Open and Crowdsourced Data to Predict and Characterize Perceived Cycling Safety

Vanessa Frias-Martinez
iSchool and UMIACS
University of Maryland

The Benefits of Cycling

Cycling Safety is a Concern

- ~500K cyclists injured (~700 deaths) in 2013 (CDC Injury Center)
- Increase in both number of people riding and number of deaths

Causes behind Cycling Accidents

- Inadequate infrastructure
- More drivers on the road (low gas prices)
- Smartphone use and distractions
- Increasing population in urban areas

Approaches to Increase Safety

- Vision Zero initiatives to eliminate all traffic fatalities include:
  - Proactive policy
  - Infrastructure changes
  - Education

- Initiatives have not always been successful
  - In 2018 - LA 5% increase in cyclist and pedestrian deaths

Understanding Safety Perception at the Street Level

- Safety measures focus a lot on crash numbers, which is an incomplete statistic
- We need a better understanding of perceived cycling safety at the street level

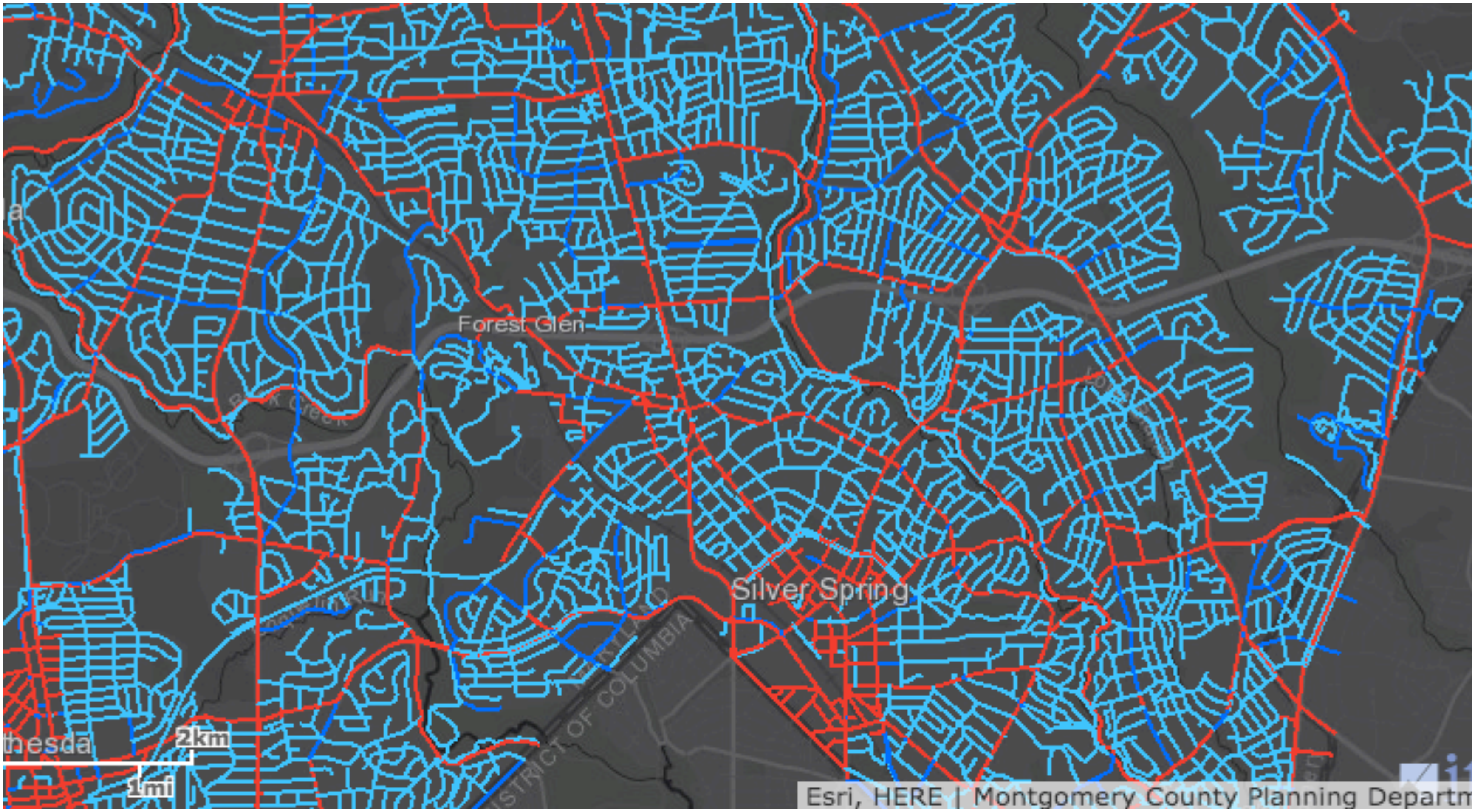
Understanding Safety Perception at the Street Level

- Identify locations where changes might be more needed (decision makers, cyclists and advocacy groups)

Understanding Safety Perception at the Street Level

- Identify locations where changes might be more needed (decision makers, cyclists and advocacy groups)
- Evaluate connectivity and cycling safety per community to reveal accessibility and equity issues

Cycling Safety Maps



Cycling Safety Maps

- Associations between Attributes and Cycling Safety Perceptions

Attributes

- Measures: traffic speed, traffic volume, frequency of parking turnover
  - Require expensive sensors that cannot be available in every street
- Observations from video recordings
  - Expensive and not scalable

Cycling Safety Perceptions

- Cycling safety perceptions associated to attributes are based on:
  - Logical intuitions (e.g., more cars, less safe)
  - Qualitative studies, generalizability not validated

Proposed Approach - Attributes

Proposed Approach – Perception Associations

New Approach to Perceived Cycling Safety Maps

Our Approach

Explore the use of Open Datasets and Open Street Maps as a source for perceived cycling safety attributes

