Artificial Intelligence, Big Data and the Future of Road Safety

George Yannis
NTUA Professor

Together with:
Apostolos Ziakopoulos, Research Associate

Department of Transportation Planning and Engineering
National Technical University of Athens

Data Quality and Best Practices of National Access Points
NAPCORE workshop

Athens, May 5, 2025

Outline

1. Introduction
2. Big Data Perspectives
3. Surrogate Safety Measures (SSMs)
4. AI + Big Data=Road Safety
5. AI in Telematics, Driver Monitoring & AVs
6. Distraction & Spatial Approach Examples
7. Pending Barriers
8. Conclusions

Introduction (1/2)

- Road transport is responsible for the majority of transport fatalities, with 1,2 million fatalities worldwide each year.
- Road safety is a field with typically high risk of important investments but not matching results.
- Absence of monitoring and accountability limits seriously road safety performance.
- Very often used to look where the data are and not where the problems and solutions are.

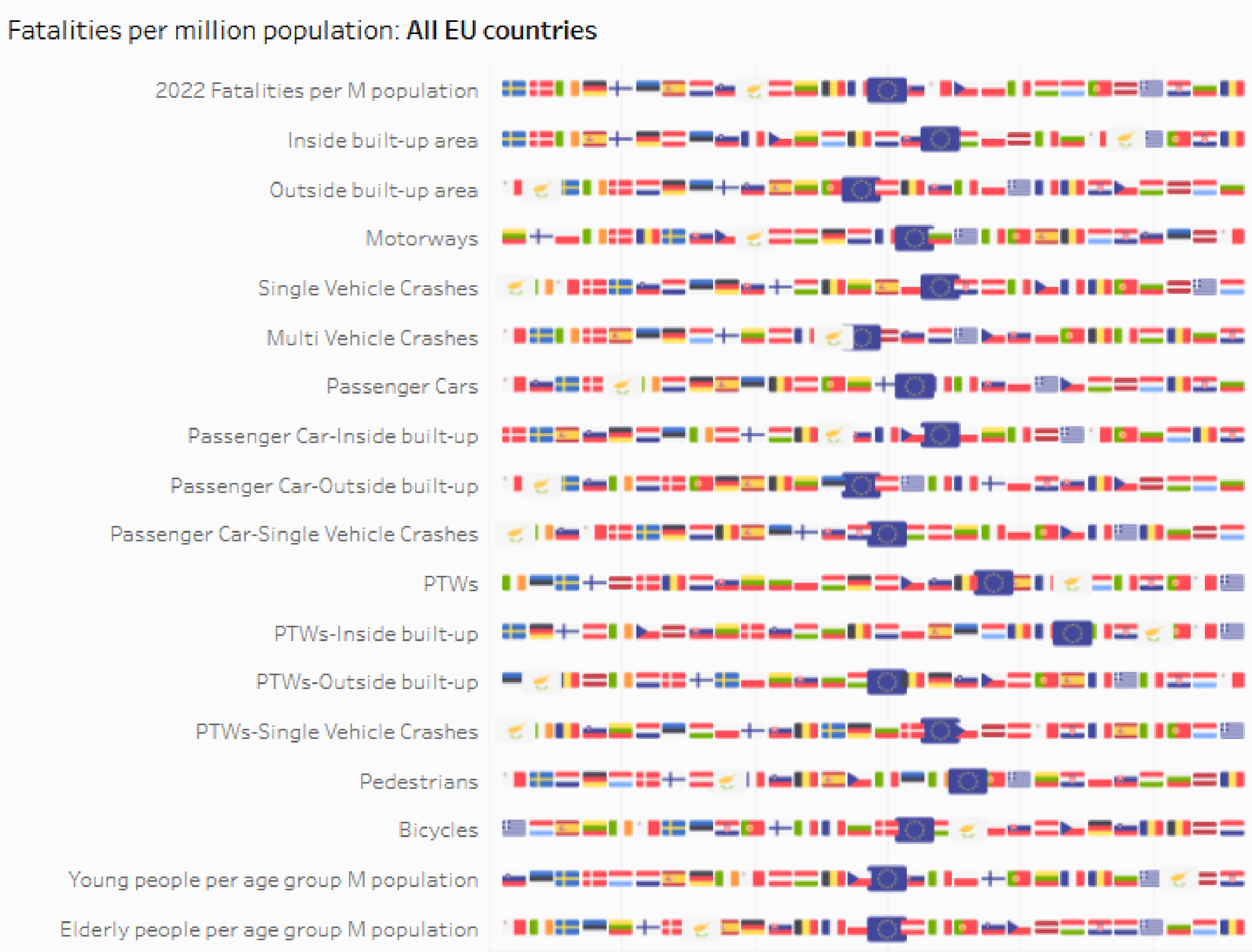
Introduction (2/2)

- Innovative data-driven solutions could contribute to a proactive approach of addressing urban road safety problems, being a core principle of the Safe System Approach.
- The rise of smartphones, sensors and connected objects offers deeper and broader transport data.
- The interpretation of these data can be made possible thanks to progress in computing power, data science and artificial intelligence.

