

Comparison of Manual and Automated Infrastructure Label Extraction for Harsh Event Prediction

Julia Alves Porto

Ivory / NTUA PhD Candidate

Together with:

Apostolos Ziakopoulos, Daniel Lopez, John Avgeros and
George Yannis



European Transport Conference

Data session 3, Video and Resilience

Porto, 9 September 2026



Outline

1. Motivation and Research gap
2. Data: infrastructure inspection + telematics
3. Target variable and baseline model
4. Building computer-based proxies
5. Modelling comparison approach
6. Results: performance and SHAP importance
7. Conclusion, limitations and future work



Motivation

- **1.2 million** road deaths worldwide every year
 - Leading cause of death among people aged 5-29 (WHO, 2023)
 - Vision zero (Tingvall and Haworth, 1999)
 - Reactive vs Proactive safety research
- **Two proactive paradigms with disconnected data**
 - **Infrastructure assessment:** Road Safety Inspection, video-based, manually coded road(side) attributes
 - **Behaviour assessment:** Surrogate Measures of Safety (SMoS): harsh braking and harsh acceleration (harsh events, HE) from telematics
- **The gap**
 - These two streams almost never share a data source, blocking simultaneous analysis of how infrastructure and environment shape driving behaviour



Research gap and objectives

- This study builds a **proof-of-concept** pipeline that **predicts HE occurrence** directly from **road-infrastructure inspection** data, and tests whether the infrastructure label extraction can be automated with **computational tools**

Objective 1

Combine video-based infrastructure coding (iRAP) with telematics-derived harsh events to predict HE occurrence.

