Predictive Modeling of Pedestrian Violations Using Ensemble Learning: A Case Study at an Urban Intersection in Athens, Greece

Stella Roussou*[1], Apostolos Ziakopoulos[2], George Yannis[3]

[1] Ph.D. Candidate, Research Associate, Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece

[2] Ph.D., Senior Associate, Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece

[3] Professor, Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece

SHORT SUMMARY

Machine learning and advanced computer vision have revolutionized the analysis of urban traffic safety. The paper discusses the predictive performance of Random Forest and XGBoost algorithms in modeling pedestrian-vehicle interactions at the Panepistimiou and Vasilissis Sofias junction in Athens, Greece, with a focus on illegal crossings. Video footage from one peak hour was processed using YOLOv8 for object detection, ResNet-50 for feature extraction, and Kalman filtering for trajectory refinement. Both have been trained using features derived from pedestrian interactions in different traffic light phases and evaluated using precision, recall, and F1-score. XGBoost outperformed Random Forest, achieving superior precision and accuracy, while Random Forest demonstrated better computational efficiency. This work underlines some of the key trade-offs and strengths of those models in an ensemble learning perspective on real-time traffic safety applications, giving actionable insights that might improve urban traffic management to support global road safety goals.

Keywords: Athens Case Study, Random Forest, XGBoost, YOLOv8, Urban Traffic Safety.

1. INTRODUCTION

The safety of urban traffic is a pressing issue all over the world, while road traffic crashes contribute to millions of deaths and injuries each year. It is an outcome of various factors interacting together, such as human behavior, infrastructure design, environmental conditions, and socioeconomic disparities. Finding solutions to these challenges is essential for achieving the ambitious global goal of reducing traffic fatalities by half by 2030, adopted under the leadership of the European Union and the World Health Organization (European Commission, 2019; WHO, 2022). In this endeavor, innovative technologies, especially in machine learning and computer vision, are increasingly vital in providing data-driven real-time traffic monitoring and safety analysis solutions.

Densely populated urban environments, where pedestrians, cyclists, and motorized vehicles coexist, present unique challenges for traffic safety management. A case study at the intersection of Panepistimiou and Vasilissis Sofias streets in Athens, Greece has been examined in this study.

High traffic density, heterogeneity in road user behavior, and a lack of real-time monitoring due to the key arterial junction with many governmental, commercial, and cultural landmarks is some of the key elements of this case study. This suggests that the lack of integrated traffic cameras and thorough data creates the need for innovative analysis to be implemented.

Computer vision approaches like object detection and tracking have become integral to traffic analysis. Models such as YOLO have shown very impressive speed and accuracy, making them capable of real-time vehicle, pedestrian, and road user detection. YOLOv8 is the most recent in this series (Ultralytics, 2024), combined with ResNet-50 for feature extraction and Kalman filtering for trajectory prediction, will provide the most potent tool in the analytical framework when it comes to complex urban traffic dynamics. These technologies are integral to the EU PHOEBE project ("Predictive Approaches for Safer Urban Environments") (Phoebe, 2024), which seeks to improve road safety by developing novel predictive modeling and simulation approaches.

While object detection provides foundational data, the prediction of such dynamic behaviors as compliance of pedestrians with traffic lights requires advanced machine learning algorithms. In this respect, ensemble learning models such as Random Forest and XGBoost are particularly appropriate, given their capability to deal with high-dimensional data and capture complex relationships. This paper intends to do an evaluation and comparison of the predictive performance of Random Forest and XGBoost models in forecasting pedestrian violations at the Panepistimiou and Vasilissis Sofias intersection. Given the power of YOLOv8 for object detection and ResNet-50 for feature extraction, the analysis will focus on the pedestrian behavior at three different traffic light phases: green, red, and intergreen. The performance of the models is evaluated with precision, recall, and F1-score metrics in order to assess their applicability to real-world scenarios.

The structure of the paper is as follows: The Methodology section demonstrates the data collection process, the preprocessing techniques applied, and the ensemble learning models used. In the Results and Discussion section, the different models are compared with their strengths, trade-offs, and implications for real-time traffic safety systems. Finally, the Conclusion reviews the main findings, discusses their practical implications in view of urban traffic management, and addresses the directions for future research. This research also contributes to the higher aim of improving the safety of urban traffic in advanced machine learning applications, especially where there is low monitoring infrastructure.