The importance of monitoring

- Total number of fatalities allows for initial comparisons between countries
- Road safety performance may differ considerably with exposure or per crash type
- Disaggregate data can reveal hidden problems or patterns
- Authorities can resolve them with more focused interventions

Available: https://www.nrso.ntua.gr/nrso-ec2/



Date: July 2024, Sources: CARE , Processing: NTUA
2022 data for all EU countries except for Ireland, Latvia (2020) and Malta (2021)

Big Data, Broad Horizons (1/2)

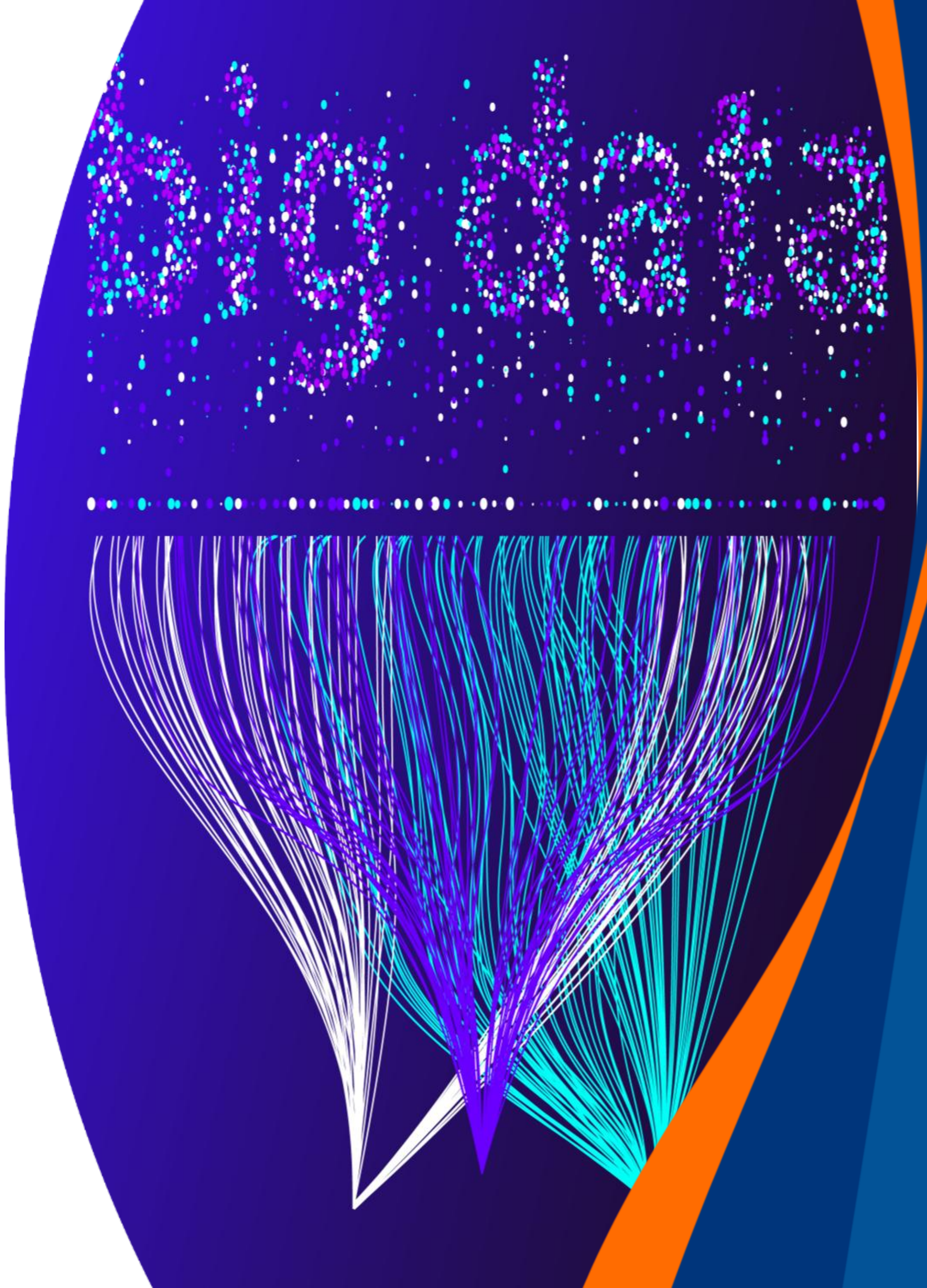
- A wealth of big data becomes available.
- Differentiations per road user category and focus on niche analyses (e.g. VRUs, professional drivers, freight vehicles etc.).
- A multitude of data sources:
- Mobile Phone data:
  - Sensor Based Data (e.g. Google Maps, Here, Waze)
  - Cellular Network Data (e.g. mobile phone operators, etc.)
- Vehicular On-Board Diagnostics data (e.g. OEM industry)
- Camera data:
  - On-vehicle (internal, dash-cam and peripheral)
  - On the road (cities, operators, police)
- Data from Car Sharing Services (e.g. Uber, Lyft, BlaBlaCar)
- Data from Micromobility Operators (e.g. Bolt, Lime, Voi, Tier, Dott)

Big Data, Broad Horizons (2/2)

- **Telematics** companies (e.g. OSeven, ZenDrive, Octo, Floow)
- **Private agency sensor** data (e.g. INRIX, Waycare)
- **Travel Card** data (e.g. Public transport)
- **Public authority** sensor or **traffic measurement** data (e.g. Ministries, Public Transport Authorities, Cities, Regions)
- **Weather** data (e.g. OpenWeatherMap, AccuWeather, etc.)
- **Census** data (e.g. Eurostat, National Statistics)
- **Digital map** data (e.g. OpenStreetMap, Google Maps, etc.)
- **Shared mobility** data (e.g. GPS, routing, etc.)
- **Social Media** data (e.g. Facebook, Twitter/X)
- **Research oriented** data (e.g. instrumented vehicles

Big Data, Big Issues

- The consequences of using data which are **not always representative** of the whole population (bias towards some user groups) should be assessed and properly corrected.
- It is easy to **wrongly consider** a dataset as unbiased if it covers a specific dimension in detail (e.g. covering different road users) while it can fail in another (e.g. not covering exposure).
- **Desired conclusions** should not drive the research approach or outcomes.
- There is a high risk for decision makers to be **misled** by the opportunistic analysis of seemingly low-cost data in absence of **qualified data scientists and statisticians**.



