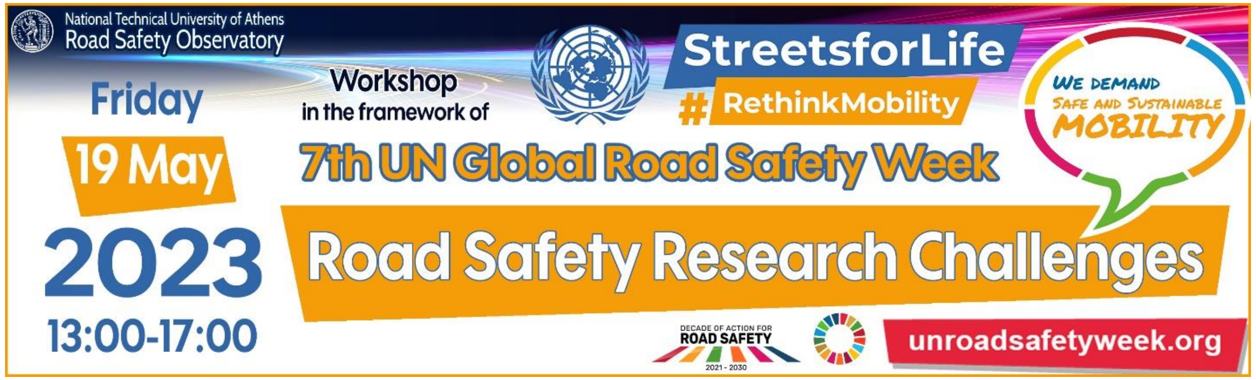


Modelling road infrastructure safety

Dimitrios Nikolaou
Transportation Engineer, PhD Candidate

Together with:
A. Dragomanovits, A. Ziakopoulos, A. Deliali,
I. Handanos, C. Karadimas, G. Yannis

The i-safemodels project

International Comparative Analyses of Road Traffic Safety Statistics and Safety Modeling

➢ Project partners:

- NTUA Department of Transportation Planning & Engineering (www.nrso.ntua.gr)
- OSeven Telematics (www.oseven.io)
- Tongji University (https://en.tongji.edu.cn)
- Third country partners: University of Central Florida (US), Purdue University (US), Loughborough University (UK), German Aerospace Center, DE

➢ Duration of the project:

- 42 months (October 2019 – April 2023)

➢ Operational Programme:

"Competitiveness, Entrepreneurship and Innovation" (EPAnEK) of the National Strategic Reference Framework (NSRF): Greece - China Joint R&D Projects.

European Union
European Regional Development Fund

ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΡΕΥΝΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

