Streetscape quality assessment related to pedestrian crashes using supervised learning

Project Type: Functional Analysis-narrative

Student: Huan Zhou, Instructor: Chester Harvey

1. Motivation

Streetscape quality refers to the visual condition of street environment, mainly related to infrastructures (e.g., roads, sidewalks, and crosswalks), buildings, and landscaping of streets (Ye et al., 2019). Poor streetscape design and quality (e.g., open enclosure, no sidewalks, missing crosswalks, and no traffic signals) may increase the risk of pedestrian crashes by creating hazardous conditions (Harvey & Aultman-Hall, 2015, Nguyen et al., 2024). For example, Baltimore City accounts for 31% of total pedestrian crashes in Maryland (Figure 1) (Zero deaths Maryland, 2023). It has poor streetscape quality, like faded markings, pavement patch, absent crosswalks, and limited greenery (Figure 2), making it a crash hotspot. This study tries to assess streetscape quality in relation to pedestrian crashes using supervised learning at a large scale. By spotting poor streetscape qualities and helping urban planners improve infrastructure, it is highly possible to decrease the risk of pedestrian crashes and make streets safer for residents.

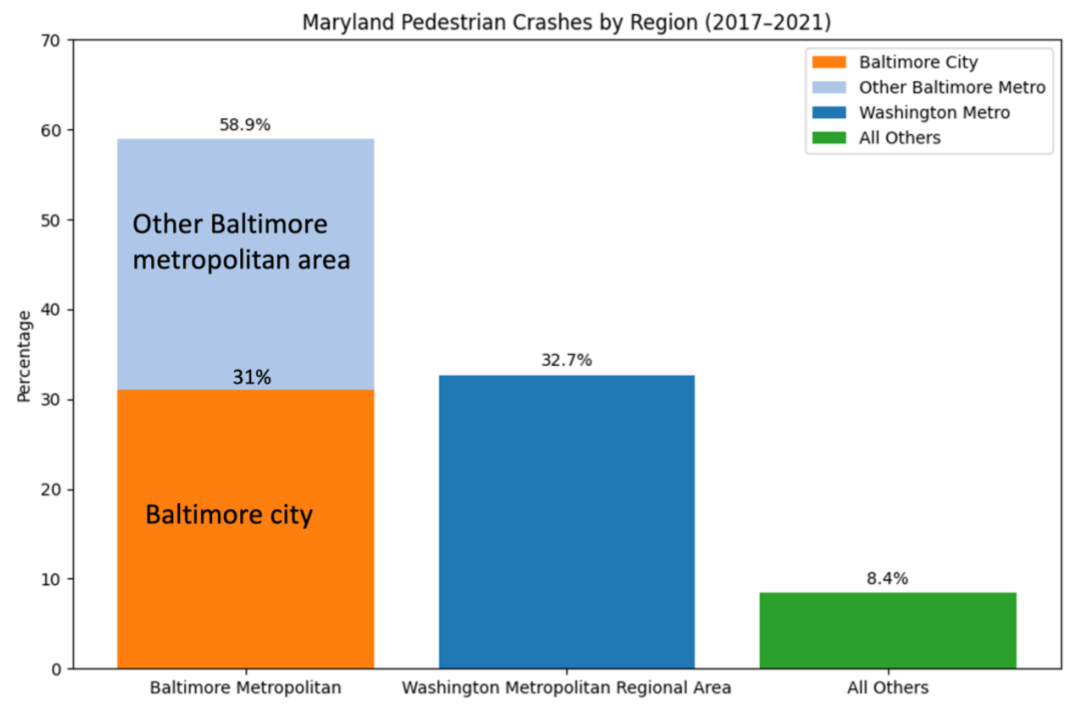

Figure 1. Maryland pedestrian crashes by region (2017-2021), Source: (Zero deaths Maryland, 2023)