How Open are Big Data?

- Fragmentation of data ownership and a lack of interoperability between datasets and platforms.
- Different interests of the various road safety stakeholders in data, creating differing requirements for data access.
- Data ownership varies by who generates and collects the data and they may be not willing to share data due to privacy, legal liability, IP, competition, or cost related issues.
- Road safety data are often ethically or commercially sensitive.
- The diversity of data sources affecting data quality.
- Systems capacity process big data on traffic and behaviour (real-time, etc.).
- Lack of expertise in machine learning, data mining, and data management with a road safety context.

Surrogate Safety Measures (SSMs)

- Big Data → SSMs, e.g. traffic conflicts, harsh driving events, spatial/temporal headways, and many others.
- Readily available for proactive analyses before crashes occur or in areas with limited or no crash data availability.
- SSMs show less underreporting; can even aid with crash reporting.
- Research on the validation of surrogate safety metrics is essential…
  1. to reveal which metrics not only are correlated with reported crashes but also have predictive capabilities
  2. to forecast the number of fatalities and/or injuries
  3. to determine how these metrics can integrate crash participant fragility, speed, mass and crash type consequences
- More than before, data must not be misused/misinterpreted.

AI in Driver Monitoring

- In-cabin AI can prevent fatigue and distraction by monitoring eye movement, gaze patterns, head or hand position, and reaction times (personalized by driver).
- AI can predict personalized proactive safety measures by analyzing historical driver data aiming at predicting potential safety risks (e.g., aggressive driving or stress).
- AI can be employed in AVs to continuously monitor the driver attentiveness in real-time especially during Take Over Requests (TORs).
- AI can personalize the AV experience by adapting the Human-Machine Interface (HMI) based on the driver preferences and patterns.

AI in Automated Driving

- Depth perception (e.g., LiDAR, radar, etc.)
- Data fusion from environment data (from cameras, lidar, radar, etc.)
- Object recognition and movement prediction.
- Dynamic Decision-Making algorithms (real-time trajectory planning, optimization and response).
- Vehicle-to-Everything (V2X) communication between a vehicle and any entity that may affect, or may be affected by, the vehicle (data exchange).
- Machine Learning for personalized adaptation in-cabin and driving experience.

AI Advances in Road Safety Risk Estimation

- Methods related to Artificial Neural Networks are the most promising for road safety, contributing to ADAS.
- Apart from incident detection, all other problems addressed are mode-specific.
- Knowledge could be transferred from the safety field of AVs to other modes.
- Pattern recognition has received heightened attention (e.g. 85% accuracy of pedestrian detection from video recording using Convolutional Neural Networks)
- However, it remains a challenge to detect and block intentional malicious manipulation of training datasets.

AI + Big Data = Road Safety

AI facilitates the proactive management of traffic safety in various ways:

- Collection of data on road infrastructure conditions and traffic events through wide and broad-scale sensors and systems such as real-time computer vision.
- Identification of high risk locations proactively, through predictive multi-layer models.
- Enabled by multiparametric big data, AI pushes the limits of pattern recognition and reaction times beyond human capabilities and may thus uncover new crash-prone road configurations.
- Recent developments in the field of so-called “explainable AI (XAI)” begin to cope with the challenge of the “black box” phenomenon.

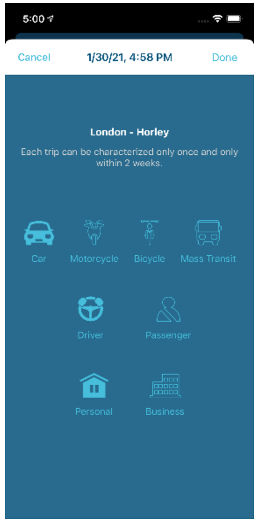
AI in Telematics & Driver Monitoring

- The insurance industry is heavily investing in telematics-based algorithms, offering reduced premiums for safer driving.
- AI and data fusion technologies used in all stages of road safety data collection, transmission, storage, harmonization, analysis and pattern detection.
- Personalized feedback can be obtained almost instantaneously.

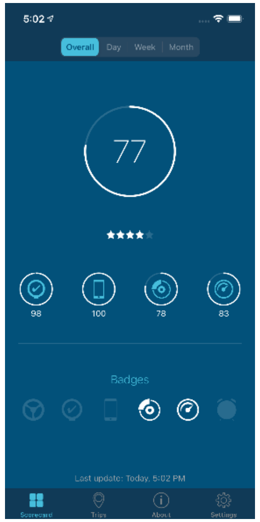
Distraction investigation (1/2)

- An investigation of factors influencing distraction from mobile phone use in naturalistic driving.
- A smartphone application developed by [OSeven telematics]() with 6 feedback phases was the basis for data collection for 87 frequent car drivers.
- Utilizes motion sensors (e.g. accelerometer and gyroscope), position sensors (e.g. magnetometer), global navigation satellite system (GNSS) receivers etc.
- A number of metrics are recorded that can be used as SSMs.