EPAnEK 2014-2020
OPERATIONAL PROGRAMME
COMPETITIVENESS•ENTREPRENEURSHIP•INNOVATION

Objective

i-safemodels

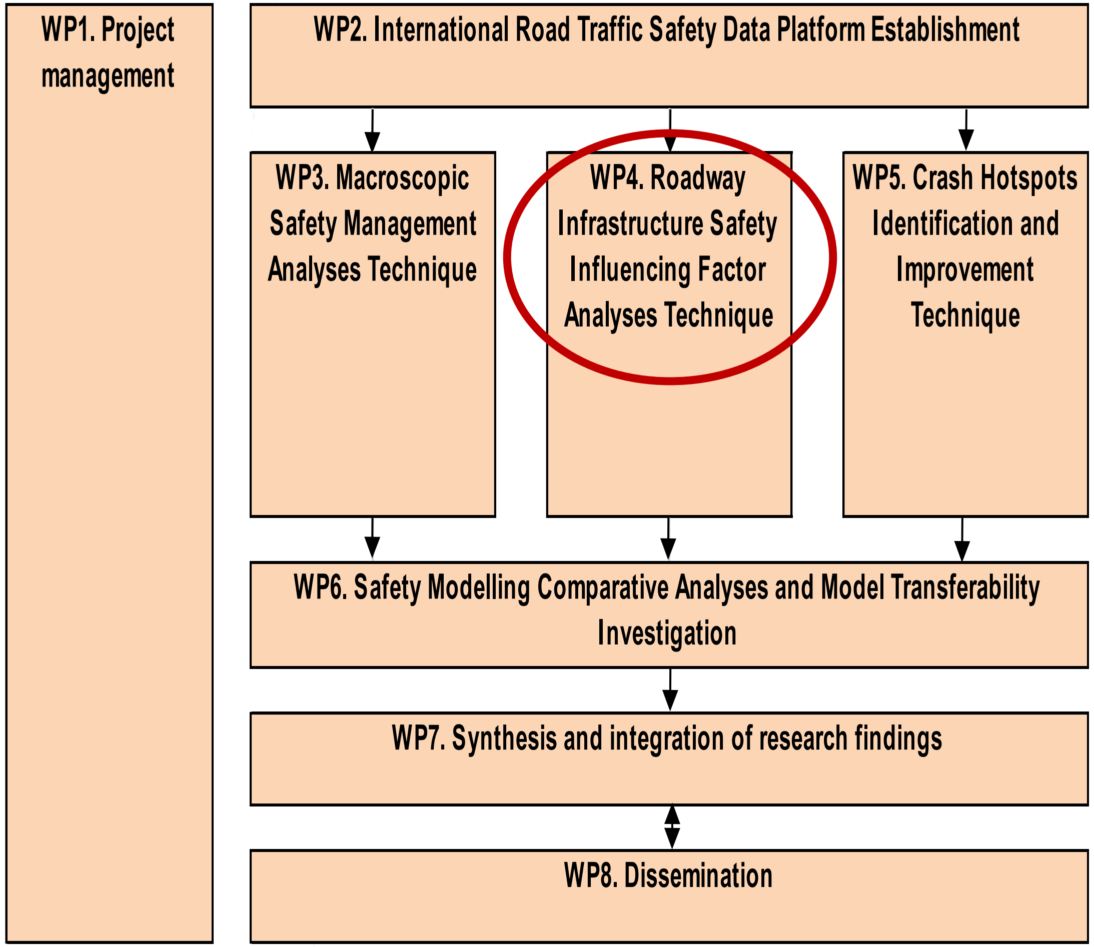
- Development of advanced road safety standardization models at both macroscopic (e.g., country, region) and microscopic levels (roadway segments/sites).

WP4. Road Infrastructure Safety

- Investigate the relationship between road crash frequency in motorway segments and various explanatory variables based on road design characteristics and SSMs.

- Create risk level clusters of the motorway segments based on crash and traffic data.

- Compare the classification performance of five machine learning techniques for predictions of crash risk levels of motorway segments.



Background

- Linking of road safety indicators and road safety outcomes at a global level is a challenging research question that involves various levels, pillars, and modelling relationships.
- At a more microscopic level, Crash Prediction Models (CPMs), including Safety Performance Functions (SPFs) and Crash Modification Factors (CMFs) are essential tools for transport decision makers, to quantitatively predict crashes, analyze injury severity, identify hotspots, and assess safety countermeasures.
- CPMs require detailed data on crashes, geometric characteristics and traffic attributes.
- Apart from such characteristics, in recent years increased attention has been given to Surrogate Safety Measures (SSMs), which are parameters that describe attributes of the network or of the vehicle movement on roads and do not stem directly from or rely on crash data.

Data

- Injury and PDO road **crashes** (Olympia Odos Operation SA)
- **Traffic** (Olympia Odos Operation SA)
- Road **geometry** characteristics (Open GIS, CAD, Google Earth)
- **Naturalistic driving** metrics - SSMs (OSeven)

668 segments (200-600m length) of the Olympia Odos motorway.

- Average AADT (2018-2020): 10,786 vehicles/day
- Average trips per segment (6/2019-12/2020): 2,035 trips
- Road Crashes (2018-2020) 80 injury & 1,270 PDO

Table: Road crash, traffic, geometry and driver behaviour variables per segment.