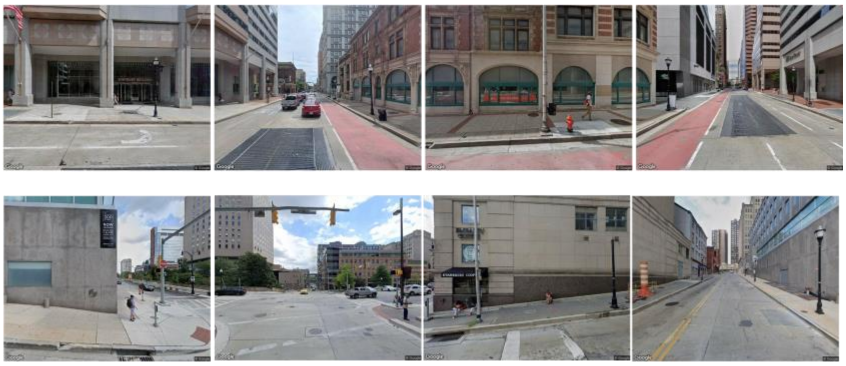

Figure 2. Street view image samples of Baltimore City

2. Central question

The study aims to develop a supervised learning model to assess streetscape quality related to pedestrian crashes at scale. It contains three central questions: i) What indicators can be used to represent streetscape quality linked to pedestrian crashes? ii) How to assess the streetscape quality on a large scale? iii) What are the implications, theoretical strengths and weaknesses of the study?

3. Approach

The methods include four key steps: downloading Street view images (SVI), labeling SVI, training the multi-label classification model, and evaluating (Figure 3). First, Baltimore City's SVI are download from Google Maps Platform using Street View Static API. Second, the study labels images manually based on strict rubric, and employs ChatGPT to generate pseudo-labels. Third, Vertex AI's Auto Machine Learning is used to train the model. Finally, the model's performance is evaluated on a test set of labeled images, based on precision and recall value.

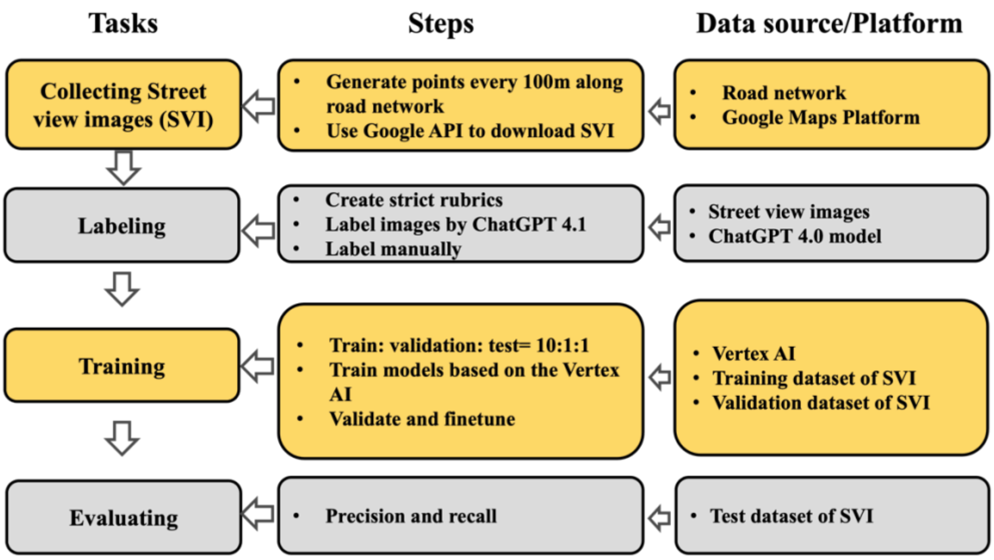

Figure 3. The method for streetscape quality assessment using supervised learning

3.1 Download Street view images

The study creates sampling points every 100 meters of road networks. At each generated point, street view images are captured in four different directions (90°, 180°, 270°, and 360°) from Google Maps Platform using API. The field of view refers to the extent of the scene captured in the image, which is 90 degrees. The pitch is 0°, which means the camera is level with the ground.