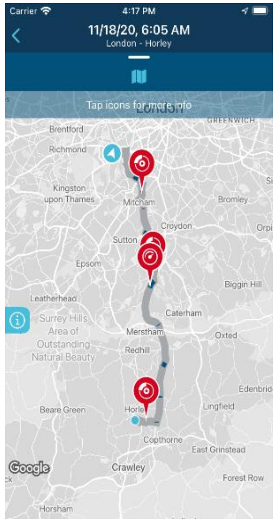
Phase 1 – No feedback



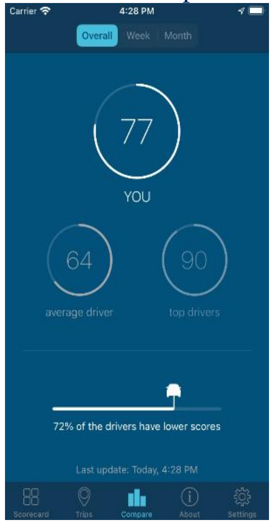
Phase 2 – Scorecard



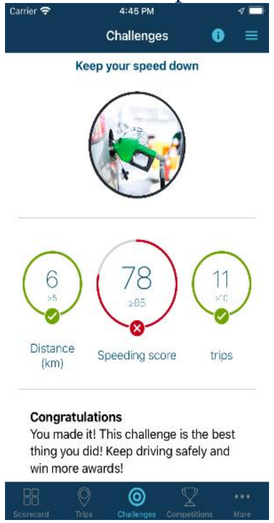
Phase 3 – Maps



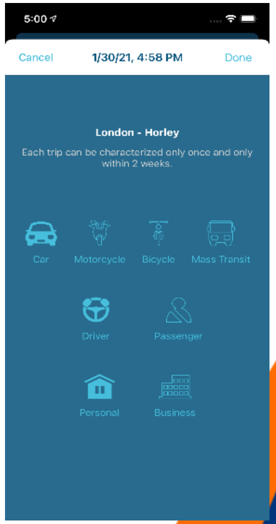
Phase 4 – Comparison



Phase 5 – Competitions



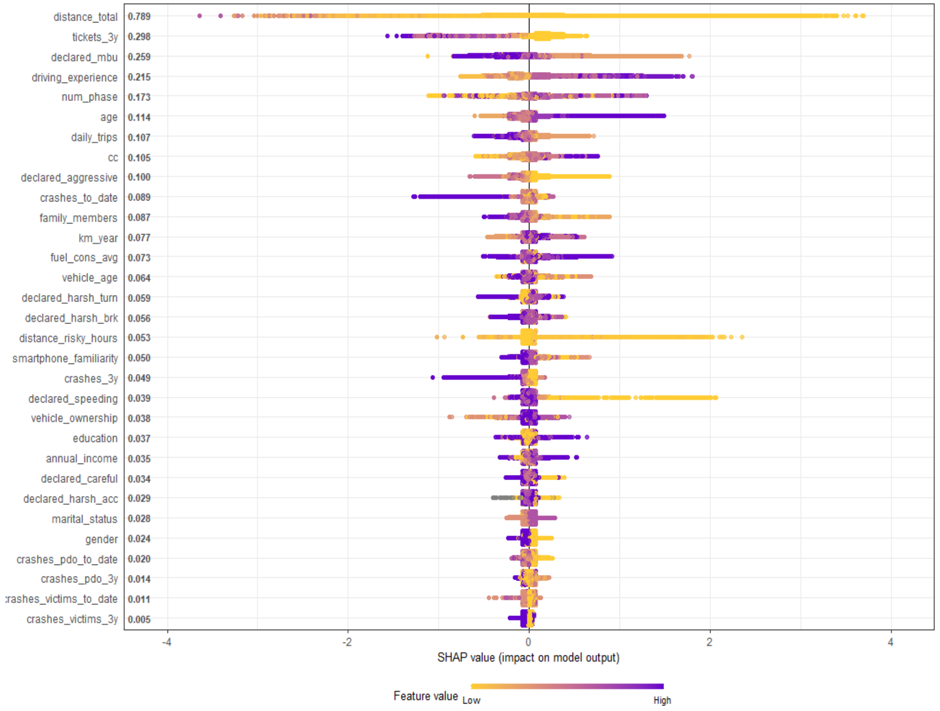
Phase 6 – No feedback

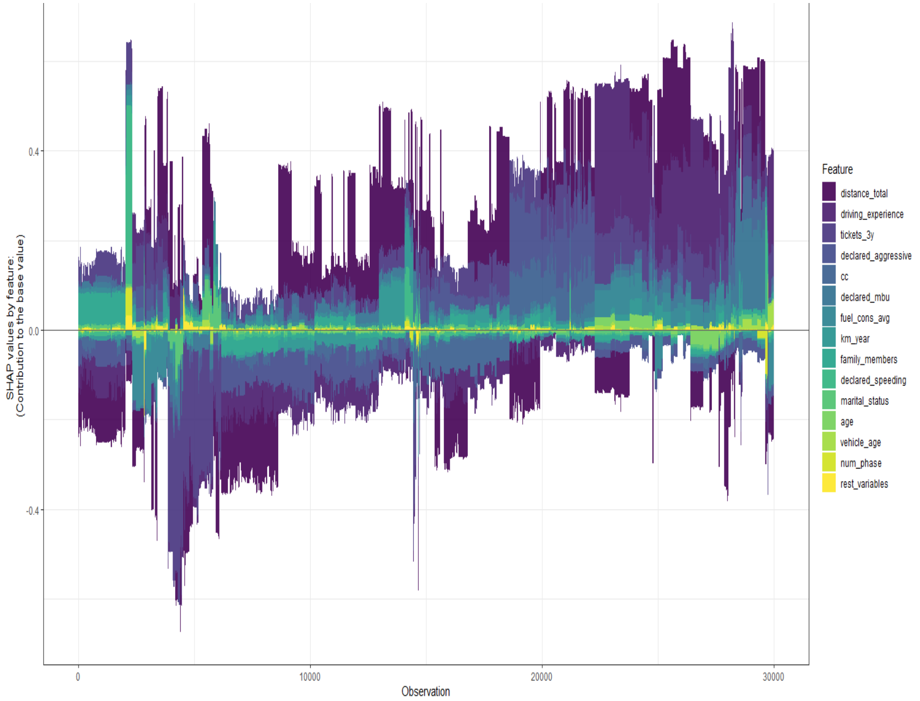