Variable	Abbreviation	Min.	Max.	Mean
Number of through lanes	lanes	2	3	-
Length of motorway segment (km)	len_seg	0.20	0.60	0.53
Average Annual Average Daily Traffic Volume of motorway segment (veh/day) 2018-2020	avg_AADT_18_20	6,511	22,079	10,786
Posted speed limit (km/h)	speed_limit	90	130	121.7
Number of Total Road Crashes (Injury & Property Damage Only) 2018-2020	TotCr18_20	0.00	13.0	2.02
Number of Total Road Crashes (Injury & Property Damage Only) by segment length 2018-2020	TotCr18_20_len_seg	0.00	30.0	3.9
Curve 1 – Radius R (m)	Curve1	0.00	50,000	2,129
Curve 1 – Length of curve in segment (m)	Lcurve1_in_seg	0.00	600.0	218.2
Lane width (m)	lane_width	3.55	3.95	3.92
Paved inside shoulder width (m)	pav_ins_sh_width	0.50	1.75	0.69
Median width (measured from near edges of traveled way in both directions) (m)	median_width	2.25	23.50	4.96
Distance from edge of inside shoulder to barrier face (m)	dist_edginssh_barf	0.00	0.75	0.04
Paved outside shoulder width (m)	pav_out_sh_width	0.25	4.50	2.77
Distance from edge of outside shoulder to barrier face (m)	dist_edgoutsh_barf	0.00	3.25	0.82
Number of recorded trips	rec_trips	173	5,068	2,035
Average speed (all trips) (km/h)	avg_speed	77.0	153.0	115.9
Average number of harsh accelerations per trip (%)	avgha_pertrip_perc	0.00	9.83	0.21
Average number of harsh brakings per trip (%)	avghb_pertrip_perc	0.00	3.91	0.21
Average number of speeding events per trip (%)	avg_sp_ev_pertrip_perc	1.28	88.61	25.79

Methodology

- **Negative Binomial Regression**
  (count data, overdispersion)

- **Hierarchical Clustering**
  (agglomerative approach)

- **Classification Algorithms**
  (Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors)

- **Classification Performance Metrics**
  (Accuracy, Precision, Recall, F1-Score, and Macro-averaged metrics)

- **SHAP values**
  (model-agnostic method, measure of contribution)

Crash Frequency Regression Model

- Dependent variable of the developed NB regression: "TotCr18_20"

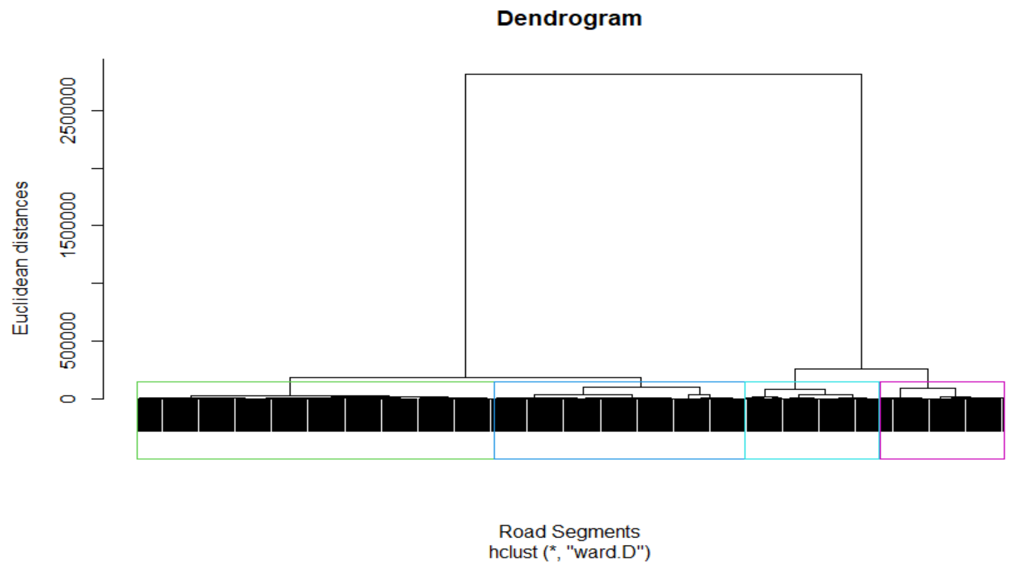
Independent variables	Estimate	Std. Error	z value	Pr(|z|)	VIF
(Intercept)	-1.23636	0.199	-6.216	<0.001	-
avg_AADT_18_20	0.00007	0.000	12.394	<0.001	1.017
avgha_pertrip_perc	14.75934	4.192	3.521	<0.001	1.071
avghb_pertrip_perc	30.00911	6.770	4.433	<0.001	1.037
len_seg	1.93453	0.330	5.856	<0.001	1.055
AICc	2333.837				