3.2 Label street view images

The project identifies seven indicators that are hypothesized to be associated with pedestrian crashes (Appendix Table 1). The indicators focus on the physical presence (e.g., streetlights) and conditions (e.g., poor, fair, good for sidewalks) of elements. Then, the study labels 100 images manually based on the rubrics (Appendix Tabel 2), and employs gpt-4.1 model to generate pseudo-labels for 1,100 images.

3.3 Train the multi-label classification model

First, the study prepares train, validation, and test dataset. The training and validation datasets contain 1,100 street view images in a 9:1 ratio, labeled by the GPT-4.1 model. The test dataset contains 100 manually labeled street view images. Then the study uses Vertex AI's Auto Machine Learning Model to train a multi-label image classification model. The input and output of the model are shown in Figure 4.

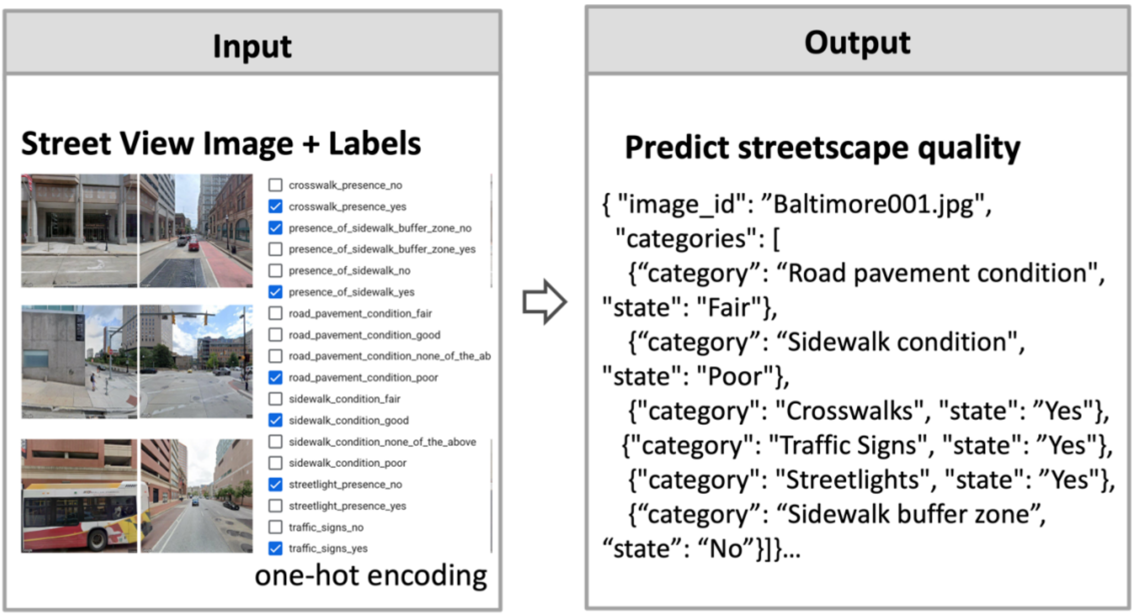

Figure 4. The input and output of the model

3.4 Evaluate the model performance

The study evaluates the model performance using AuPRC (average precision), confidence threshold, recall, and precision. The detailed explanation of evaluation metrics is shown in Appendix Table 3.

4. Results

The study area is a census tract in Baltimore city, which has the highest pedestrian crash rate (crash number/population) in Maryland. Approximately 1,300 images from this census tract are downloaded for this study. 1,100 images are labeled by ChatGPT 4.1 for training, and 100 images are manually labeled for testing. This study used Vertex AI's Automatic Machine Learning to train a multi-label image classification model.