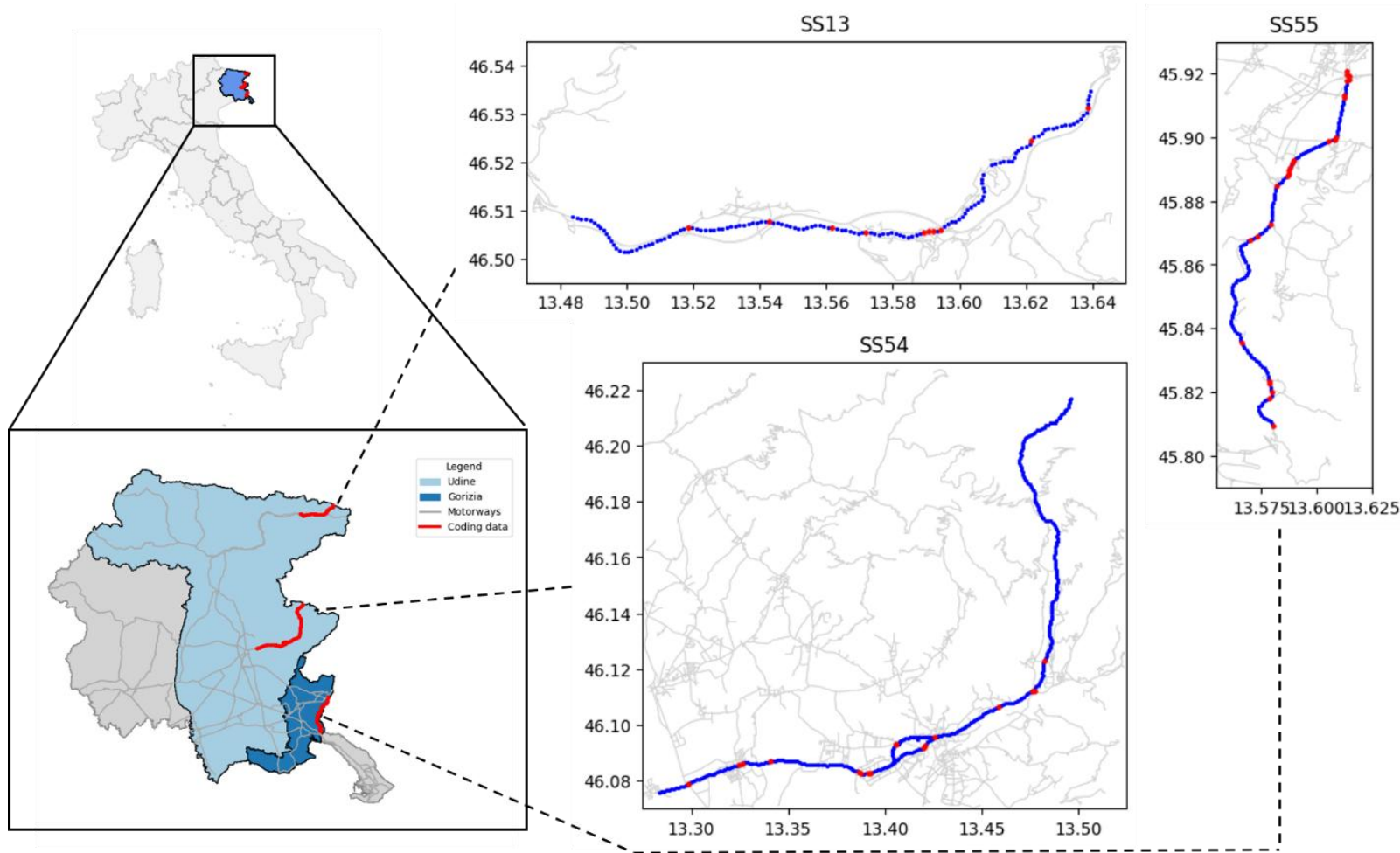
Objective 2

Test whether replacing manual iRAP coding with computer-vision / open-data proxies changes model predictability and attribute importance.

- Proof-of-concept scope: ~64km of road, 3 routes, northern Italy, 322 trips
- Not intended as a generalizable causal model of crash risk



Data: summary and study area



Road	km	Trips	HE Segm.
SS13	14.3	75	10
SS54	34.8	30	17
SS55	15.6	217	24

92% of segments recorded zero HE: strongly zero-inflated target

- **Manually coded iRAP**
 - CSV attributes at 100-m intervals, 9 videos (2304x1296, 29.97 fps) and matching KML GPS tracks. Surveyed 17-18 Sept. 2024, rural tertiary roads
- **OSeven telematics**
 - 322 trips (2021-2025), 1-second GPS + HE flags, joined to the nearest iRAP segment (100-m threshold)



Target variable and base model

- **Three candidate targets x four model families** were screened after iRAP variable curation
 - Zero-variance exclusion
 - High correlation exclusion
 - Semantic aggregation
 - VIF analysis

Target variable	Econometric	Random Forest	XGBoost	CatBoost
Has event (TP count)	0	2	3	2
Number of events (TP count)	0	4	5	1
Speed ratio (R^2)	0.079	0.120	0.151	0.090

- Selected: **XGBoost** predicting the **number of HE per segment** (Poisson count objective)
 - Highest TP count
 - $R^2 = 0.223$



Linking video, GPS and telematics

- 100-m segments, every iRAP segment retained
- Per segment: **trip count**, telematics-point count, **summed HE** and **mean speed** computed from joined points (closest, 100-m limit)
- For the automated pipeline, KML tracks are split into matching 100-m segments, with **gap-filling and boundary duplication** so adjacent segment geometries stay continuous
- **Frame ranges are mapped** to each segment for image-based attribute extraction
 - 2 of 600+ segments were dropped for insufficient points.



Building computer-based proxies

- Six computational sources, retained if it reached $\geq 50\%$ model accuracy or $|\text{Pearson } r| \geq 0.20$ against the manually coded iRAP variable

Semantic segmentation

- DeepLabV3+ (Cityscapes)
- Buildings, sidewalk, vegetation, signs/lights/poles, walls

Object detection

- YOLO11m on DORIE dataset
- Guardrail, bollard

Lane model

- Dual-branch lane segmentation
- Mean detection confidence as a delineation-quality proxy

Curvature

- Cubic B-spline fit to the KML trajectory, 10 knots
- Minimum radius per 100-m segment