- Crash frequency is positively correlated with the average AADT, showing that as traffic volume increases, the number of road crashes increases as well.
- Harsh accelerations and harsh brakings have a positive relationship with the dependent variable, indicating that as the number of these two harsh driving behaviour events increases, crash frequency also increases.
- This finding confirms that harsh driving behaviour events present a statistically significant positive correlation with historical crash records indicating that these metrics can be meaningfully considered as reliable SSMs.
- Lastly, crash frequency is higher for motorway segments with higher length, as length serves as an exposure parameter.

Definition of Crash Risk Levels

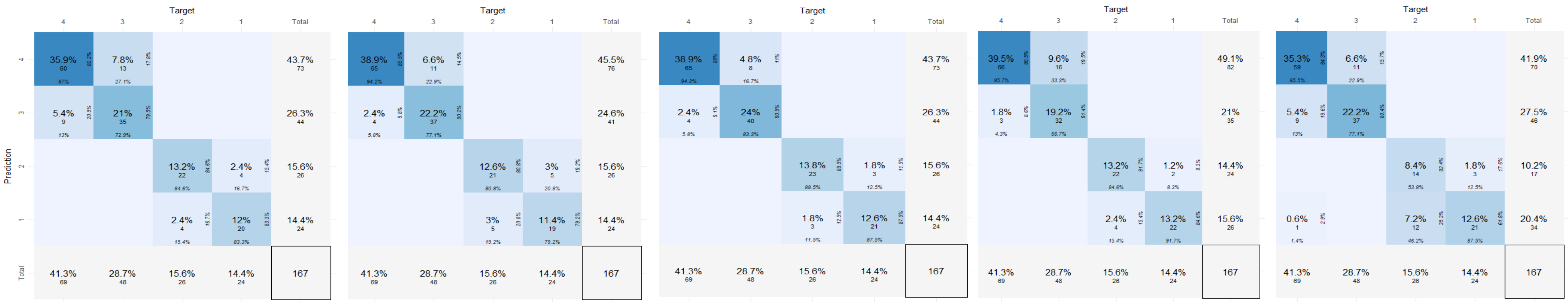
Agglomerative hierarchical clustering

- The Euclidean distance between single observations of the dataset and Ward's minimum variance method as the linkage criterion were used.
- The variables considered for the formation of the risk level clusters of the motorway segments correspond to the number of total road crashes by segment length and the respective AADT of each segment.
- The selection of the number of clusters was based on the produced dendrogram.
- Four distinct clusters representing crash risk levels of the examined segments emerged from the hierarchical clustering procedure, ranging from more risk-prone, potentially unsafe locations to more safe locations.



Crash Risk Level	Count of Segments	Average “TotCr18 20 len seg”	Average “avg AADT 18 20”
1	96	7.57	20,876
2	104	4.55	17,218
3	193	3.25	8,086
4	275	2.76	6,726
Total	668	3.87	10,786

Crash Risk Level Classification Models



Logistic Regression (LR) | Decision Tree (DT) | Random Forest (RF) | Support Vector Machine (SVM) | K-Nearest Neighbor (K-NN)

- **Response variable:** Crash Risk Level
  Predictors: lanes, lane_width, Curve1, Lcurve1_in_seg, median_width, pav_ins_sh_width, pav_out_sh_width, dist_edginssh_barf, dist_edgoutsh_barf, speed_limit, avg_speed, avg_sp_ev_pertrip_perc, avghb_pertrip_perc, avgha_pertrip_perc
- The **training subset (75%)** was used to train the models, while the **test subset (25%)** was used to evaluate their performance.
- **Overall accuracies**: 89.3% for RF, 85.1% for DT, 85.1% for SVM, 82.1% for LR and 78.4% for K-NN.
- **RF classification model** was the best performing model, based on both the overall accuracy and the per-class metrics.