4.1 Overall performance

Table 1 shows that the model achieves an average precision of 0.687 (i.e., the area under the precision-recall curve in Figure 5), which is acceptable. At the 0.5 confidence threshold, the model has a precision of 68.5%, meaning that 68.5% of its predictions for a given class are correct. It also reaches a recall of 67.4%, indicating that it identifies 67.4% of all actual positives of that class.

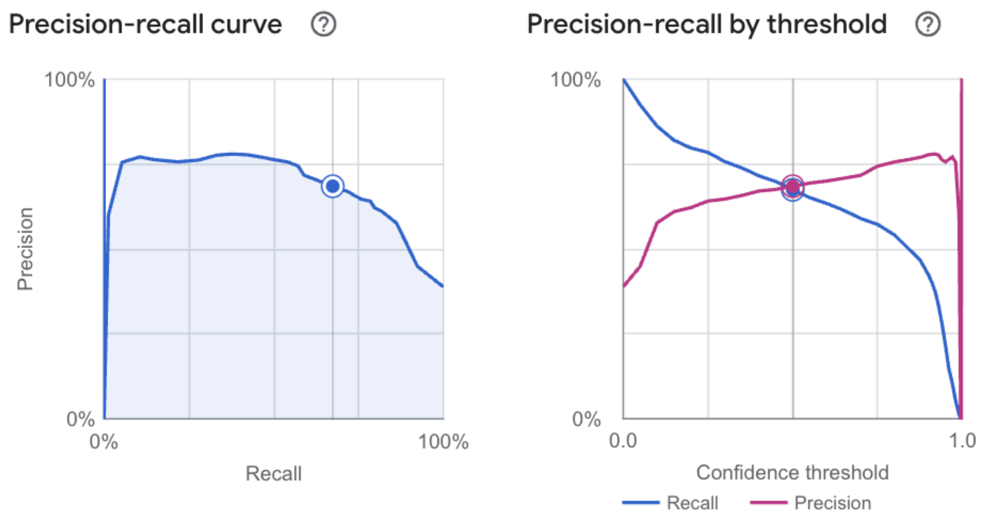

Figure 5. Precision-recall curve for all the labels (Source: Vertex AI platform)

Table 1. Evaluation results of overall performance

Evaluation metrics	Value
Average precision	0. 687
Precision	68.5%
Recall	67.4%
Total images	1,197
Training images	988
Validation images	109
Test images	100

4.2 Each label's performance

As shown in Table 2, performance varies across individual labels. The model performs very well in detecting the presence of physical street features, such as sidewalks, crosswalks, and traffic signs, with precision scores exceeding 0.75. For instance, the model identifies the presence or absence of sidewalks with perfect precision of 1.0 (Figure 6).

However, the model fails to assess subjective condition, such as road or sidewalk condition, with precision scores less than 0.50. For example, the model struggles to identify poor road pavement condition, achieving an average precision of only 0.346, with a confidence threshold below 0.4 (Figure 7).

Table 2. Evaluation results of each label's performance

Labels	Precision	Labels	Precision
presence_of_sidewalk_no	1	sidewalk_condition_good	0.662
presence_of_sidewalk_yes	1	streetlight_presence_no	0.638
crosswalk_presence_no	0.919	presence_of_sidewalk_buffer_zone_no	0.548
crosswalk_presence_yes	0.796	road_pavement_condition_good	0.438
presence_of_sidewalk_buffer_zone_yes	0.787	road_pavement_condition_fair	0.406
traffic_signs_yes	0.76	sidewalk_condition_poor	0.376
traffic_signs_no	0.759	road_pavement_condition_poor	0.346
streetlight_presence_yes	0.708	sidewalk_condition_fair	0.31