TomTom API

- Traffic sample size (volume proxy) and posted speed limit per road segment

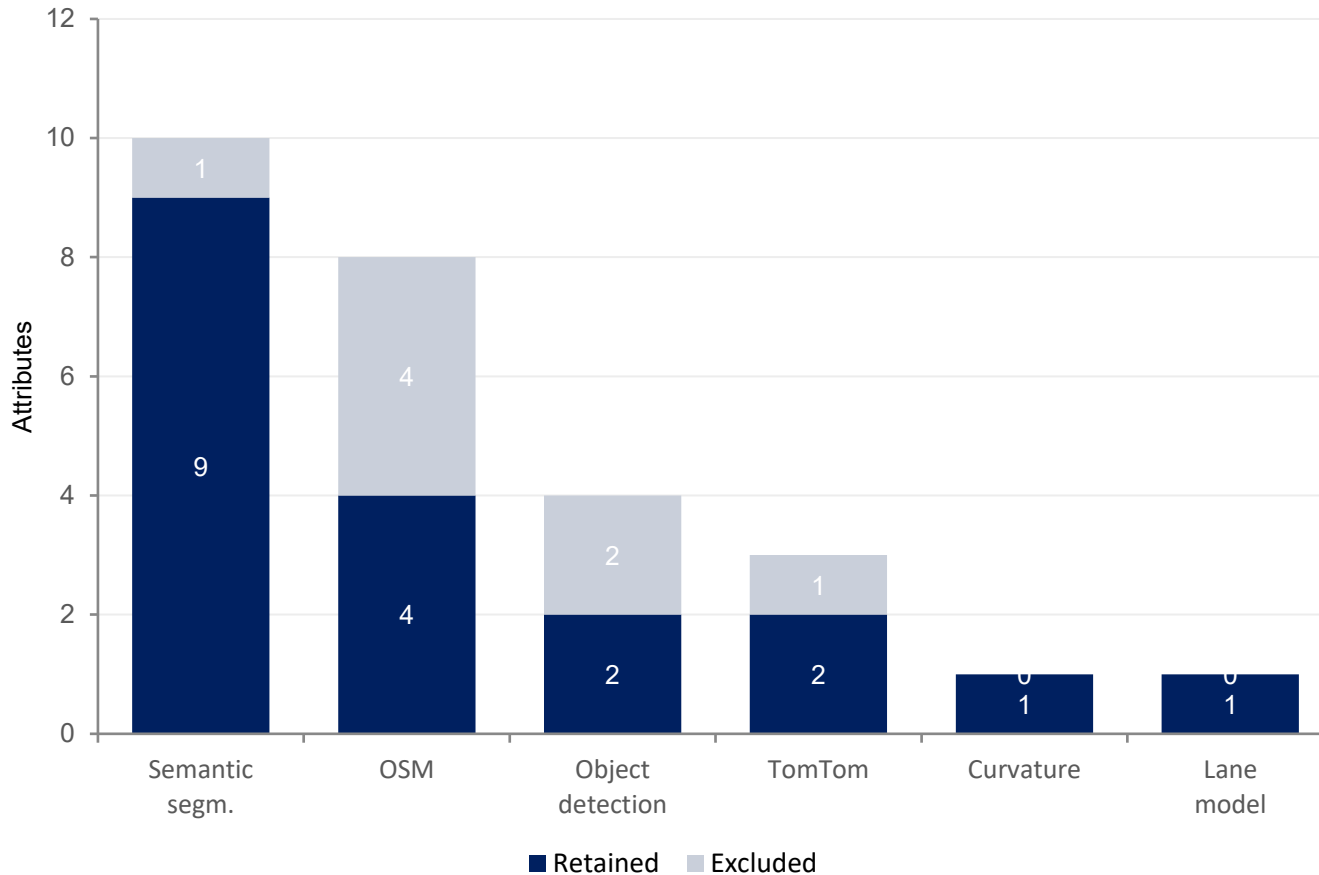
OpenStreetMap

- Buildings and roundabout
- Graph-based intersection



Building computer-based proxies

Proxy attributes by source technique



➤ 19 of 27 candidate attributes retained

➤ Excluded for failing the quality threshold:

- Light poles (44% IoU)
- Intersecting-road volume (~ 0 rho)
- Residential land use (~ 0 rho)
- Merge lanes (~ 0 rho)
- 3/4-leg intersections (0.18, 0.08 rho)
- Bollards (23% mAP)
- Urban/rural land use (~ 0 rho)

➤ Excluded for absence of fair proxy:

