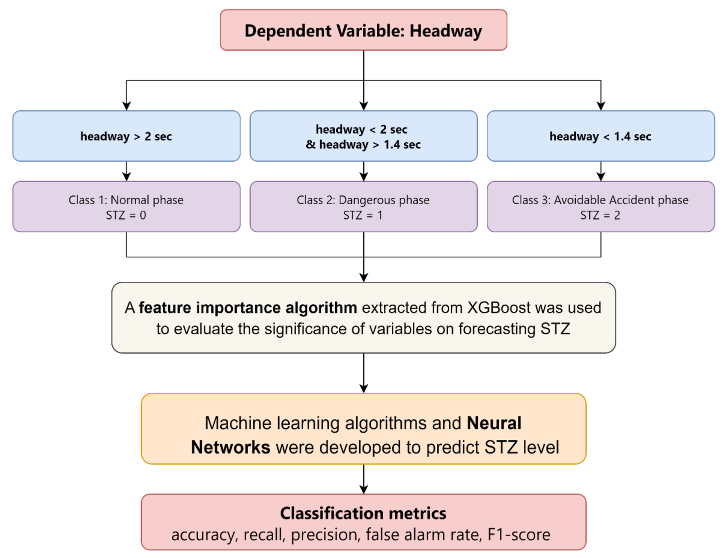


Figure 5: Proposed methodology for the definition of the STZ headway

Feature Importance Results

As shown in Figure 6, the analysis identified that duration, average speed, vehicle age, time indicator, overtaking, gearbox, forward collision warning and car wipers as **the most critical features** for the on-road dataset. On the other hand, distance travelled and fuel type were found to be less significant. As per simulator dataset, it was revealed that time to collision, average speed, duration, hands-on event and fatigue found to be the most influential factors among all examined indicators. Conversely, parameters such as lane departure warning was less significant, while forward collision warning had a negligible impact on STZ headway.

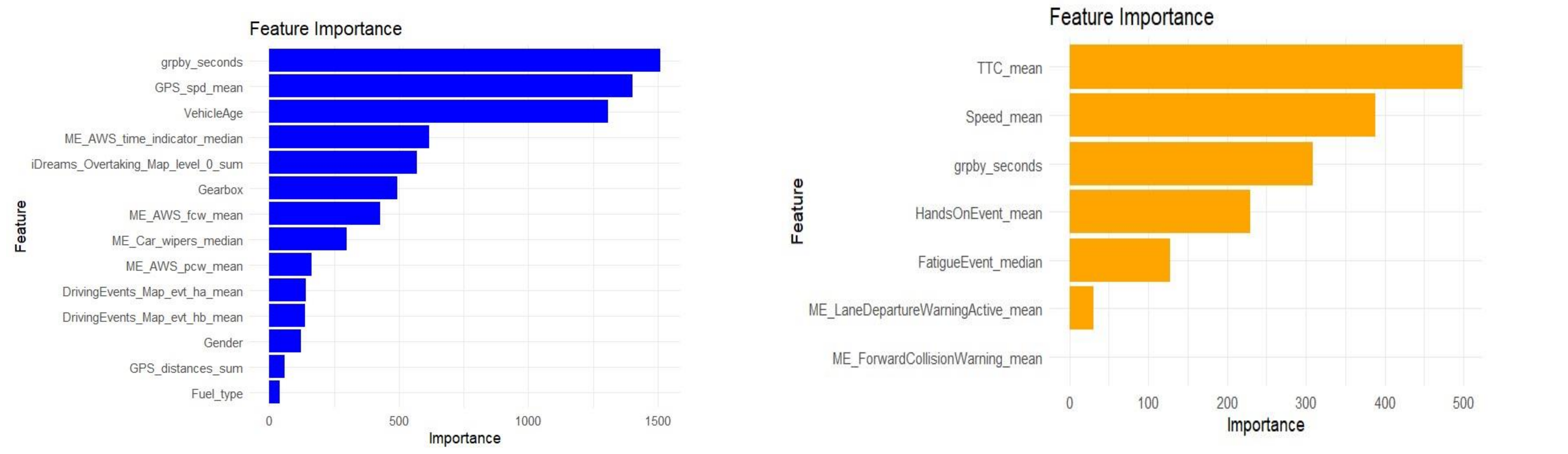


Figure 6: XGBoost feature importance of independent variables (a) on-road (b) simulator experiment

Based on the feature importance and the significance of the relevant indicators, a dataset of 998,358 rows from the on-road and 745,251 rows from the simulator experiment was used and a feed-forward multilayer perceptron NN model was implemented. The data were split into 80% train and 20% test in order to evaluate the models. The model was run with deep neural networks, making use of **two hidden layers** (represented by circles in the middle of the diagram) where the computations take place. Each hidden layer node receives inputs from the previous layer, processes them, and passes the output to the next layer. STZ1 headway refers to normal phase, STZ2 headway refers to danger phase, while STZ3 headway refers to avoidable accident phase.