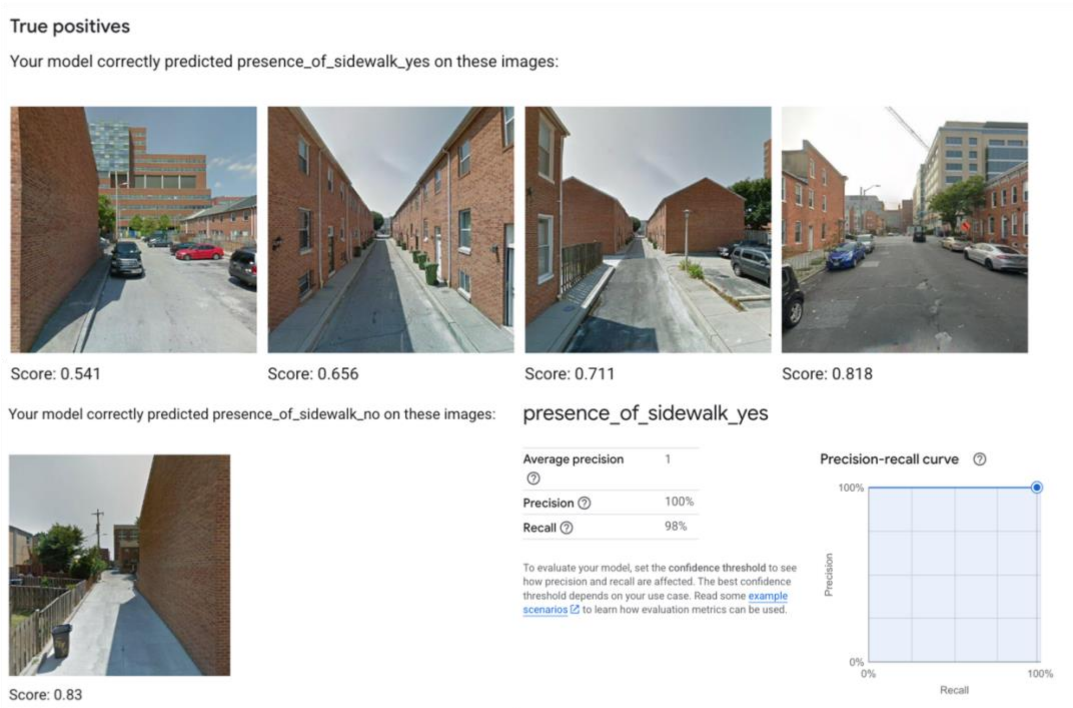

Figure 6. The presence of sidewalk label's performance (Source: Vertex AI platform)