Distraction investigation (2/2)

- Explainable XGBoost tree ensemble **ML algorithms** with SHAP values were trained.
- Higher total trip distance, number of tickets & feedback **decrease mobile phone use**.
- Higher driver age & experience, annual kilometers & engine capacity **increase mobile phone use**.

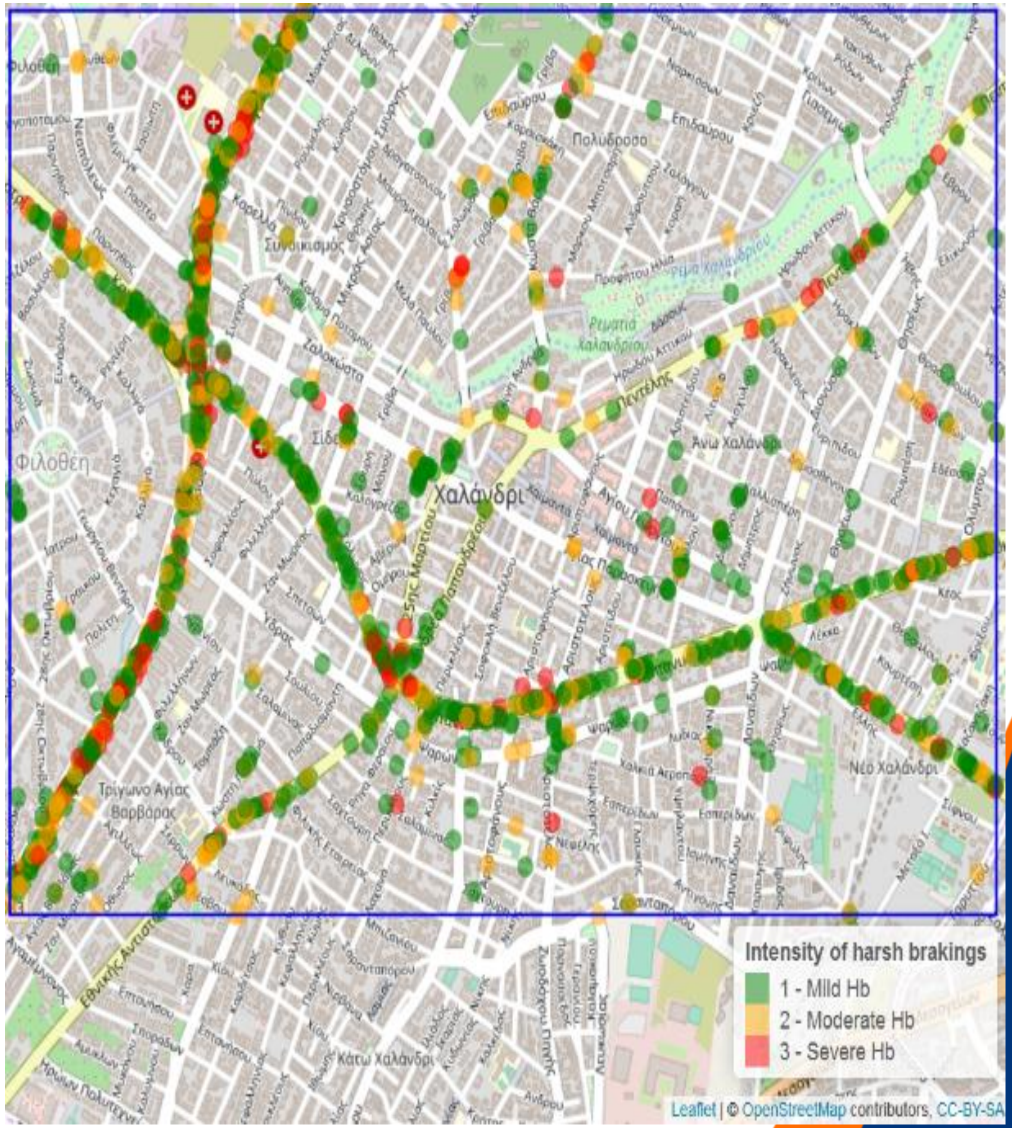
...all in a proactive analysis!





