

Modelling Pedestrian Signal Violations Using Video-Based Behavioural and Traffic Data

George Yannīs¹, Stella Roussou¹, Apostolos Ziakopoulos¹

¹ National Technical University of Athens, Department of Transportation Planning and Engineering, Athens, Greece

Introduction

Pedestrian safety remains one of the most pressing challenges in urban transportation. A significant share of pedestrian crashes occur at intersections and are linked to non-compliant behaviour, crossing during the red signal, or outside the marked crossing area, raising collision risk.

Non-compliance is rarely a simple rule violation as it is shaped by waiting time, traffic volume, vehicle speed, and environmental conditions. Signal indication itself is not the only driver of the crossing decision, **perceived risk and the available gap in traffic are equally decisive**, pointing to non-compliance as tactical behaviour rather than random rule-breaking.

Traditional approaches, observational surveys and crash-data analysis, are limited, since crashes are rare events that reveal little about the interactions and near-misses preceding them. **Surrogate safety measures and behavioural analysis have therefore become central to road-safety research**, allowing risky interactions to be studied before a crash occurs.

Computer vision and video analytics now make it possible to **extract detailed trajectory and behavioural data directly from video**, capturing movement, speed, signal compliance, and conflict indicators such as Time-to-Collision (TTC). These data form the basis of the analysis presented here.

Objectives

- Examine the traffic and behavioural factors influencing pedestrian non-compliance at a signalized intersection using video-derived data.
- **Develop and compare statistical and machine-learning models, logistic regression and Random Forest, to predict illegal pedestrian crossings.**
- Support the development of proactive, video-based road-safety assessment tools.

Study Area

The present analysis **focuses on the Omonoia Square intersection**, one of the busiest pedestrian-vehicle interaction points in central Athens. The intersection operates under a fixed-time signal plan with a 90-second cycle, comprising **15 seconds of pedestrian green**, 7 seconds of intergreen (yellow), and 68 seconds of pedestrian red, meaning pedestrians face a red signal for the large majority of the cycle, a condition that plausibly encourages non-compliant crossing when gaps in traffic appear.

Methodology

A computer-vision pipeline was created and processed the video footage in four stages:

- **Object detection**, YOLOv8 detects pedestrians and vehicles in each frame (pre-trained COCO weights, no fine-tuning).
- **Feature extraction**, ResNet-50 generates a feature vector for every detected object.
- **Tracking**, a Kalman filter predicts and updates object positions, the Hungarian algorithm associates detections across frames.
- **Signal-phase & TTC**, trajectories are matched to the signal phase, Time-to-Collision is computed.
- **A crossing is logged as illegal** if the pedestrian crosses while the light is red, legal if green. Vehicle speed is derived from consecutive bounding-box positions via a homographic pixel-to-metre transform.

Two complementary statistical approaches were then applied:

- **point-biserial correlation** to test the association between continuous traffic variables and the binary legal/illegal outcome, and
- **binary logistic regression** to estimate the probability of a violation as a function of traffic and interaction conditions.

A **Random Forest classifier** (100 trees, maximum depth 10, minimum samples per split = 2, class weights balanced to address the imbalance between legal and illegal crossings, hyperparameters tuned via grid search) was developed in parallel, both to predict violations and to rank the relative importance of each predictor.



Figure 1: Omonoia Square in Athens

Results

The analysis confirms that **pedestrian non-compliance at Omonia Square is not random, but systematically linked to traffic conditions**. Point-biserial correlation revealed significant associations between violations and vehicle speed, TTC, signal status, and vehicle presence, consistent with earlier findings that crossing decisions depend on gap acceptance and perceived risk rather than the signal alone.

The logistic regression results show that the **probability of an illegal crossing increases when vehicle speed is lower**, TTC is higher, and no vehicle is actively approaching, in other words, pedestrians cross illegally when they judge there is enough time and space to do so safely. The 1st logistic model achieved a Pseudo R2 of 0.643, where the 2nd was 0.907.

Pedestrian presence in the crossing area				
Variable	Coef.	Std.Err.	z	P> z
const	-2.518	0.069	-36.375	0.000
MinTTC_wins_log	0.961	0.037	26.05	0.003
vy	0.384	0.024	15.681	0.000
GroundPlaneCentroid_1_veh	0.077	0.036	2.1	0.000
magnitude_veh	-0.238	0.028	-8.42	0.000
VehicleStatus	-0.292	0.129	-2.272	0.000
ResolvedTrafficLightStatus	4.581	0.160	28.568	0.023