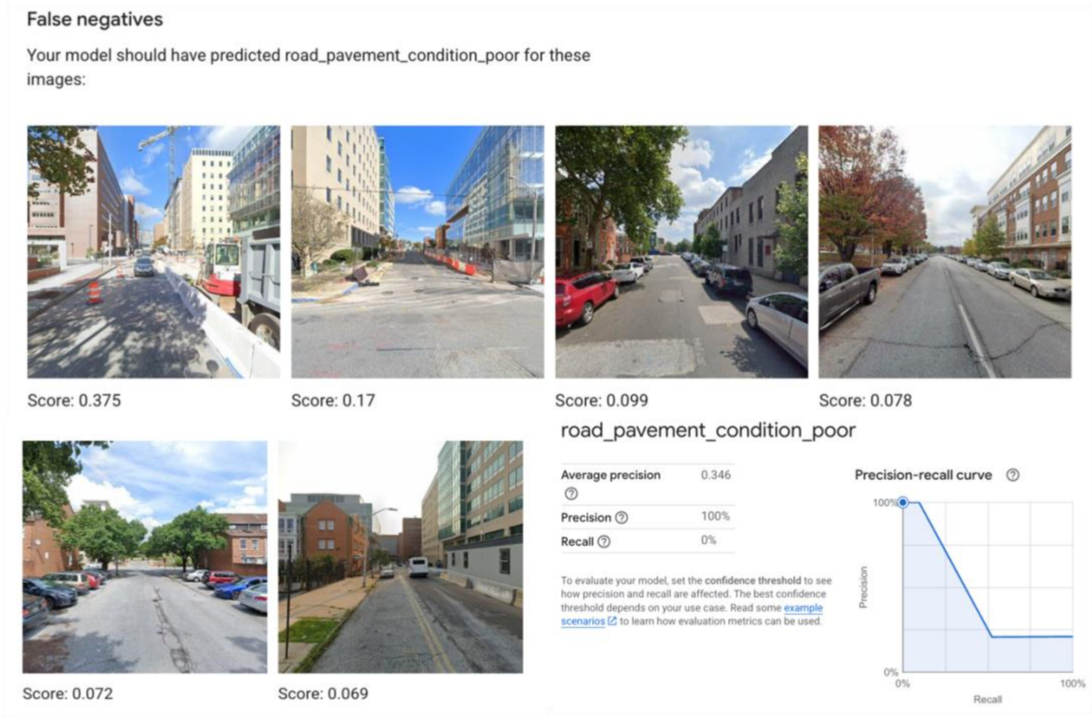

Figure 7. Road pavement condition label's performance (Source: Vertex AI platform)

5. Discussion

5.1 The implications and limitation of the model

The results highlight the potential of using street view images and machine learning to assess streetscape quality at scale. The model performed well in identifying objective, easily observable features such as the presence of sidewalks, crosswalks, and traffic signs, achieving precision scores above 0.75. It can be applied to identify and map disparities in streetscape infrastructure, finding underserved areas that may be exposed to greater pedestrian risks. The results also show the potential of using AI to label a large number of images.

However, the model struggles with subjective condition assessment (i.e., road pavement and sidewalk condition), where precision scores fall below 0.5. This may be because the rubric of condition assessment is subjective, ambiguous, and hard to tell the different conditions. Human should verify and refine condition assessment results.

5.2 Theoretical strength and weakness of the study

On one hand, the findings show the theoretical strength of using AI for scalable, objective streetscape assessment and spatial audits. By using street view images and Auto Machine Learning tools, public agencies can rapidly assess streetscape quality at large areas. The approach enables the identification of infrastructure disparities, like missing sidewalks or lack of crosswalks, which are often correlated with higher pedestrian risk and lower access to safe mobility. In this way, data science offers a method to detect and prioritize low-quality environments that may affect underserved communities. This aligns with the goals of environment justice, where the quality of the built environment should not determine one's risk exposure.

However, there are theoretical weakness in both tools and data. While the model identifies the presence or absence of physical infrastructure, it performs poorly on subjective or condition-based assessments, such as evaluating the quality of road pavement or sidewalks. The model fails to capture condition inequity. Moreover, the dataset is biased. Machine learning models only rely on what is visible and well-documented data. Street view images coverage is more frequent in downtown, while underserved neighborhoods or rural areas may have outdated or no images. The places most in need of investment may be least represented in the training data, which reinforces existing inequalities and unequal resource allocation.

In summary, while the study highlights how AI can be effective in assessing streetscape qualities, it lacks human judgment and underserved neighborhood data. Future work can address these limitations by integrating community knowledge and diverse data sources, which ensures that smart city tools promote rather than undermine urban equity.

6. Reference

Harvey, C., & Aultman-Hall, L. (2015). Urban streetscape design and crash severity. *Transportation Research Record*, *2500*, 1–8. https://doi.org/10.3141/2500-01

Nguyen, Q. C., Alirezaei, M., Yue, X., Mane, H., Li, D., Zhao, L., Nguyen, T. T., Patel, R., Yu, W., Hu, M., Quistberg, D. A., & Tasdizen, T. (2024). Leveraging computer vision for predicting collision risks: A cross-sectional analysis of 2019-2021 fatal collisions in the USA. *Injury Prevention*. https://doi.org/10.1136/ip-2023-045153

Ye, Yu, Zeng, Wei, Shen, Qiaomu, Zhang, Xiaohu, & Lu, Yi. (2019). The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images. *Environment and Planning B*, *46*(8), 1439–1457. https://doi.org/10.1177/2399808319828734

Zero deaths Maryland (2023), Pedestrian Crash Statistics 2022. Accessed May 11, 2025. https://zerodeathsmd.gov/news/pedestrian-crash-statistics-2022/#:~:text=From%202017%20to%202021%2C%20one,Baltimore%20and%20Washington%20metropolitan%20areas.