- Property access points
- Paved shoulder
- Median type
- Object distance

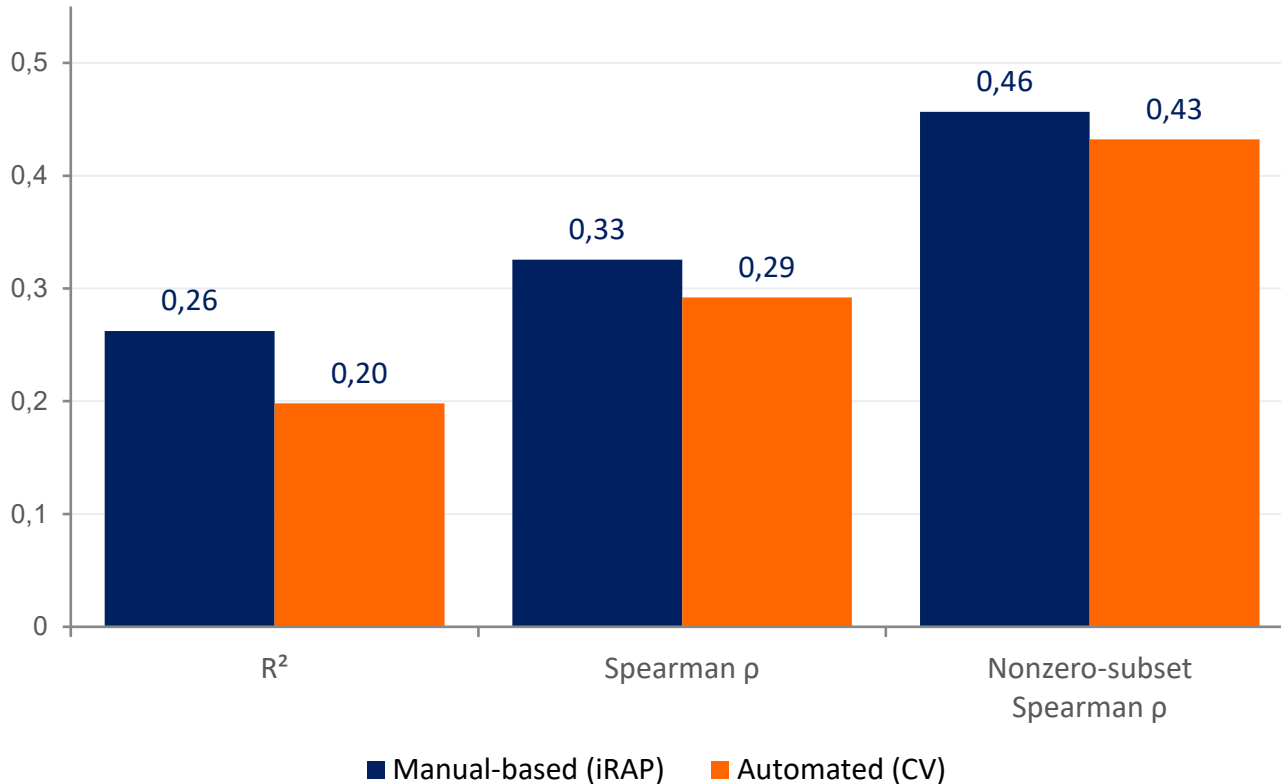


Modelling comparison approach

Model	XGBoost regression, Poisson count objective – number of HE per segment
Features	Manual: 23 iRAP attributes Automated: 21 CV / open-data proxies. Both keep mean speed & trip count.
Validation	Repeated stratified 5-fold CV, 10 repetitions → 50 train/test splits, identical folds & seed for both models
Metrics tracked	R^2 , MAE, RMSE, Spearman rho, zero-prediction ratio, and R^2 and Spearman rho restricted to the nonzero subset, SHAP values, SHAP top-5 important features