Table 1: Results of 1st Logistic Regression Model

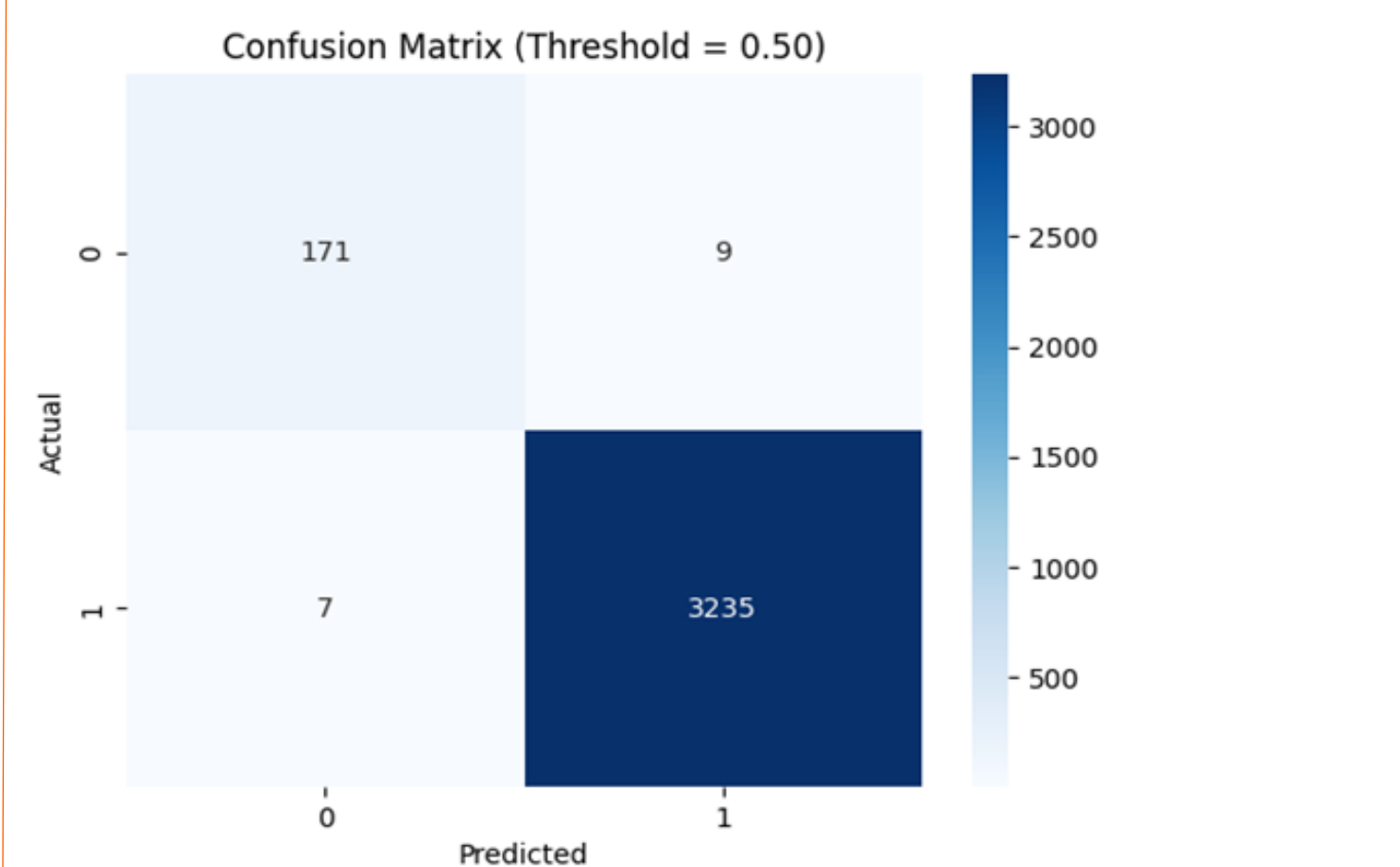
Legal or illegal pedestrian crossing				
Variable	Coef.	Std.Err.	z	P> z
const	5.726	0.145	39.606	0.000
MinTTC_wins_log	-0.213	0.072	-2.965	0.003
Confidence	0.842	0.09	9.347	0.000
vy	-0.379	0.044	-8.59	0.000
GroundPlaneCentroid_2_veh	-0.445	0.048	-9.214	0.000
magnitude_veh	0.798	0.092	8.713	0.000
VehicleStatus	-0.292	0.129	-2.272	0.023
ResolvedTrafficLightStatus	4.581	0.16	28.568	0.000
IsInCrossingArea	-3.039	0.091	-33.383	0.000