7. Appendix

Appendix Table 1. Hypotheses of street view indicators related to pedestrian crashes

Indicators	Hypothesis
Road Pavement Condition	Poor road pavement conditions reduces driver maneuverability, increase pedestrian crash risks.
Presence of Sidewalk	The absence of sidewalks forces pedestrians to share space with vehicles, increasing their exposure to moving traffic and elevating crash risk.
Sidewalk Condition	Poor sidewalk conditions (e.g., uneven, broken surfaces) increase tripping risk, discourage sidewalk use and push pedestrians onto roadways, leading to higher chances of vehicle-pedestrian conflicts.
Sidewalk Buffer Zone	The presence of a sidewalk buffer zone (e.g., trees, landscaping, bike lanes) physically separates pedestrians from vehicles, reducing the likelihood of crashes.
Crosswalk Presence	Marked crosswalks provide predictable crossing locations for pedestrians, improving driver awareness.
Traffic Signs Presence	Visible traffic signs alert drivers to pedestrian activity, enhancing compliance with traffic rules and reducing crash risks.
Streetlight Presence	Adequate streetlighting improves visibility for both pedestrians and drivers at night, reducing crash risk in low-light conditions.

Appendix Table 2. Rubric of labeling streetscape quality related to pedestrian crashes

Indicators	Labels	Description	Criteria
Road pavement condition	Poor, Fair, Good	Quality of the road surface	Poor: Large potholes, deep cracks, uneven surfaces, under construction, large patches. Driving conditions are very uncomfortable and bumpy Fair: Several areas of distress, with more than one form of cracking, rutting, shoving, etc. Driving conditions are somewhat uncomfortable, with vibrations and slight leaning in curvature, etc. Good: Small areas of cracks, potholes, bleeding, etc., minor surface vibration, but driving conditions are generally smooth and comfortable.
Presence of sidewalk	Yes, No	Presence of sidewalk	
Sidewalk condition	Poor, Fair, Good	Quality of pedestrian sidewalks	Poor: Large crack, uneven surfaces, missing sections, too narrow, or under construction, hazardous conditions for pedestrians, potentially leading to trips and falls. Surface is very uncomfortable to walk on, with significant obstructions or debris. Fair: Several areas of distress, such as cracking, unevenness, or minor obstructions, but surface is passable, and walking conditions are somewhat uncomfortable. Good: Minor cracks, smooth surface, well-maintained, and surface is comfortable for walking, with minimal obstructions.
Sidewalk buffer zone	Yes, No	Physical barriers between the sidewalk and roadway (i.e., street trees, landscaping, and bike lanes)	
Crosswalk presence	Yes, No	Presence of pedestrian crossings	
Traffic signs presence	Yes, No	Visibility of traffic signs	
Streetlight presence	Yes, No	Visibility of streetlights	

Appendix Table 3. Evaluation metrics of the model performance

Evaluation metrics	Explanation
AuPRC (Average Precision)	The area under the Precision-Recall (PR) curve. It ranges from 0 to 1, with higher values indicating better overall model performance in balancing precision and recall.
Confidence Threshold	A minimum confidence score used to filter predictions. Only predictions with scores equal to or above this threshold are returned.
Recall	The proportion of predictions with this class that the model correctly predicted.
Precision	The proportion of classification predictions produced by the model that were correct.