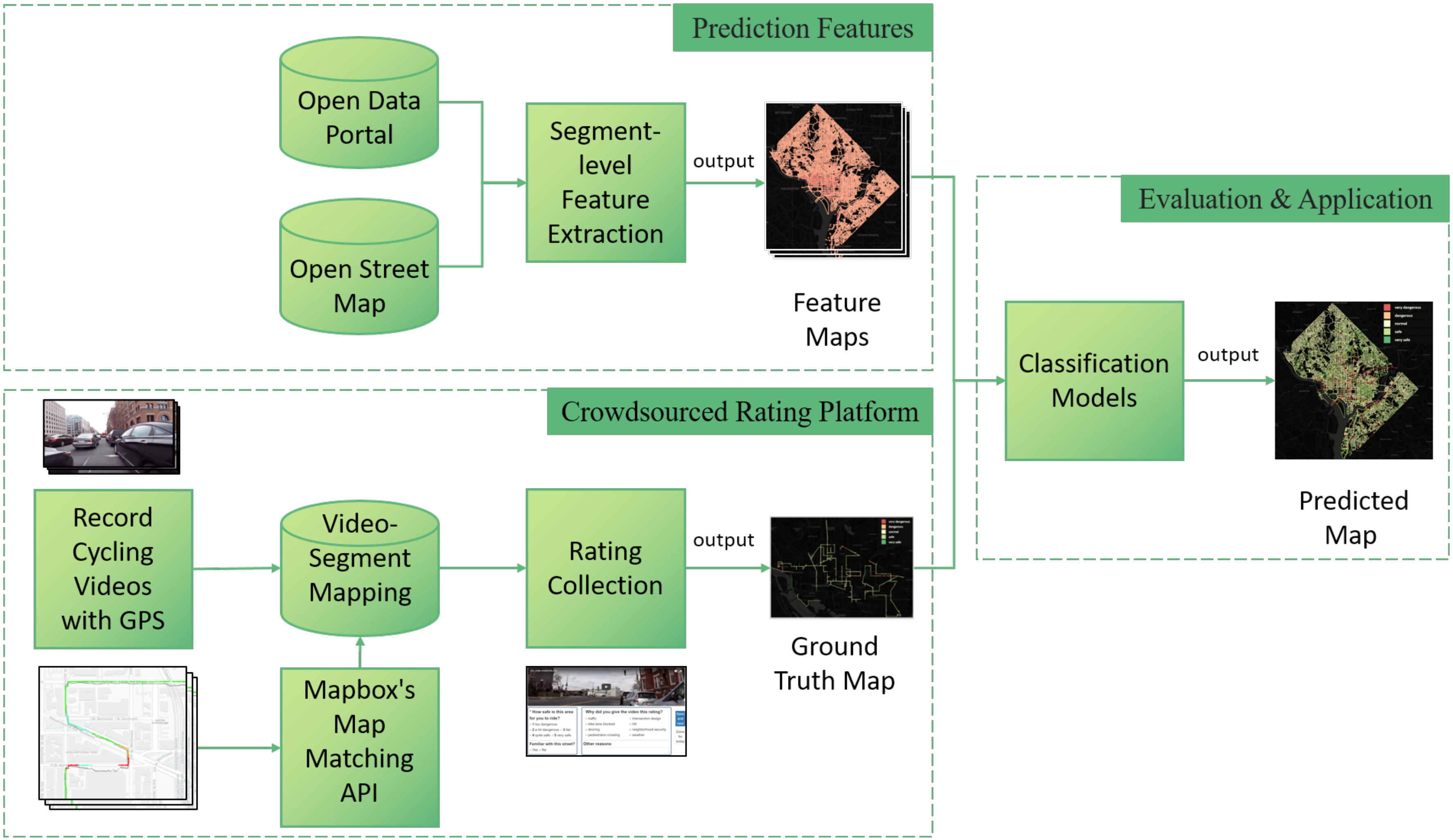
Open Data

- Lowering the bar to comprehensive cycling safety maps:
  - Open Data Repositories: 2600 cities worldwide (some cities have the data, but not public)
  - Open Street Maps: 4 million small- to mid-sized cities

Our Approach

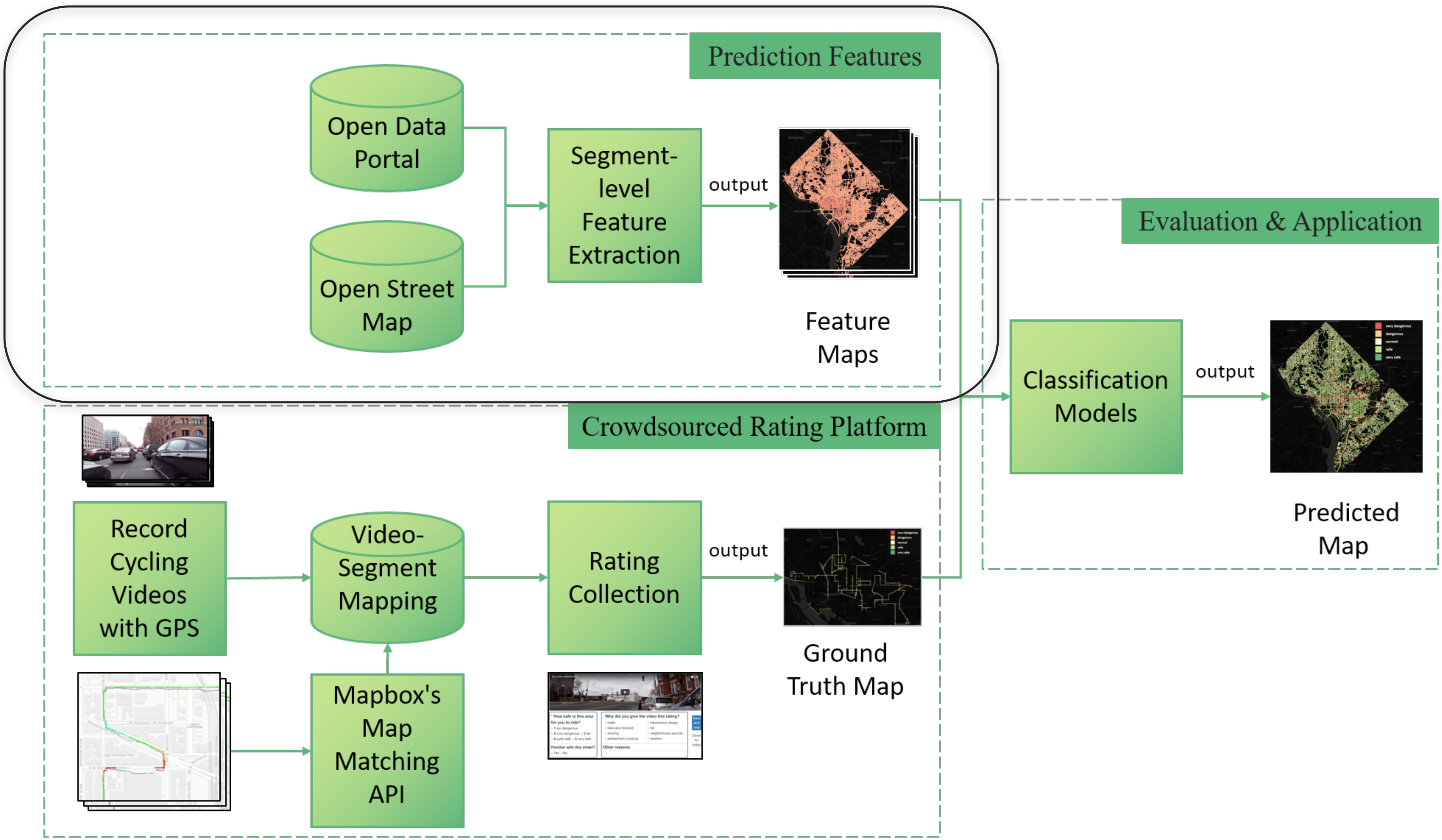
Crowdsource cycling safety perceptions from cyclists (ground truth) and build a ML model to test associations between attributes and safety perceptions