Results: model comparison



➤ MAE

- Manual: 0.173
- Automated: 0.174
- Nearly identical

➤ Zero-prediction ratio

- Manual: 97%
- Automated: 98%
- True: 92%

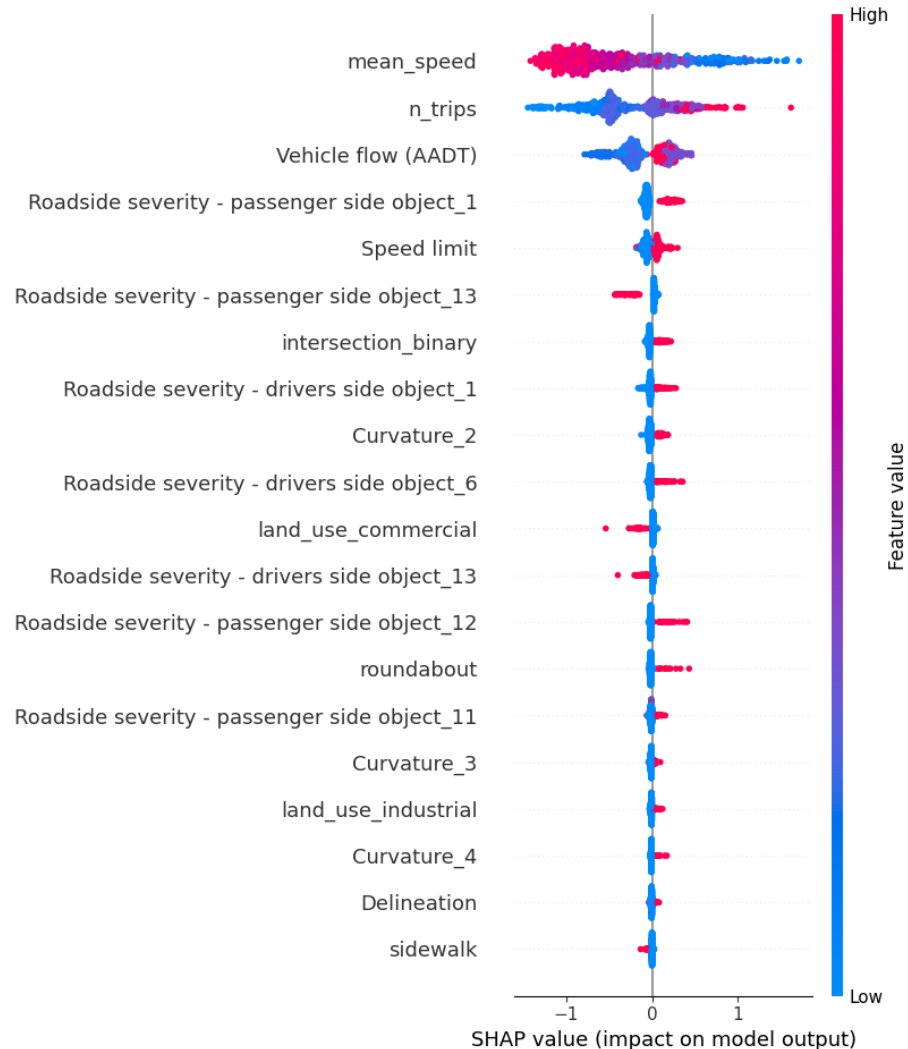
➤ Naïve mean-prediction baseline

- Manual: 0.264 MAE
- Automated: 0.271 MAE

- The automated pipeline **approximates, but does not exceed**, the manually coded on raw predictive power