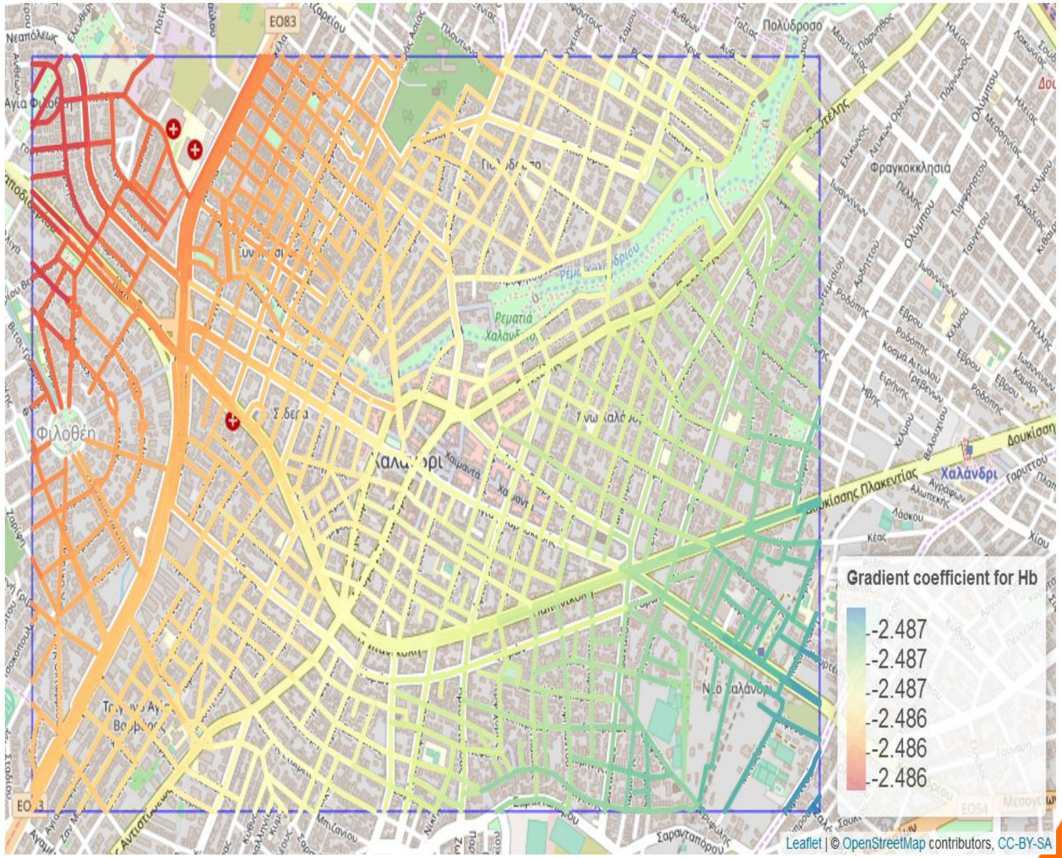
Network Spatial investigation (1/3)

- Smartphone **driving behavior** data & OpenStreetMap **geometric data** are exploited and map-matched.
- **Harsh braking** counts are spatially analyzed in an urban road network.
- 869 road segments (removal of 14 footways) with 4.293 nodes (of which, 49 road with traffic lights, 80 with pedestrian crossings)
- 3.294 trips from 230 drivers, 1.000.273 driving seconds (average trip duration 304s) during 2 months
- 1.348 **harsh brakings** (& 921 **harsh accelerations**…)



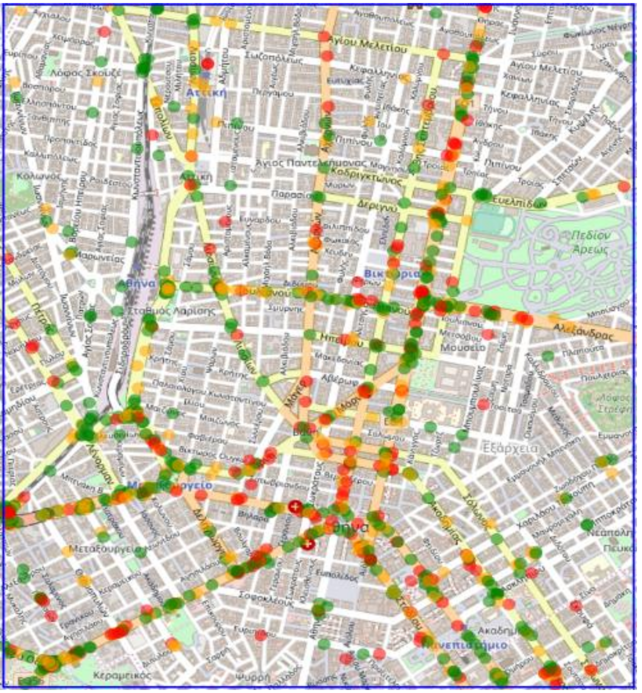
Network Spatial investigation (2/3)

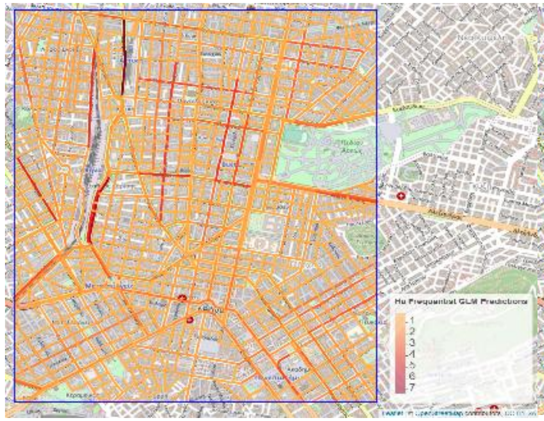
- Statistical models GWPR, CAR, and machine learning XGBoost models (randomly and spatially cross-validated) were trained.
- After adjustments, counts are predicted in another network to assess transferability.
- 87.6% accuracy of harsh braking frequencies was achieved, in a fully proactive analysis.
- Indicative correlations:
  Segment length and pass counts are positively correlated with HBs.
  Gradient and neighborhood complexity are negatively correlated with HBs