Framework


Prediction Features
Open Data Portal
Open Street Map
Segment-level Feature Extraction
output
Feature Maps
Crowdsourced Rating Platform
Record Cycling Videos with GPS
Video-Segment Mapping
Mapbox's Map Matching API
Rating Collection
output
Ground Truth Map
Evaluation & Application
Classification Models
output
Predicted Map

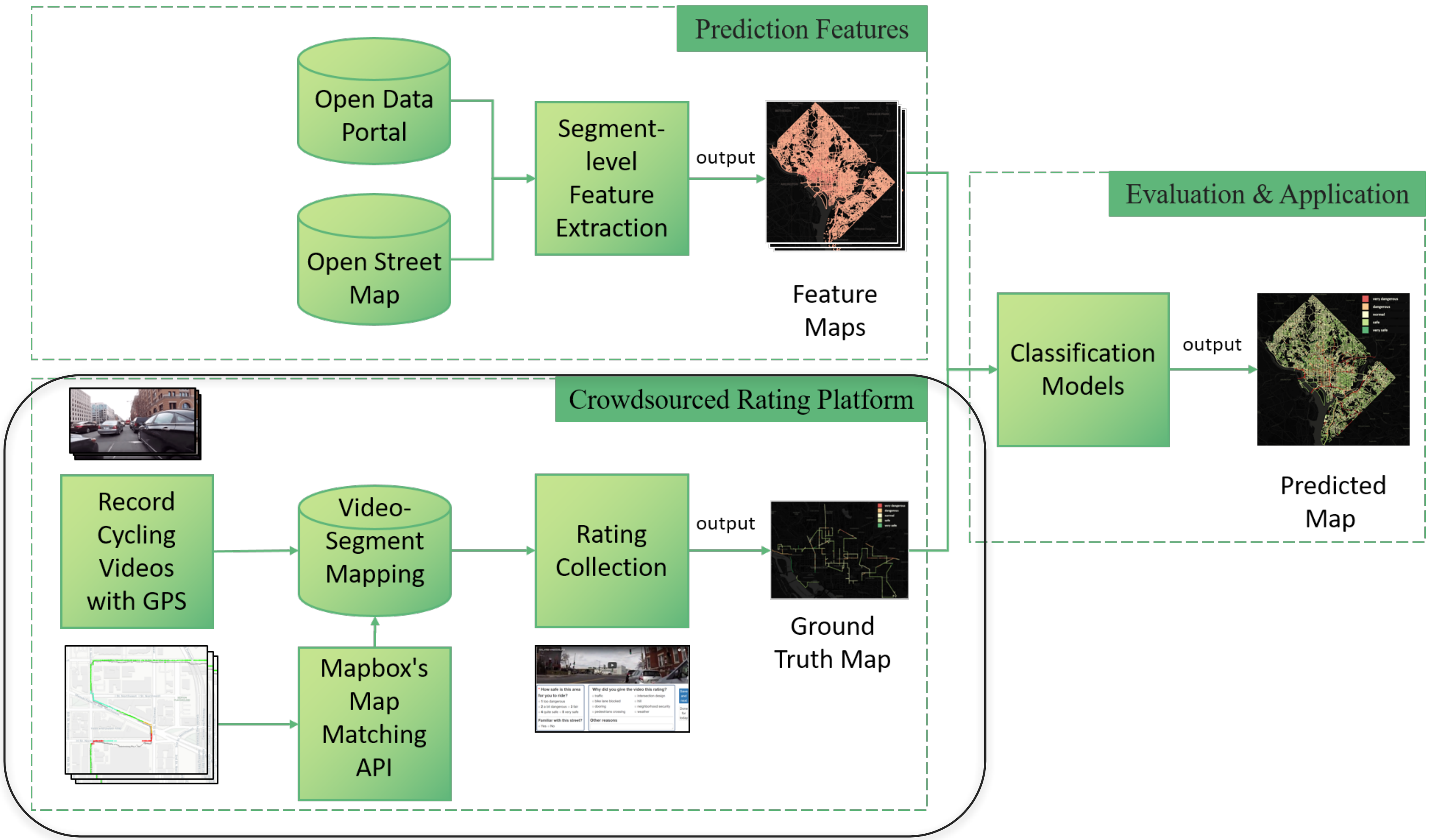
Framework

A. Perceived Cycling Safety Attributes

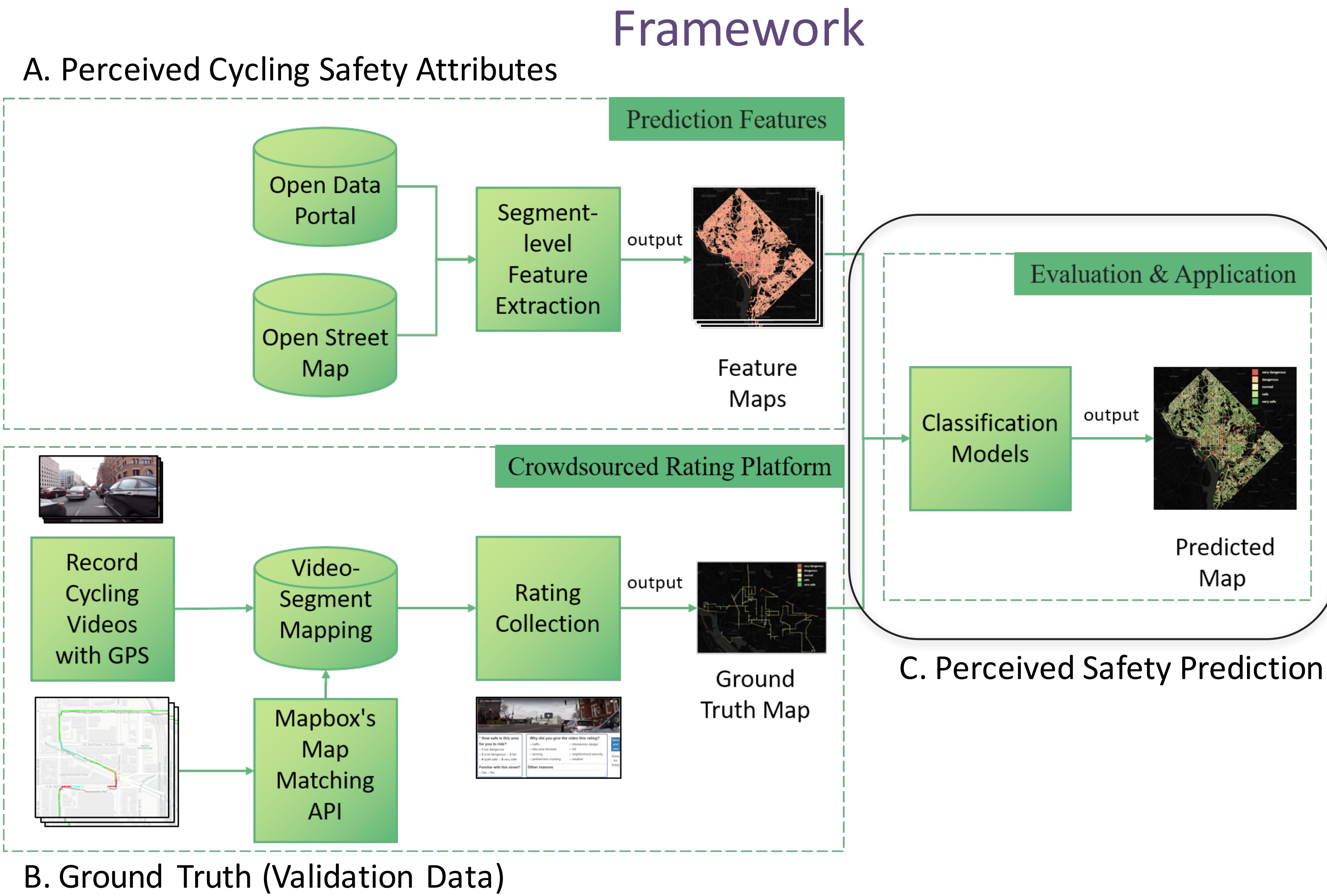


Framework

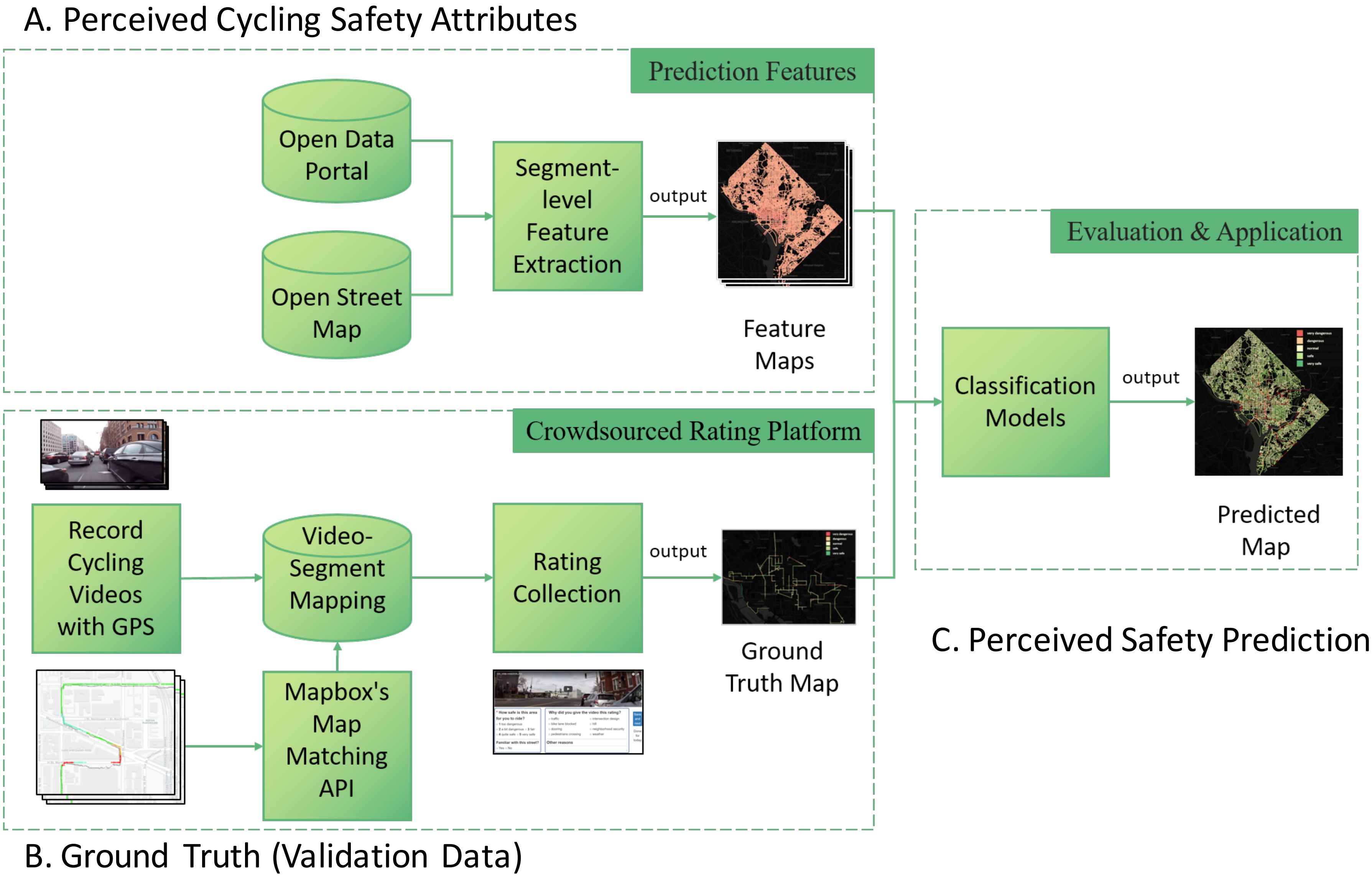
A. Perceived Cycling Safety Attributes



B. Ground Truth (Validation Data)


Framework
A. Perceived Cycling Safety Attributes
Prediction Features
Open Data Portal
Open Street Map
Segment-level Feature Extraction
output
Feature Maps
B. Ground Truth (Validation Data)
Crowdsourced Rating Platform
Record Cycling Videos with GPS
Video-Segment Mapping
Mapbox's Map Matching API
Rating Collection
output
Ground Truth Map
C. Perceived Safety Prediction
Evaluation & Application
Classification Models
output
Predicted Map

Framework



WASHINGTON, D.C.