Table 2: Results of 2nd Logistic Regression Model

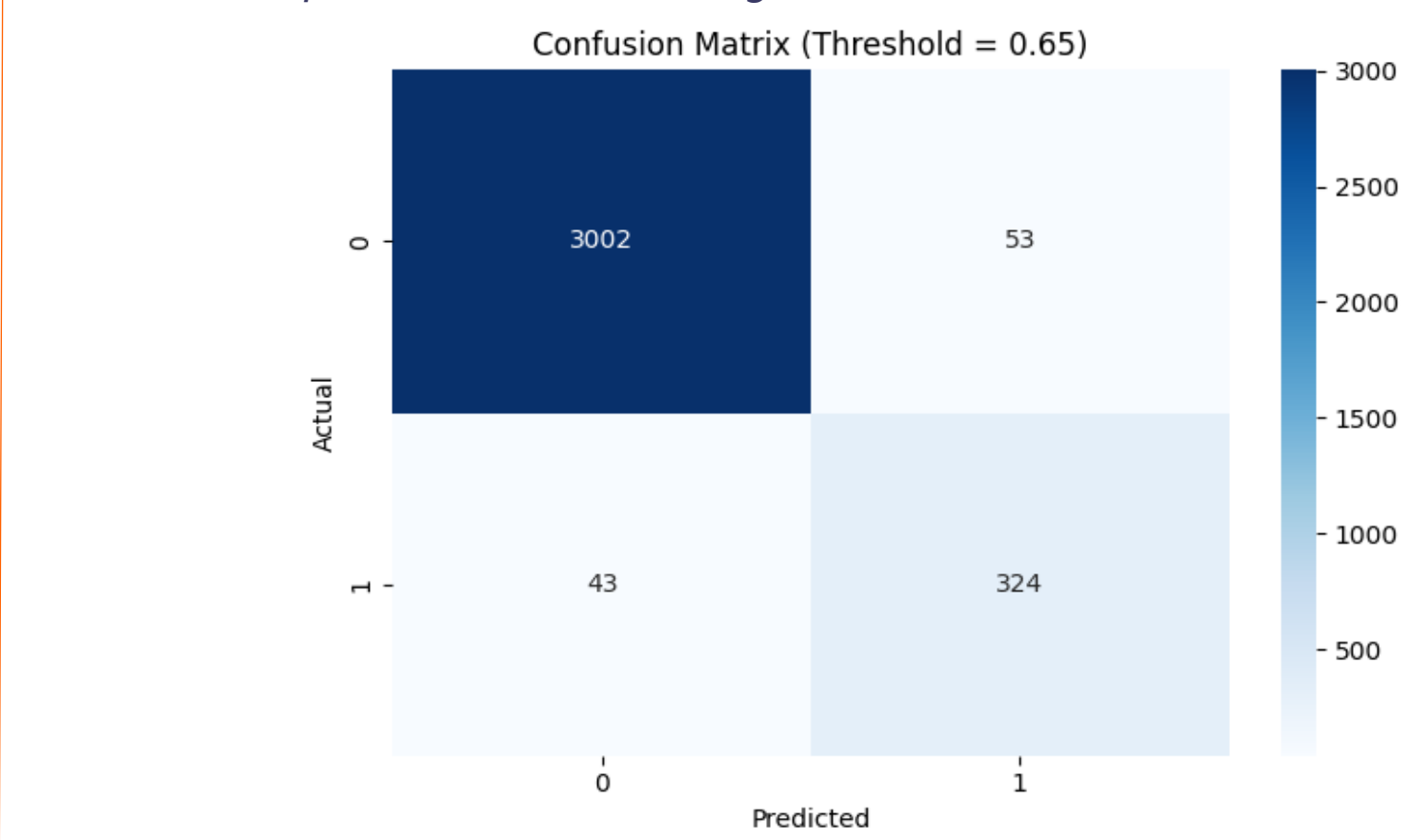
The Random Forest model achieved high classification accuracy in predicting violations. **Variable-importance analysis confirmed signal status, TTC, vehicle speed, and vehicle presence as the most influential predictors**, with whether the pedestrian was already inside the crossing area emerging as the single strongest predictor of the legal/illegal outcome.

Pedestrian presence in the crossing area		Legal or illegal pedestrian crossing	
GroundPlaneCentroid_2	0.234	IsInCrossingArea	0.579
PedestrianStatus	0.181	ResolvedTrafficLightStatus	0.136
TrafficLightStatus	0.106	MinTTC_wins_log	0.034
MinTTC_wins_log	-0.08	magnitude_veh	0.028
Confidence_veh	0.017	VehicleStatus	0.0002
IsInCrossingArea_veh	0.01		
VehicleStatus	0.01		

Table 3: Results of Random Forest Feature Importance



(a) Pedestrian presence in the crossing area



(b) *Legal or illegal pedestrian crossing*

Conclusions

- **Pedestrian non-compliance is significantly influenced** by vehicle speed, Time-to-Collision, signal status, and vehicle presence, **not random behaviour**.
- Pedestrians make tactical, risk-based crossing decisions, exploiting **perceived safe gaps rather** than strictly obeying the signal.
- Logistic regression and Random Forest both predict illegal crossings with high accuracy, supporting video-derived data for proactive safety analysis.
- **Waiting time before crossing was not captured by the current pipeline**. Future work should add it, along with more intersections and active countermeasures.

Acknowledgments

The present research was carried out within the research project “PHOEBE - Predictive Approaches for Safer Urban Environment”, which has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101076963



Contact Information:

George Yannis, Professor, NTUA

Department of Transportation Planning and Engineering

Email: geyannis@central.ntua.gr

Website: <https://linktr.ee/georgeyannis>