2. METHODOLOGY

Object Detection and Feature Extraction

The first step in processing is the detection of objects within the video frames. The YOLOv8 model is employed for this purpose, and it formulates object detection as a single regression problem. For an input image extracted as a frame from the video, YOLOv8 (Ultralytics, 2024) outputs a set of bounding boxes B=$(B1, B2, \dots, B_n)$, where each bounding box B_i is defined by:

$$B_i = (x_i, y_i, w_i, h_{i,} c_i) \qquad \text{(Eq. 1)}$$

Where (x_i, y_i) are the coordinates of the center of the box, $w_i, and\ h_i$ are the width and height of the bounding box, and c_i is the class probability score indicating whether the object in the box is a pedestrian or vehicle.

Once the objects are detected, ResNet-50 is employed for feature extraction. The feature vector F_i corresponding to each detected object B_i is extracted from the output of ResNet-50, represented as:

$$F_i = ResNet - 50(B_i) \quad \text{(Eq. 2)}$$

These then form the basis for tracking and re-identification. For pedestrian tracking, the system relies on the application of Kalman filters (Kalman, 1960). A Kalman filter has two main steps: Prediction and Update. The Kalman Gain balances the uncertainties in the prediction and the measurement. In addition, it also makes use of the Hungarian algorithm for data association. This algorithm solves optimally the association between the predicted object locations and new detections to get a minimal cost function regarding object-to-track association.

Once objects are detected and tracked, the system analyzes pedestrian behavior in relation to traffic signal phases. The pedestrian behavior is classified as either legal or illegal depending on whether they cross the pedestrian line during a red pedestrian light. The behavior status of tracked pedestrian is $S_{ped}(t)$, where:

$$S_{ped}(t) = \begin{cases} illegal & \text{if } S_{light}(t) = red \\ legal & \text{if } S_{light}(t) = green \\ unknown & \text{if } S_{light}(t) = intergreen \end{cases}; \quad (3)$$

Each illegal crossing event is logged, and the number of violations is counted. It then generates annotated outputs in video and data file annotations after the detection of illegal crossings by pedestrians. This includes the video output with marked illegal crossings, along with time stamps, bounding boxes, and phases of the traffic signal corresponding to every detected event. The data files include illegal crossing counts along with violation times and locations.

Random Forest

Random Forest is an ensemble learning algorithm that builds multiple decision trees during training and combines their outputs to improve predictive performance. Its robustness to noise and overfitting makes it particularly suited for datasets with imbalanced classes. Key features of the RF implementation include:

- Bagging (Bootstrap Aggregating): By training each tree on a different random subset of data, RF reduces variance and enhances generalization.
- Feature Importance Ranking: RF inherently ranks features by importance, aiding interpretability and highlighting key predictors, such as pedestrian speed and traffic light phases.
- Optimization: Hyperparameters, including the number of trees, maximum tree depth, and minimum samples per split, were tuned using grid search. Class weights were adjusted to counter class imbalance, ensuring that minority classes (e.g., illegal crossings) were not overlooked.

XGBoost

XGBoost, a gradient boosting framework, excels in predictive accuracy through iterative optimization and regularization. Its ability to handle missing data, parallelized computations, and focus on minimizing errors at each iteration makes it ideal for complex, high-dimensional datasets. Key characteristics of the XGBoost model include:

- Gradient Boosting: Trees are sequentially built, with each subsequent tree focusing on correcting errors made by the previous ones.
- Regularization: To prevent overfitting, L1 (Lasso) and L2 (Ridge) regularization techniques were applied.
- Weighting Imbalanced Classes: The weight parameter was tuned to balance the model's minority class of illegal crossings.

- Hyperparameter Tuning: Parameters, including the number of estimators, maximum depth, learning rate, and subsampling ratios, were optimized through grid search to achieve the best balance between accuracy and computational efficiency.

3. RESULTS AND DISCUSSION

The dataset for this study was collected by strategically setting up smartphone cameras at the key point of Panepistimiou Street and Vasilissis Sofia's junction, a central and heavily trafficked location in Athens, Greece. This location of this urban arterial was chosen as part of the EU PHOEBE project to provide a representative sample of urban traffic scenarios, including various types of vehicles, pedestrians, and cyclists (Phoebe, 2024). The absence of integrated traffic cameras and comprehensive street view data in the area necessitated manual video capture. This was attained by using field researchers who shot video footage using commercially available smartphones and tripods and compiling them in external databases at the end of each workday data collection.