A. Perceived Cycling Safety Attributes

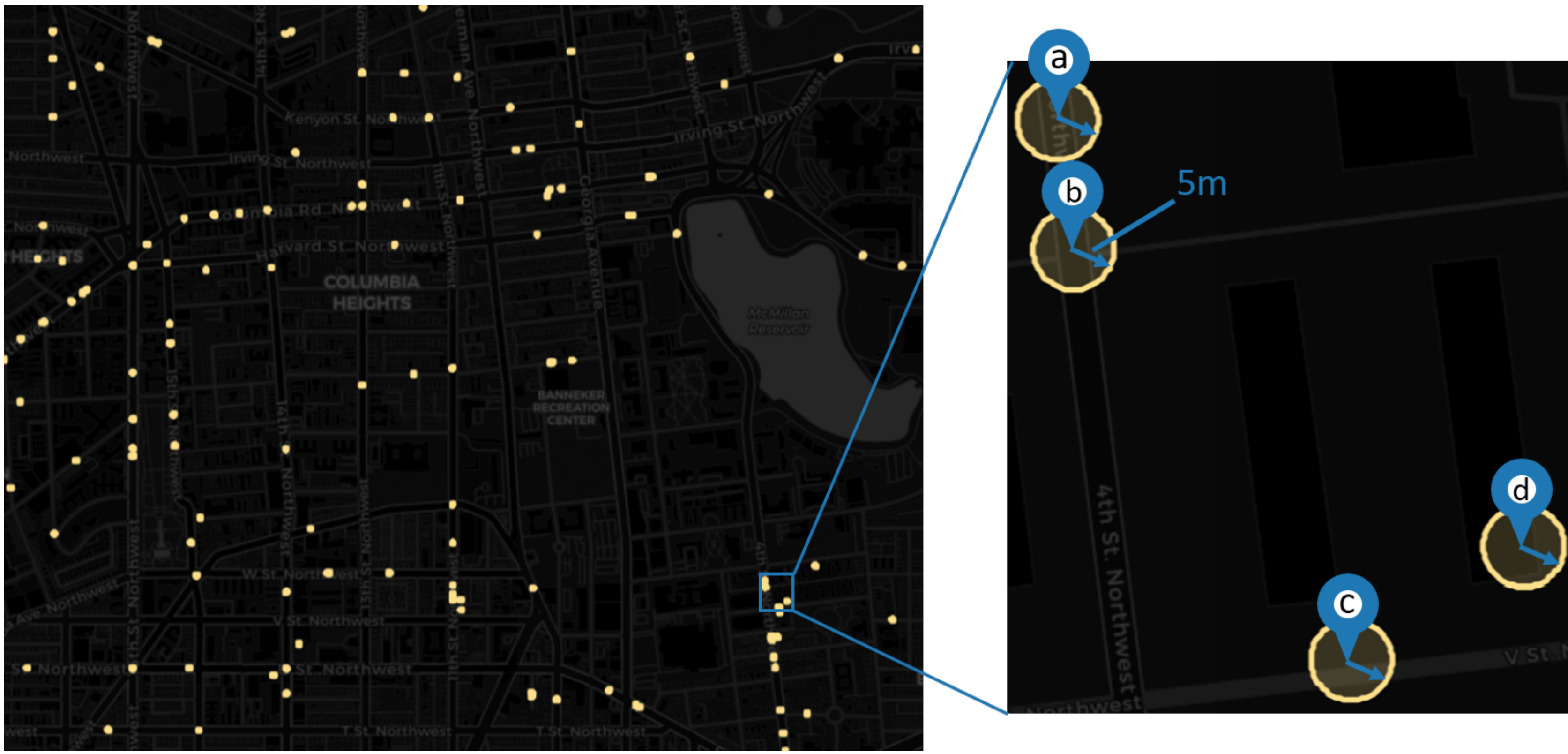
Open Data

- Qualitative research on cycling safety factors has identified that these factors play a role in safety perception:
  - Social fabric e.g., crime rates (Open Datasets)
  - Built environment e.g., presence of cycling facilities (Open Street Maps)

Social Attributes

- Crime rates
- Points of interest
- Bicycle crashes
- 311 requests related to street conditions
- Parking and moving violations

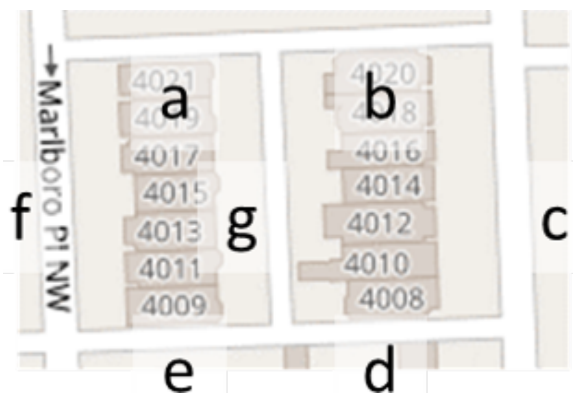
Impact Buffer



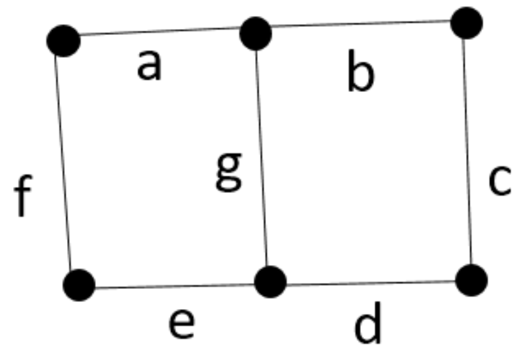
Built Environment Attributes

- Road network characteristics
- Presence of cycling facilities
- Graph-based road network features