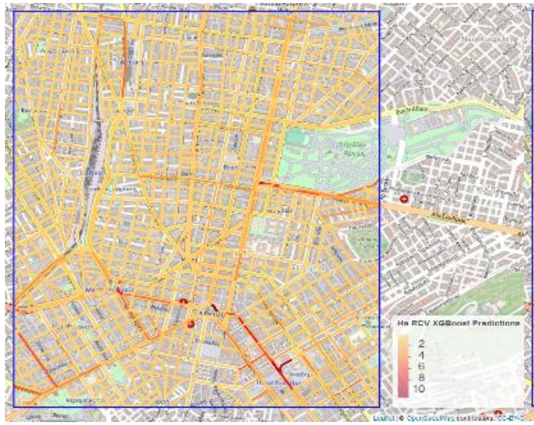


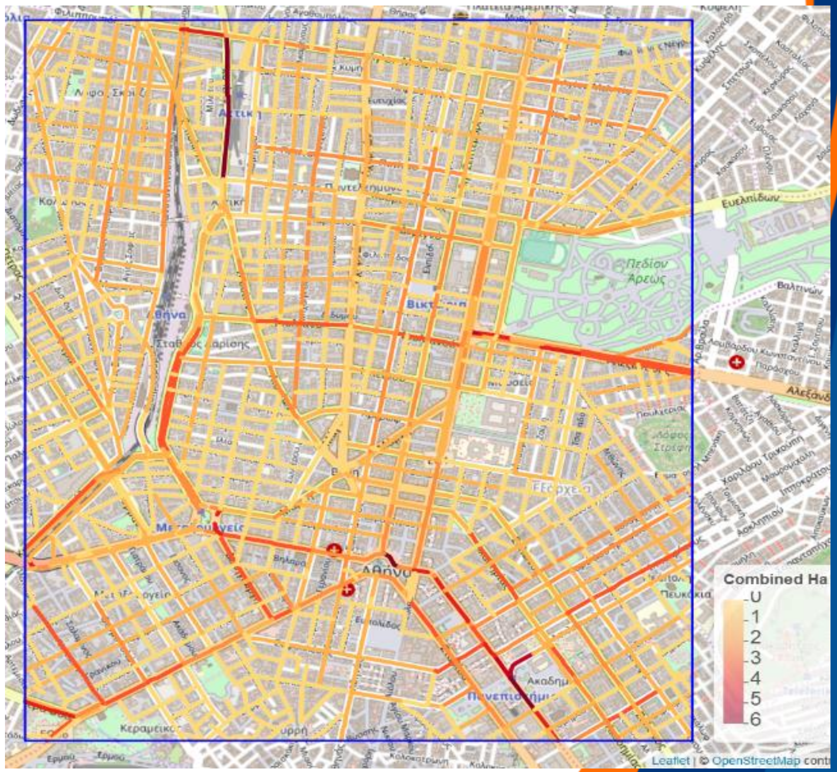
Network Spatial investigation (3/3)

Model weaknesses are covered and strengths are enhanced with combined predictions.



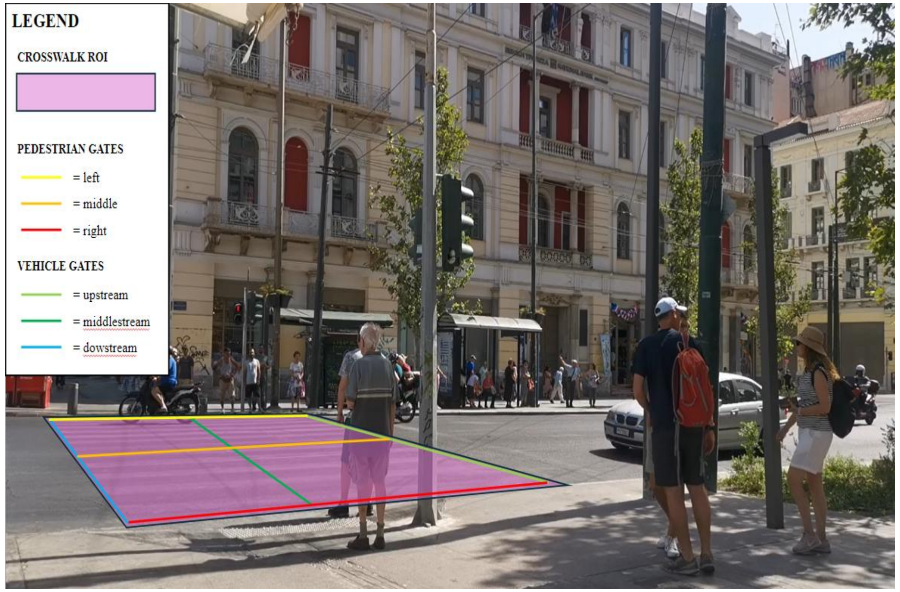






Pedestrian behavior from video recognition

- Pedestrians are analysed with a **multi-step** tracking logic:
  1. Object detection,
  2. Consistency tracking,
  3. Matching of detected and tracked objects
  4. Tracking of movement behind occlusions
- **Traffic light status** is determined
- **Illegal crossings** are inferred based on time and traffic light color
- **Time-to-Collision** with oncoming traffic is calculated
- **Higher accuracy of speed monitoring** by authorities can be achieved.