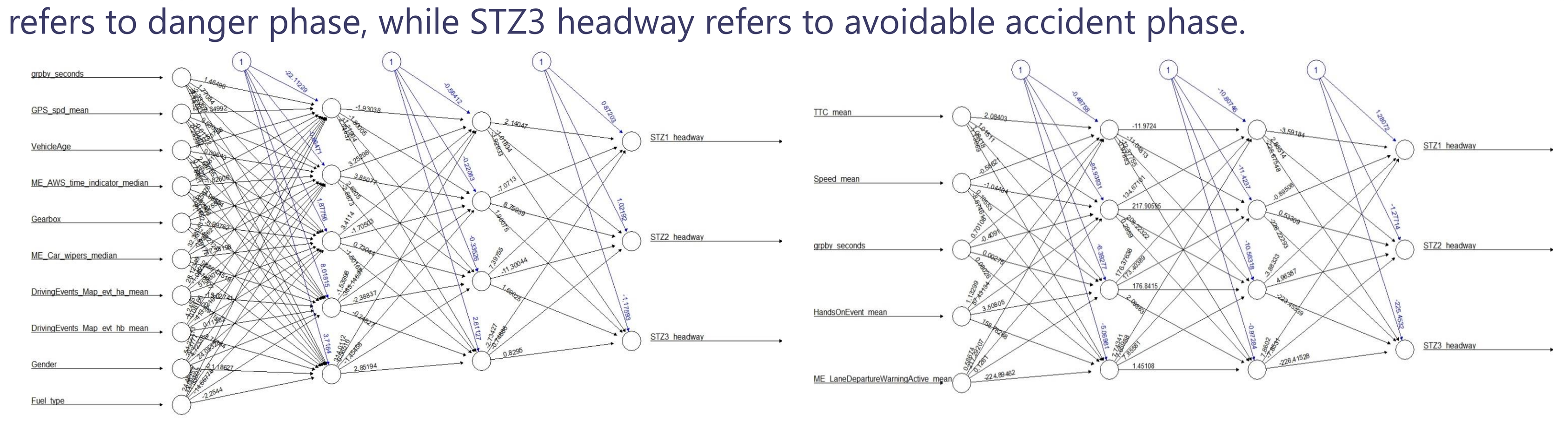


Figure 7: The multi-layer Neural Network model layout for STZ headway (a) on-road (b) simulator experiment

Classification Results

Table 1 provides the assessment of classification model for on-road and simulator dataset. Focusing on the results of on-road experiment, the classifiers achieved 81.7% accuracy, 80.8% precision, 83.4% recall and an F1-score of 81.9%. The **overall accuracy indicates that the model is 81.7%** accurate in making correct predictions, while the recall of 83.4% demonstrates the model's ability to detect safety-critical classes (i.e. "dangerous" and "avoidable accident") effectively.

In the simulator experiment, the overall model metrics were impressive, with an **accuracy of 89.8%**, precision of 91.2% and recall of 90.6%. These metrics indicated that the model was highly accurate in making correct predictions and excels in identifying positive samples, as evidenced by its high precision. The model's ability to detect safety-critical classes effectively was also demonstrated by its high recall. This performance suggested a well-rounded and effective predictive capability for headway in the simulator environment.

Table 1: Evaluation metrics for NN for headway

Model Fit measures	0	1	2	Total
On-road experiment				
Accuracy	0.863	0.852	0.819	**0.817**
Precision	0.845	0.805	0.803	**0.808**
Recall	0.891	0.818	0.744	**0.834**
F1 Score	0.867	0.811	0.773	**0.819**
False alarm rate	0.317	0.413	0.348	**0.392**
Simulator experiment				
Accuracy	0.907	0.973	0.915	**0.898**
Precision	0.876	0.968	0.853	**0.912**
Recall	0.899	0.946	0.842	**0.906**
F1 Score	0.887	0.957	0.847	**0.899**
False alarm rate	0.287	0.114	0.257	**0.153**

**0 refers to normal phase, 1 refers to dangerous phase, 2 refers to avoidable accident phase*

The evaluation of the three machine learning classifiers (DT, RF, kNN) **revealed varying performance** across the two datasets. Figure 8 presents the comparison of classifier metrics of the three machine learning techniques for headway.

In the on-road experiment, RF exhibited higher performance, leading in **satisfactory accuracy (86.9%)** and precision (88.7%), while showing competitive recall scores (90.7%). DT and kNN showed similar performance, though kNN tended to lag slightly behind in precision.