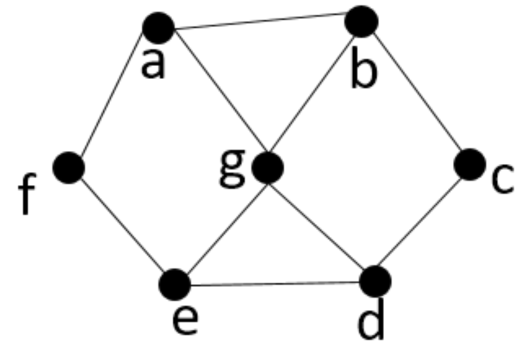
Graph-based Features



(a) Road Network



(b) Primal Graph

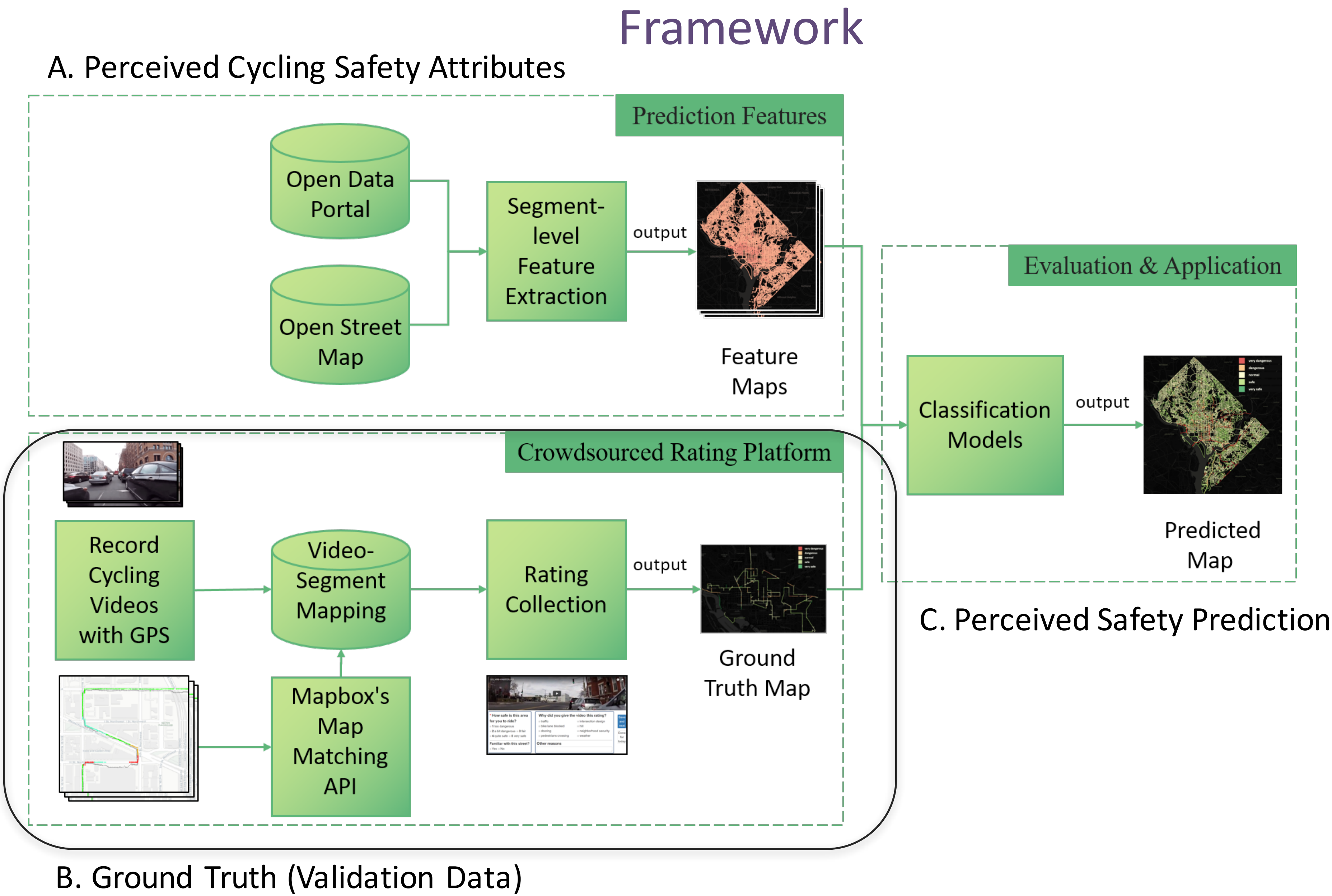


(c) Dual Graph

Attributes for DC

- 63 built environment features
  - 11 road network types
  - 39 graph-based (centrality measures)
  - 13 cycling facilities types
- Social features: monthly average across types (6) and monthly average per type (148)
  - 11 types for crime data
  - 11 types for crash data
  - 72 for 311 requests
  - 10 POIs
  - 36 types of parking violations
  - 8 moving violations