The entire data collection period for a number of key point locations had a duration of 4 days, spread across 2 weeks to record two typical days per week (Tuesdays and Thursdays) for each location. The intent was to collect video footage encompassing a peak traffic hour of each workday when the city center is at its busiest (thus presenting the highest challenge to the algorithms), specifically on a weekday from 9 am till 10 am and a non-peak traffic hour later in the same day at 8 pm till 9 pm. Ultimately, more than 8 hours of video were collected for each location. Indicative images depicting one of the recorded locations can be found in Figure 1 below, taken by the authors.

Figure 1: Panepistimiou and Vassilissis Sofias Urban Junction in Athens

The video footage used for this study was generated with cameras from smartphones attached to tripods at key pedestrian crossing points in Athens, thus providing a flexible and cost-effective solution for data collection. The collection was made so that representative samples of pedestrian behavior could be gathered for different traffic conditions and light phases. It also avoids the need for fixed infrastructure to make it applicable within any urban environment with limited monitoring systems.

For the research study, an enhanced video analysis algorithm for tracking pedestrians while crossing the zebra section illegally. Figure 2 demonstrates the critical component of the algorithm, precisely the one of the traffic light detection. Recognition of the intersection traffic light color status-that is, red, green, or intergreen- represented valuable recognition relevant to the modeling of illegal crossings made by pedestrians. This was achieved by the use of pre-trained models and further incorporating region-based classification of the frames of the image.

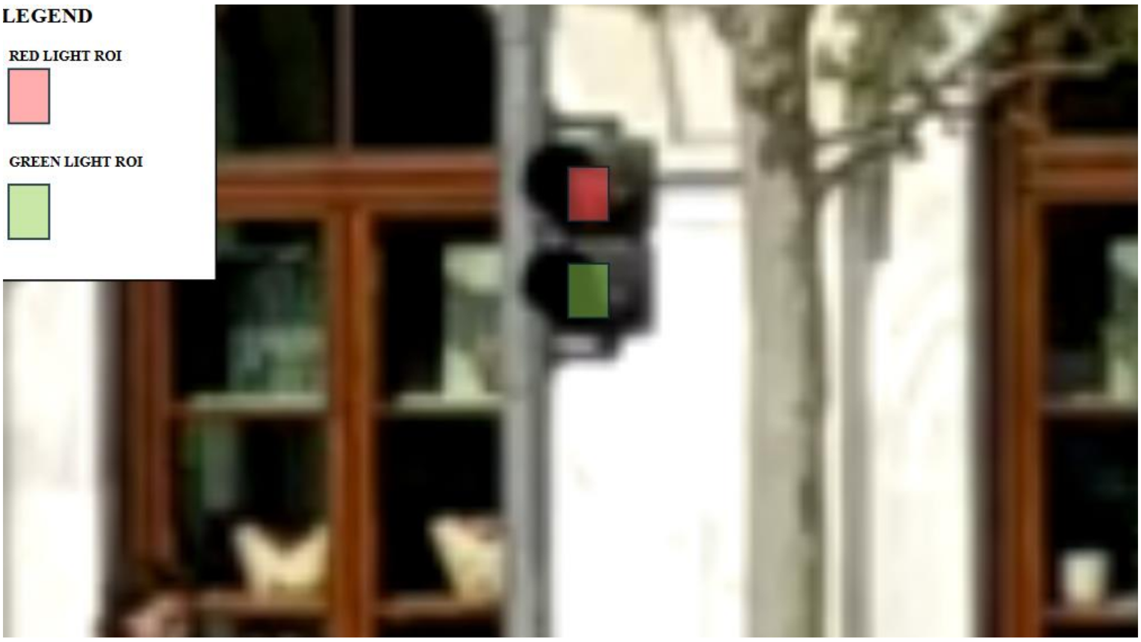

Figure 2: Traffic Light Detection

The Region of Interest (ROI) for this study was set up to be centered exactly on the marked pedestrian crossing zones, as presented in Figure 3. By defining the pedestrian zones as ROIs, the area that was under examination was reduced to only the zone of interest where pedestrians may flow, thus helping to filter irrelevant information and achieve higher accuracy regarding the detection of violations. This approach detected an illegal pedestrian crossing when a pedestrian entered an already defined ROI during red or intergreen phases. The algorithm tracked the position and movement of pedestrians across the intersection and compared these movements with the current status of the traffic light to determine whether a violation had occurred.