	LR	DT	RF	SVM	K-NN
Crash Risk Level	**Precision (%)**				
1	83.3	79.2	87.5	84.6	61.8
2	84.6	80.8	88.5	91.7	82.4
3	79.5	90.2	90.9	91.4	80.4
4	82.2	85.5	89.0	80.5	84.3
Macro-averaged	82.4	83.9	89.0	87.0	77.2
Crash Risk Level	**Recall (%)**				
1	83.3	79.2	87.5	91.7	87.5
2	84.6	80.8	88.5	84.6	53.8
3	72.9	77.1	83.3	66.7	77.1
4	87.0	94.2	94.2	95.7	85.5
Macro-averaged	82.0	82.8	88.4	84.7	76.0
Crash Risk Level	**F1 score (%)**				
1	83.3	79.2	87.5	88.0	72.4
2	84.6	80.8	88.5	88.0	65.1
3	76.1	83.1	87.0	77.1	78.7
4	84.5	89.7	91.5	87.4	84.9
Macro-averaged	82.1	83.2	88.6	85.1	75.3

SHAP values

- The SHAP values can be positive (green bars) or negative (red bars) for each crash risk level, depending on whether the feature has a positive or negative contribution to the prediction for that class.
- To create a representative instance of motorway segments, the median values of the continuous predictors were used.
- It can be observed that this representative motorway segment is more likely to belong to the lowest crash risk level, which corresponds to overall safer locations with lower traffic volumes and road crashes by segment length than the motorway segments between the first and the third crash risk level.
- The harsh acceleration related variable does not make a significant contribution to the prediction of the segment crash risk level.
- The results of this investigation suggest that harsh brakings may be more pertinent than harsh accelerations for predicting the crash risk level of motorway segments overall.