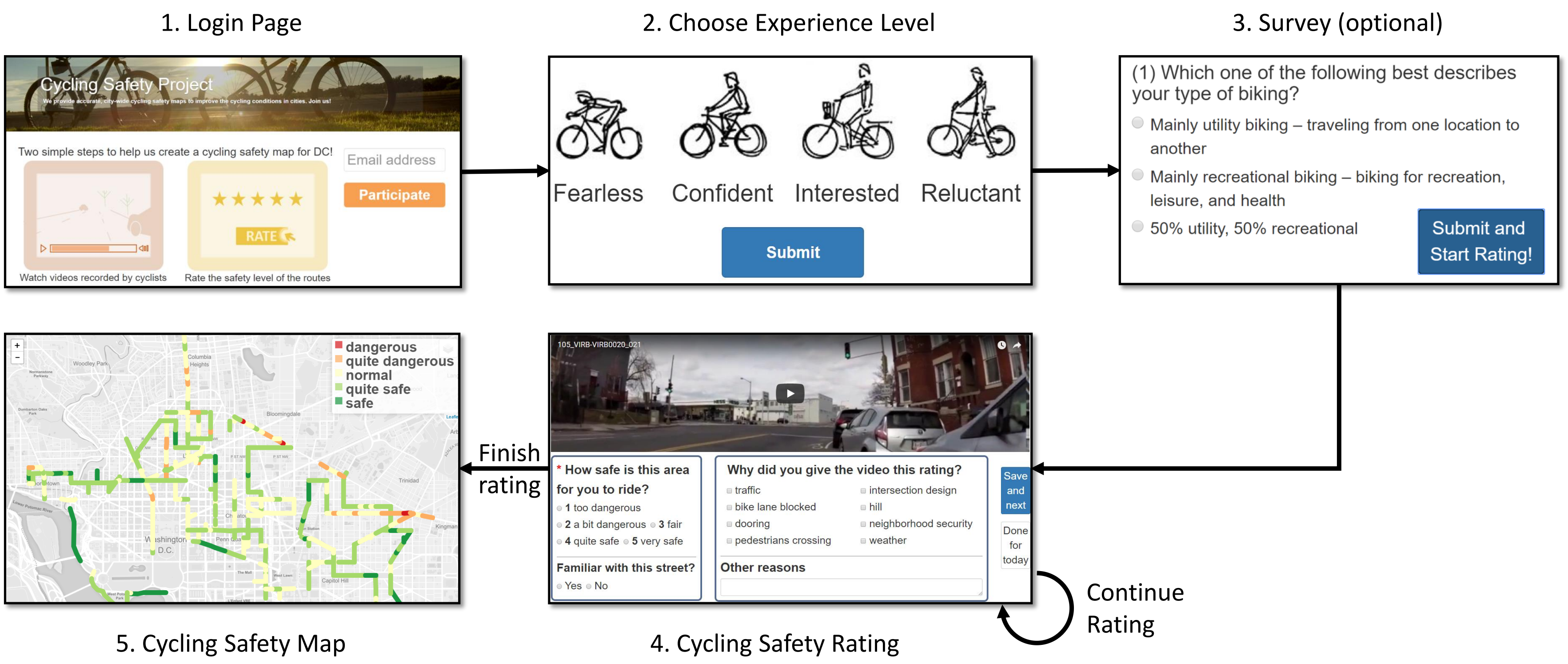
B. Ground Truth Data Collection


Framework
A. Perceived Cycling Safety Attributes
Prediction Features
Open Data Portal
Open Street Map
Segment-level Feature Extraction
output
Feature Maps
Evaluation & Application
Classification Models
output
Predicted Map
C. Perceived Safety Prediction
Crowdsourced Rating Platform
Record Cycling Videos with GPS
Video-Segment Mapping
Rating Collection
output
Ground Truth Map
Mapbox's Map Matching API
B. Ground Truth (Validation Data)

Ground Truth Collection

- Recorded cycling videos in Washington, D.C
- Built a webpage to crowdsource cycling safety perceptions
- WABA promoted our project in cycling events
- Collected cycling safety perceptions from cyclists

Crowdsourced Safety Perceptions



Cycling Safety Tool

From Videos to Segments

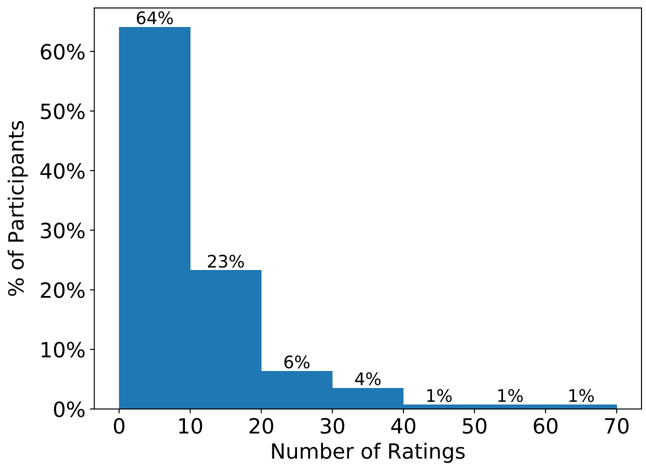
- Videos are rated multiple times by cyclists
- Each segment might appear in multiple videos
- Final segment label (1-5) is averaged across video ratings and weighted by % of street segment present in video

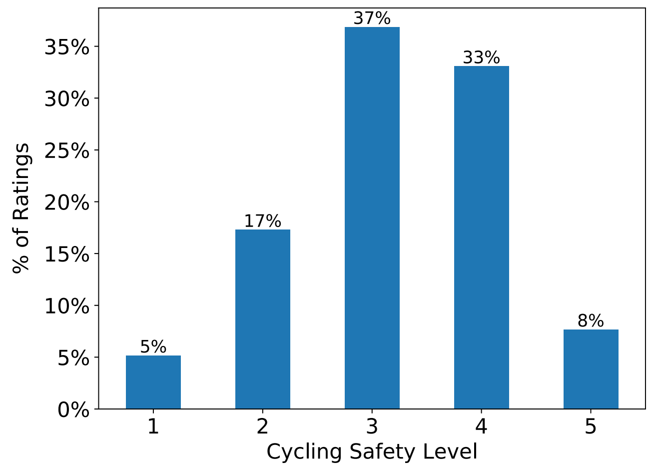
Open source Code

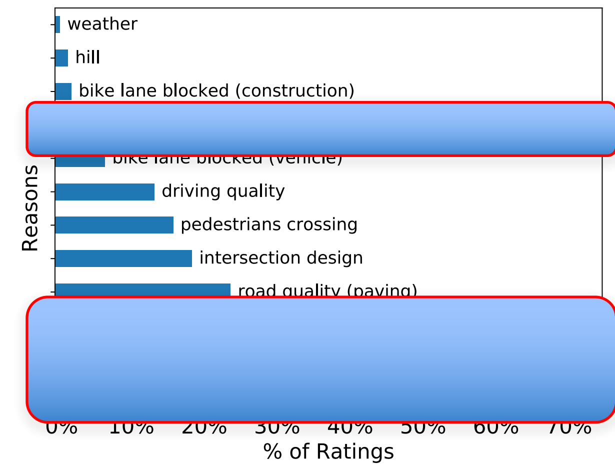
Personal and Rating Features

Personal Features	Safety Ratings	Rating Reasons
Usual trip purpose	1: too dangerous, I would never ride there	Traffic
Age	2: a bit dangerous, I wouldn't ride here unless I have to	Bike lane design (or lack of)
Ethnicity	3: fair, I need to be cautious to ride here	Bike lane blocked (vehicle)
Education level	4: quite safe, I would easily ride here	Dooring (car door might hit cyclist)
Marital status	5: very safe, even a kid could ride here	Pedestrians crossing
Gender		Intersection design
Driver's license		Driving quality
Access to car		Road quality (paving)
Household income		Hill
Length of residence in city		Neighborhood security
Type of biking		Weather