Results: SHAP importance – manual model



➤ **Top predictors** present in 100% of top-5 ratings

➤ Mean speed

➤ Number of trips

➤ Vehicle flow

➤ **Infrastructure signal**

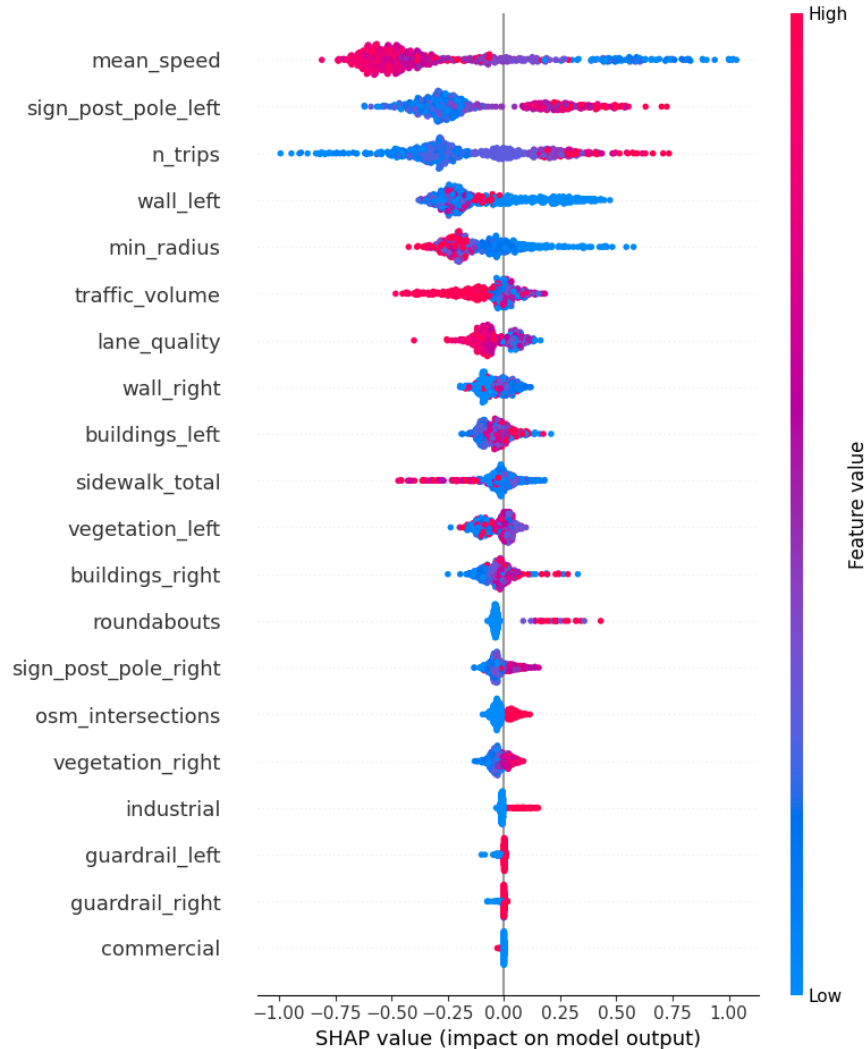
➤ Guardrail presence, passenger side (70%)

➤ Speed limit (56%)

➤ Building presence (40%)



Results: SHAP importance – automated model



➤ **Top predictors** present in 100% of top-5 ratings

➤ Mean speed

➤ Number of trips

➤ ~~Vehicle flow (18%)~~

➤ **Infrastructure signal**

➤ Sign / post / pole on the left (98%)

➤ Wall on the left (96%)

➤ Minimum radius (76%)



Results summary

- Similar predictive power, different explanatory story

	Manual-based (iRAP)	Automated (CV)
Top-5, both models	mean_speed, n_trips	mean_speed, n_trips
Distinctly influential	Guardrail (passenger side) Traffic volume (AADT) Speed limit	Sign / pole / light (left) Wall presence (left) Curvature (min. radius)
Encoding logic	Forced single category from 14 roadside classes (here condensed in 7 classes)	Independent, continuous output per object class
Non-significant, both models	Land use (commercial / industrial) and sidewalk presence	



Results discussion

Encoding structure, not proxy quality

- iRAP forces a single mutually-exclusive category per location while the CV pipeline records every object type independently and continuously
- Continuous outputs let the model capture co-occurrence effects that categorical, one-hot coding structurally does not
- Guardrail detection is highly accurate (97% mAP) as a proxy yet nearly irrelevant to the CV model's predictions, showing strong proxy validity $\neq$ high model importance