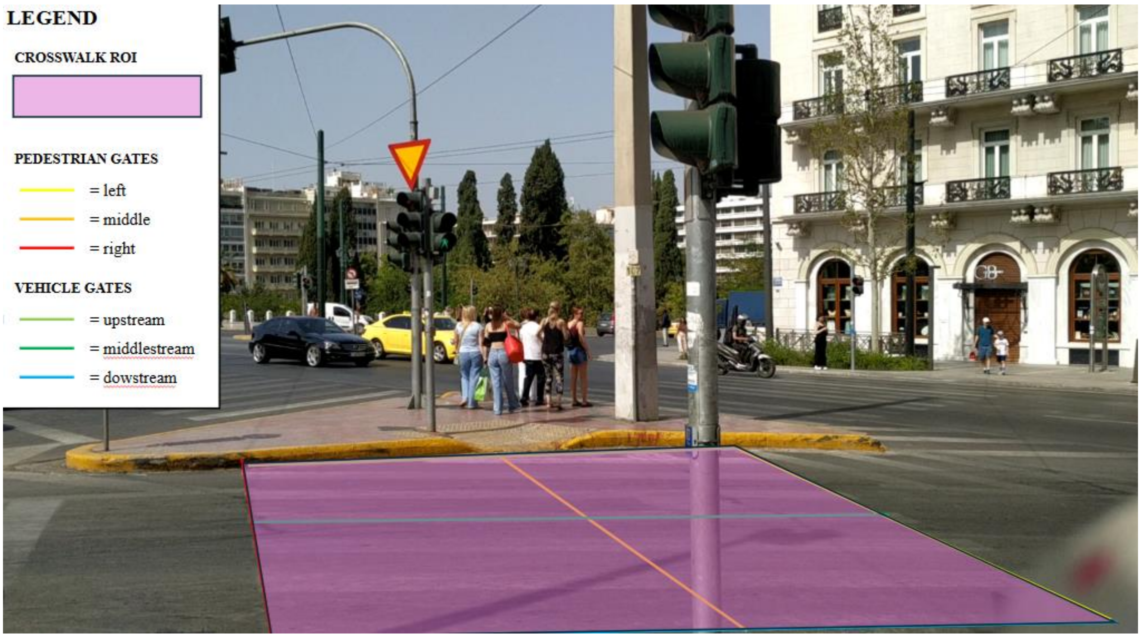

Figure 3: Region of Interest (ROI)

By focusing on the ROI and employing traffic light detection, the algorithm provided exact and actionable insights into pedestrian behavior, such as illegal crossings during different phases of traffic lights. The processed data was used for training and performance evaluation of the predictive models, thus providing a comprehensive dataset to assess pedestrian behavior and violations across urban intersections.

There are clear differences in performance, strengths, and limitations between the Random Forest and XGBoost performance on the test dataset. The Random Forest model outperformed overall with an accuracy of 88%, which includes precision of 96% for the majority class and 55% for the minority class, as demonstrated in Table 1. The recall in the minority class was 68%, leading to an F1-score of 61%, depicting a balanced ability to identify violations, keeping a relatively low rate of false positives. These results seem to depict Random Forest as a robust and interpretable model but with limitations in modeling rare events, such as pedestrian violations, even if it can generalize very well for the majority class.

Table 1: Classification Report for Random Forest Model

Metric	Class 0 (No Violations)	Class 1 (Violations)	Macro Average	Weighted Average
Precision	0.96	0.55	0.76	0.92
Recall	0.91	0.68	0.79	0.88
F1-Score	0.93	0.61	0.77	0.89
Support	139,386	11,386	-	150,772

In contrast, the XGBoost model displayed superior performance across all evaluation metrics, achieving an overall accuracy of 93%, while for the minority class alone, it posted a precision of 72%, a recall of 88%, and an F1-score of 79%. Its iterative mechanism of gradient boosting and hyperparameter optimization gave an edge for this strong performance; it adaptively self-corrects misclassifications to efficiently generalize to unseen data. More so, XGBoost regularization techniques include L2 penalty to reduce overfitting, hence giving more reliability in the model while making violation predictions on even imbalanced datasets.