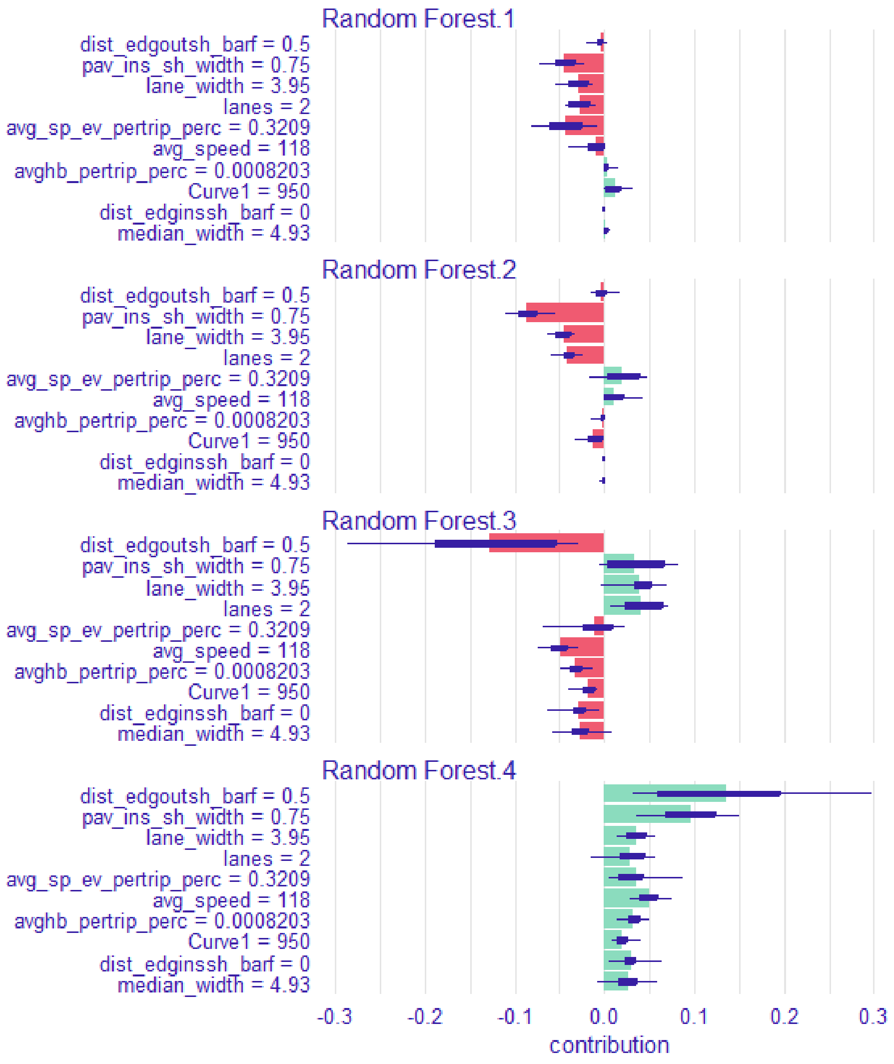


Figure: SHAP values for the RF model and a representative motorway segment

Streets for Life

- Contrary to the EU average (38%), the majority of road fatalities in Greece occurred on urban roads (54%) in 2019.
- The coexistence of all road user types and the high fatality rates in urban road environments highlight the need to apply similar road safety models for urban streets.

Taking into account:

- Vulnerable road users behaviour (pedestrian, cyclists, PTW riders)
- Pedestrian and cycling infrastructure
- Intersection design and signalization
- Traffic calming measures

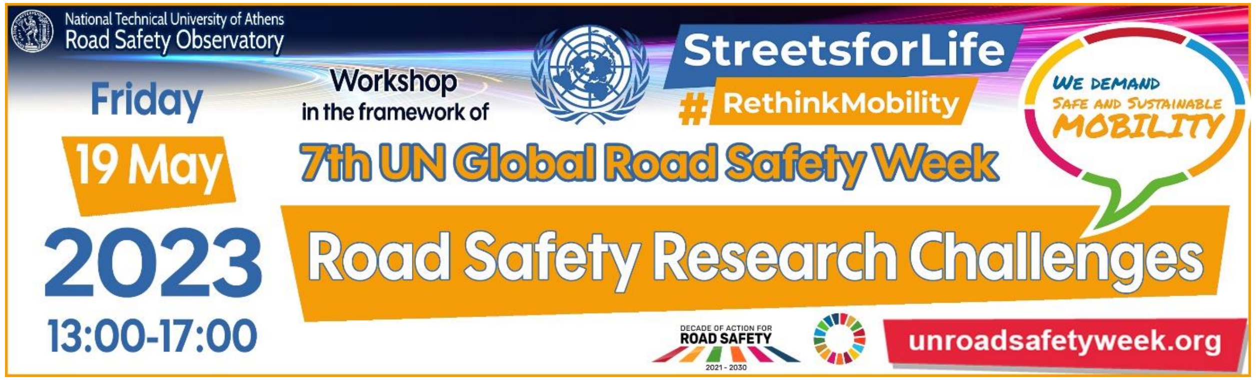
Scientific and Social Impact

- Improvement of decision-making practices
  - Development of models to be used as quantitative tools in decision making.
  - Identification of potentially hazardous road segments.
  - More effective exploitation of available funds for road safety, e.g., selection of more promising countermeasures.

- Reduction of road casualties
  - Improved road infrastructure safety management

Future Challenges

- Enhance the **integration of CPM** techniques at various levels (macro-,meso-, and micro-scopic) to enable advanced road safety modelling internationally.
- Improve road safety **data collection and management** in Greece, particularly at the microscopic level.
- Apply the analyses included in this research to **other road environments**, such as urban and rural roads which are not motorways.
- Encourage Road Authorities and Operators to use **Crash Prediction Models** as a decision-making tool for road safety.



Modelling road infrastructure safety

Dimitrios Nikolaou
Transportation Engineer, PhD Candidate

Together with:
A. Dragomanovits, A. Ziakopoulos, A. Deliali,
I. Handanos, C. Karadimas, G. Yannis