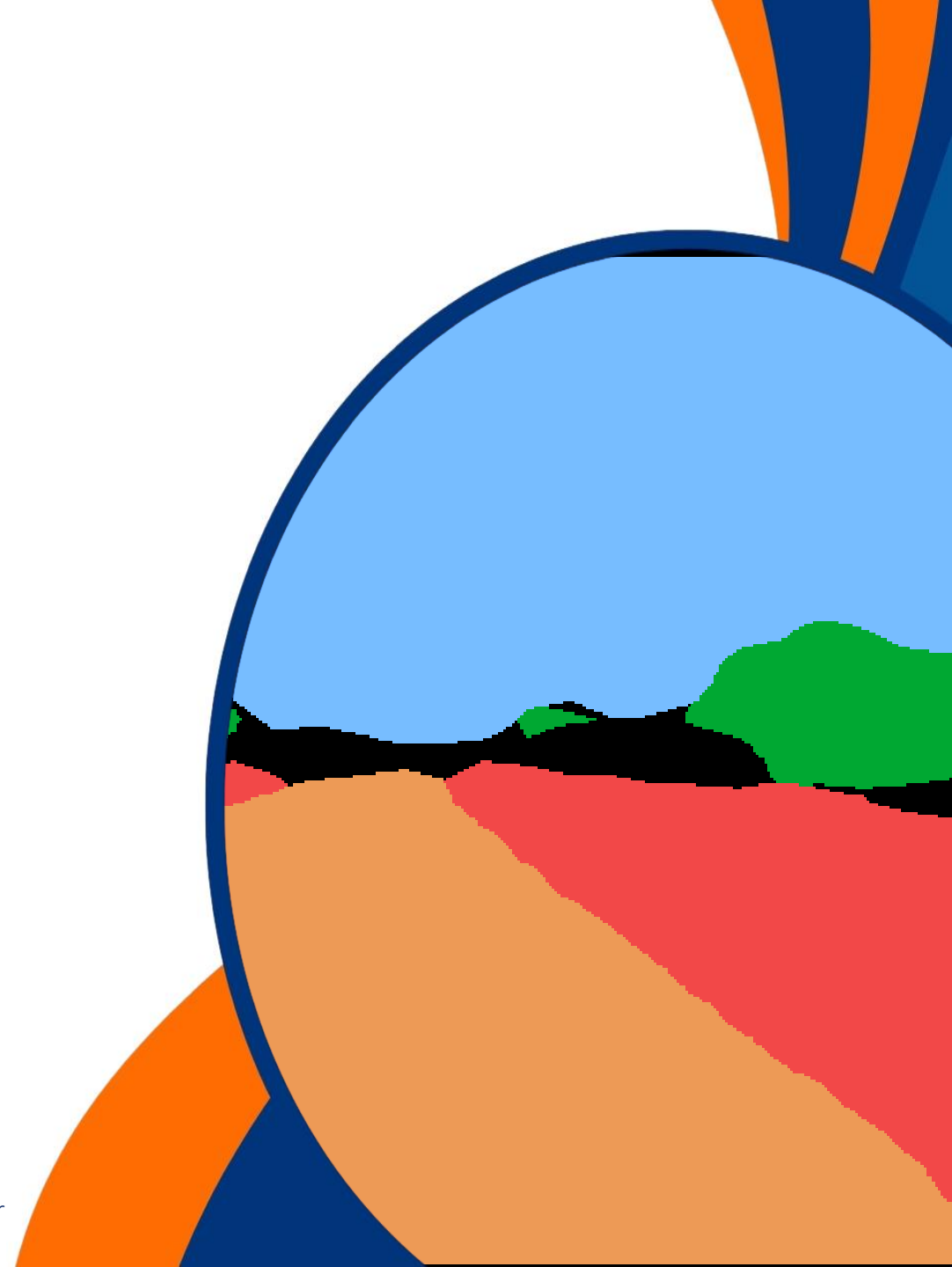
A left/right puzzle

- Manual model: guardrail & buildings matter far more on the passenger side than the driver's side
- Automated model: walls and signs/poles matter far more on the left than the right
- Telematics events were not split by travel direction, so this cannot yet be attributed to driver gaze – but it flags a promising, underexplored direction.



Conclusion

- Infrastructure inspection data can predict HE occurrence
 - Both models clearly beat the naïve baseline
- Telematics exposure (mean speed, trip count) is the dominant predictor in both models
 - Infrastructure plays a secondary role
- CV and geospatial proxies approximate manual coding in raw predictive power
 - 0.20 vs 0.26 R^2
- The two approaches had meaningful differences at the attribute level
 - Encoding logic reshapes which infrastructure features matter



Limitations and Future Work

➤ Limitations

- Small network and limited trip volume
- Results are not generalizable
- Driving direction not incorporated when aggregating telematics events
- Roadside object distance could not be reliably proxied from monocular imagery

➤ Future work

- Larger dataset with more trips per segment: move towards a full case study
- Additional proxies: intersection volume, object-distance estimation
- Incorporate driver gaze and directional analysis of roadside elements
- Apply data-balancing for the zero-inflated target



Comparison of Manual and Automated Infrastructure Label Extraction for Harsh Event Prediction

Julia Alves Porto

Ivory / NTUA PhD Candidate

Together with:

Apostolos Ziakopoulos, Daniel Lopez, John Avgeros and George Yannis



Questions?

Contact:

julia_porto@mail.ntua.gr

European Transport Conference

Data session 3, Video and Resilience

Porto, 9 September 2026