Table 2: Classification Report for XGBoost Model

Metric	Class 0 (No Violations)	Class 1 (Violations)	Macro Average	Weighted Average
Precision	0.96	0.72	0.84	0.93
Recall	0.95	0.88	0.91	0.93
F1-Score	0.96	0.79	0.87	0.93
Support	139,386	11,386	-	150,772

Both models were optimized using their respective performance levels. As for the Random Forest model, several critical hyperparameter tunings included the number of trees optimized at 200, maximum tree depth at 15, and minimum samples per split-10. Apart from this, adjusting class weights was implemented to treat the class imbalance so that the minority class receives adequate representation during training.

Extensive hyperparameter tuning has been done for the XGBoost model using grid search. The most important hyperparameters are the number of estimators, 100; maximum tree depth, 10; learning rate, optimized at 0.1; subsampling ratio, 0.8. Besides, the scale_pos_weight parameter is set to balance the class distribution within the dataset, to make it sensitive to rare events such as illegal crossings. These choices of hyperparameters were very instrumental in getting the best performing model, especially for the minority class.

The comparative analysis brought into focus huge trade-offs between the models. Whereas Random Forest yielded strong overall performance with high precision for the majority class, limitations in recall and F1-score for the minority class reduced its effectiveness for real-time applications requiring high sensitivity. Conversely, XGBoost showed more balance across all metrics and is better positioned for scenarios where accurate identification of violations is crucial. With its better recall, the XGBoost method would be suitable for real-time monitoring systems that ought to catch even rare events of illegal crossings of pedestrians while minimizing false negatives.

4. CONCLUSIONS

The study demonstrated the performance of the machine learning models of Random Forest and XGBoost, for predicting pedestrian behavior in terms of illegal crossings in an urban intersection in Athens, Greece. The models have been trained on a dataset of pedestrian-vehicle interactions across various traffic light phases by utilizing computer vision techniques, including state-of-the-art object detection using YOLOv8 and feature extraction through ResNet-50.

Eventually, the XGBoost model proved to be the most appropriate model. This model presented greater precision, recall, and total accuracy, especially with respect to finding the events of illegal crossing. It works efficiently in managing class imbalances, minimizes error with computational efficiency, and therefore befits real-time traffic safety applications. Random Forest is less precise but demonstrates robust performance and interpretability; therefore, it might act as a very useful tool when computational resources are limited or there is a need to understand feature importance explicitly.

These findings have important implications for urban traffic safety in a densely populated area like Athens, where pedestrian behavior and traffic conditions are highly dynamic. The XGBoost algorithm is able to provide actionable insights into pedestrian violations, supporting the development of advanced traffic monitoring systems that could lead to better signal compliance and improved safety measures. On the other hand, Random Forest might be helpful in exploratory analyses and identifying key predictors of pedestrian behavior.

Future research should move toward the improvement of this feature set by adding contextual variables such as weather condition, road structure characteristics, and demographics of pedestrians for further improved predictions. Similarly, fine-tuning decision thresholds could also further optimize the precision-recall tradeoff depending on the application-a system designed either for enforcement applications or public safety intervention. More such model validations using larger and more varied datasets, with subsequent field implementation of these systems, would shed light on operational reliability, scalability, and how these systems perform in the real world. These are steps toward the overarching aim of traffic risks reduction in contributing to the meeting of global safety targets.

ACKNOWLEDGEMENTS

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.

REFERENCES

European Commission. EU Road Safety Policy Framework 2021–2030 - Next Steps towards “Vision Zero”, 2019.

Kalman, R.E. A new approach to linear filtering and prediction problems. J. Basic Eng. 1960, 82, 35–45.

PHOEBE Project. PHOEBE Library. https://phoebe-project.eu/phoebe-library/. Accessed July 1, 2024.

Ultralytics. (2024). *Ultralytics YOLOv8* [GitHub repository]. GitHub. https://github.com/ultralytics/ultralytics (Accessed on September 27, 2024).

World Health Organization, Global status report on road safety 2023. WHO, 2023.Güney, E., & Bayılmış, C. (2022). An implementation of traffic signs and road objects detection using faster R-CNN. Sakarya University Journal of Computer and Information Sciences, 5(2), 216-224.