The results from the simulator were similar to those observed in the on-road experiment. In particular, in the simulator experiment for STZ headway, RF emerged as the **top-performing model with an accuracy of 90.1%**, demonstrating its ability to accurately classify driving behavior in a controlled environment. Following the DT model which also performed well scoring a notable 87.1% accuracy. Regarding kNN model, they underperformed compared to the other two, displaying a lower weighted accuracy (85%) and recall (82.6%).

Among the different algorithms, RF stranded out with the highest accuracy of 90% in STZ headway, indicating its ability to **accurately classify driving behaviors in a controlled environment**. RF also achieved a well-balanced precision (87.2%) and recall (84.1%), demonstrating its robustness and versatility.

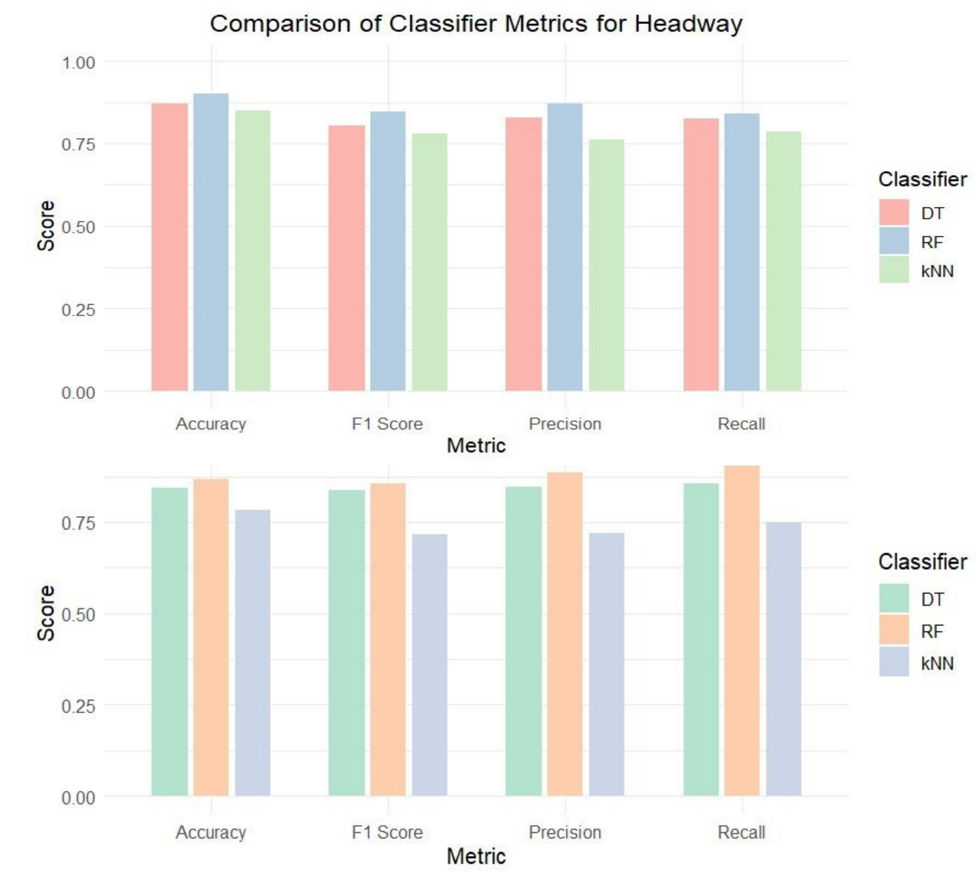


Figure 8: Comparison of classifier metrics of machine learning techniques (DT, RF, kNN) (a) on-road (b) simulator experiment

Conclusions

- The effectiveness of the NN models in **predicting headway levels was encouraging**; the level of STZ can be predicted with an exceptional accuracy of up to 89.8%.
- The three machine learning classifiers (DT, RF, kNN) had **insightful results** in terms of accuracy, precision and recall.
- Results indicated that **RF models outperformed** the DT and kNN models across all metrics, making them the most effective for predicting headway with accuracy up to 90%.
- The DT **model showed satisfactory performance**, while the kNN model consistently had the lowest but moderate scores, indicating that it is the least effective for this task.
- The performance variations observed underscored the importance of selecting the right model based on data characteristics and **precision-recall trade-offs**, essential for real-world applications.
- As per future research, imbalanced learning, factor analysis and models taking into account **unobserved heterogeneity** could be explored for the understanding of the relationship between task complexity, coping capacity and crash risk.

Acknowledgments

i-DREAMS

The research was funded by the EU H2020 i-DREAMS project (Project Number: 814761) funded by European Commission under the MG-2-1-2018 Research and Innovation Action (RIA).

Contact Information:
Eva Michelaraki, PhD, Research Associate NTUA
Department of Transportation Planning and Engineering
Email: evamich@mail.ntua.gr
Website: https://www.nrso.ntua.gr/p/evamich/

104th Annual Meeting of the Transportation Research Board, January 5–9, 2025
TRBAM-25-02616