Pending Barriers for AI

- Safe, road-worthy AI systems face significant challenges that are only hesitantly tackled:
  - Interfaceability
  - Interoperability
  - Timelessness
  - Scalability
- Absence of monitoring and accountability limits seriously road safety performance.
- To counter this, increase acceptance and public trust by monitoring and reporting.
- Research and innovation efforts on the use of AI in computer vision and risk prediction needs more support.

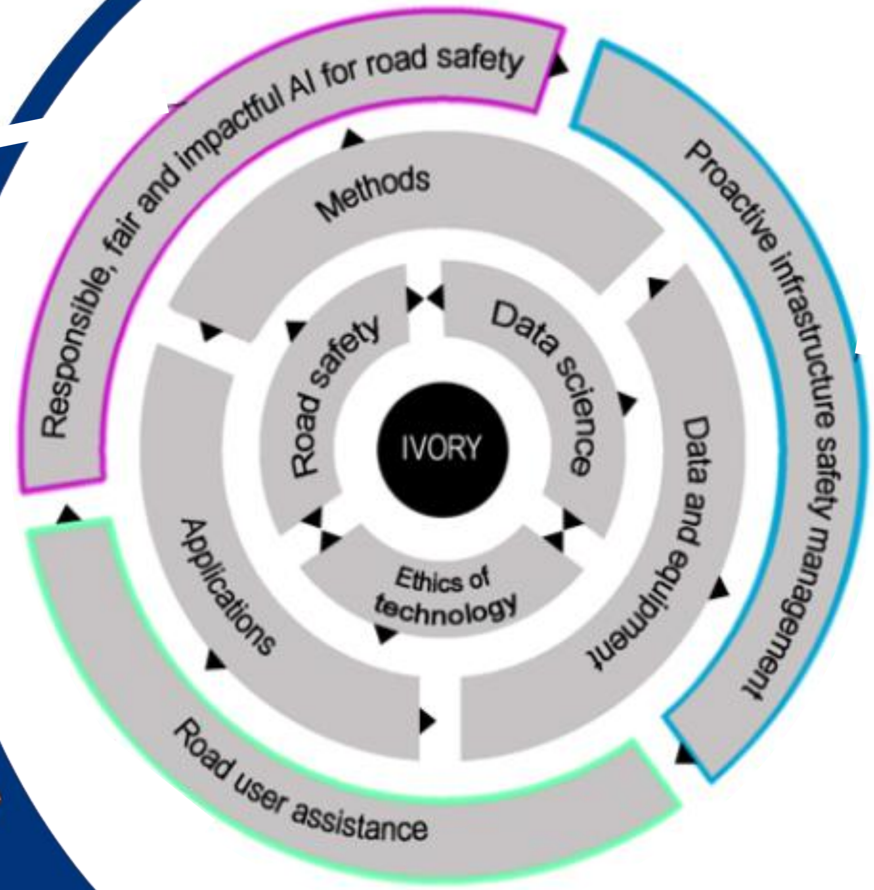
The IVORY MSCA Doctoral Network

Further research: IVORY – 'AI for Vision Zero in Road Safety': https://ivory-network.eu/

- An EC-MSCA Industrial Doctorates Network, aiming to develop a new framework for the integration of AI in road safety and create a new generation of leading researchers

Objectives to be developed:

- Responsible, fair and impactful AI for road safety
- New ways of supporting road users and human-vehicle-environment interaction by means of AI
- New scalable and equitable AI technologies for proactive infrastructure safety management
- A sustainable learning, knowledge sharing and networking framework on AI for road safety



Conclusions (1/2)

- Multiple-criteria based exploration and decision analysis to determine the most efficient Surrogate Safety Measures that can be mined or obtained from the available Big Data.

- AI modelling can reveal complex, non-linear relationships such as factors affecting drivers using a mobile – and be distracted.

- Combining high resolution multi-parametric naturalistic driving, geometric and traffic data to conduct meaningful spatial analyses at segment and network level can be proved highly useful.

Conclusions (2/2)

- Road safety practitioners can rapidly gain by copying best practices for data sharing and privacy protection from other fields.

- Completely unexplored directions remain in several road safety aspects (crowdsourcing options, measure effectiveness, data harmonization).

- Big Data and Artificial Intelligence can become efficient catalysts for achieving Vision Zero road fatalities by 2050.

Key Recommendations

- Integrate lessons learned from telematics & AI for the advent of Connected, Cooperative & Automated Mobility (CCAM)
- Balance carefully between accurate road user recording and protesting of the public due to privacy disruptions and AI-based control
- Foster dialogue between data holders and policymakers for standardization and more openness of data
- Invest into training specialized road safety-oriented computer science professionals

Artificial Intelligence, Big Data and the Future of Road Safety

George Yannis
NTUA Professor

Together with:
Apostolos Ziakopoulos, Research Associate

Department of Transportation Planning and Engineering
National Technical University of Athens

Data Quality and Best Practices of National Access Points
NAPCORE workshop

Athens, May 5, 2025