Ground Truth Collection







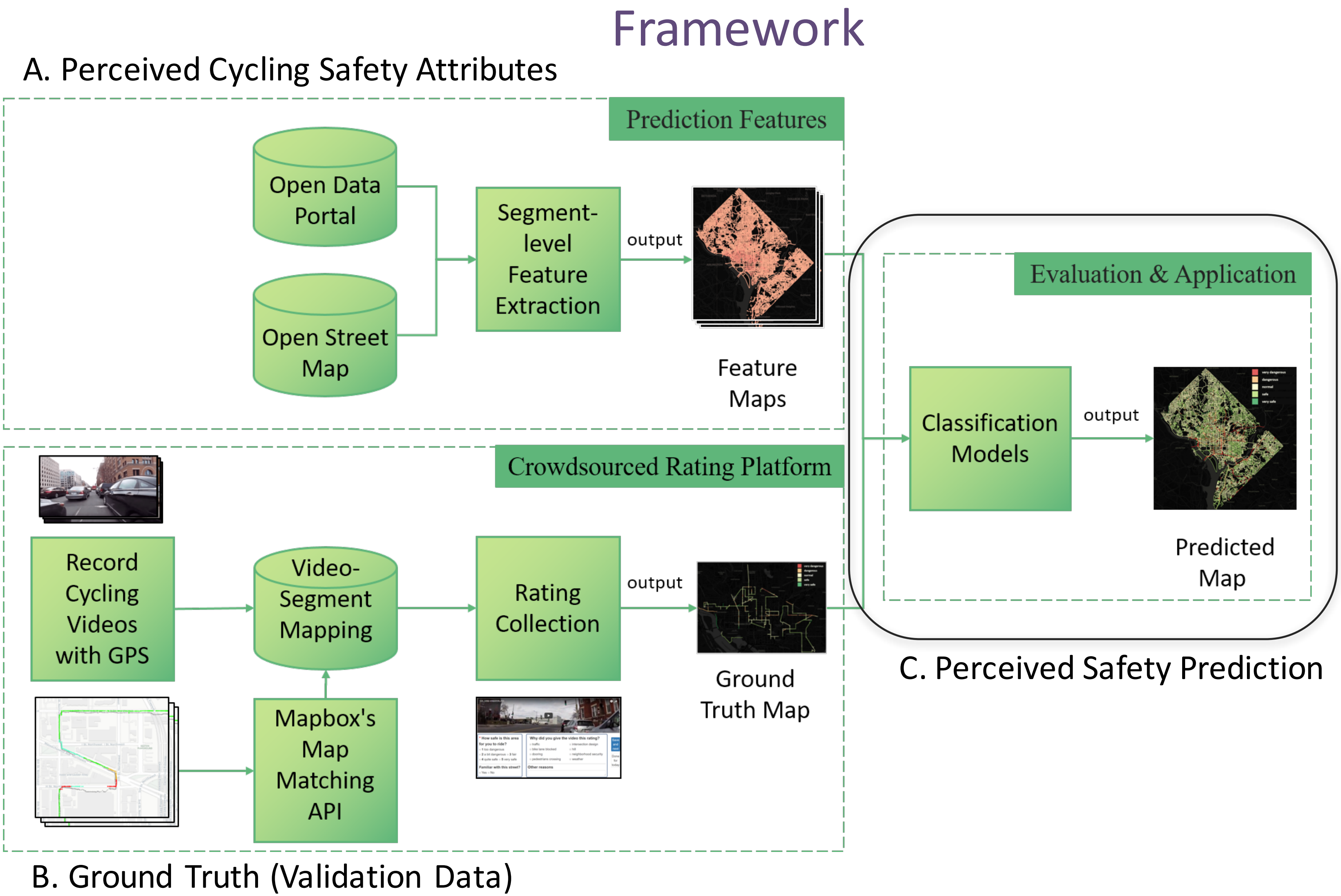
1476 ratings from 159 participants

Ground Truth Collection

Very Dangerous

Very Safe

C. Perceived Cycling Safety Prediction


Framework
A. Perceived Cycling Safety Attributes
Prediction Features
Open Data Portal
Open Street Map
Segment-level Feature Extraction
output
Feature Maps
B. Ground Truth (Validation Data)
Crowdsourced Rating Platform
Record Cycling Videos with GPS
Video-Segment Mapping
Mapbox's Map Matching API
Rating Collection
output
Ground Truth Map
C. Perceived Safety Prediction
Evaluation & Application
Classification Models
output
Predicted Map

Perceived Cycling Safety Prediction

- To assess whether open and crowdsourced data can be used to
  - predict perceived cycling safety
  - assess associations between attributes and cycling safety perceptions

Prediction Results

- Dataset:
  - Segments with features
  - Crowdsourced cycling safety labels
  - mRMR feature selection
  - 70%-30% training-testing 10 times and report averages

Prediction Results

METHOD / FEATURES	BuiltEnv	Social [total]	Social [type]	BuiltEnv+Social [total]	BuiltEnv+Social [type]
SVM	0.59/0.31	0.52/0.27	0.54/0.31	0.58/0.34	0.58/0.36
Decision Trees (DT)	0.46/0.34	0.48/0.26	0.49/0.30	0.56/0.31	0.52/0.36
Bagging DT (BAG)	0.60/0.43	0.52/0.29	0.57/0.40	0.62/0.36	**0.65/0.42**
Random Forest (RF)	**0.62/0.45**	0.54/0.30	0.57/0.39	0.63/0.37	0.63/0.41
Gradient Boosting (GBoost)	0.60/0.41	0.55/0.31	0.58/0.41	0.62/0.40	0.64/0.44
XGBoost	0.57/0.37	0.55/0.34	**0.59/0.43**	0.62/0.37	**0.65/0.44**
Baseline	0.45/0.13	0.45/0.13	0.45/0.13	0.45/0.13	0.45/0.13

Prediction Results

METHOD / FEATURES	BuiltEnv	Social [total]	Social [type]	BuiltEnv+Social [total]
SVM	0.59/0.31	0.52/0.27	0.54/0.31	0.58/0.34
Decision Trees (DT)	0.46/0.34	0.48/0.26	0.49/0.30	0.56/0.31
Bagging DT (BAG)	0.60/0.43	0.52/0.29	0.57/0.40	0.62/0.36
Random Forest (RF)	**0.62/0.45**	0.54/0.30	0.57/0.39	0.63/0.37
Gradient Boosting (GBoost)	0.60/0.41	0.55/0.31	0.58/0.41	0.62/0.40
XGBoost	0.57/0.37	0.55/0.34	**0.59/0.43**	0.62/0.37
Baseline	0.45/0.13	0.45/0.13	0.45/0.13	0.45/0.13

Prediction Results

METHOD / FEATURES	Social [total]	BuiltEnv+Social [total]	BuiltEnv+Social [type]
SVM	0.52/0.27	0.58/0.34	0.58/0.36
Decision Trees (DT)	0.48/0.26	0.56/0.31	0.52/0.36
Bagging DT (BAG)	0.52/0.29	0.62/0.36	**0.65/0.42**
Random Forest (RF)	0.54/0.30	0.63/0.37	0.63/0.41
Gradient Boosting (GBoost)	0.55/0.31	0.62/0.40	0.64/0.44
XGBoost	0.55/0.34	0.62/0.37	**0.65/0.44**
Baseline	0.45/0.13	0.45/0.13	0.45/0.13

Prediction Results

Improving Predictions

- Imbalanced dataset
  - Over/under-sampling with SMOTE
  - XGBoost only improved 1%
- Spatial Autocorrelation with Moran's I
  - Enhance feature vector with spatially autocorrelated features from nearby segments (<150m)
  - Improved macro F1 scores by 4%

Improving Predictions

- Weighting safety labels by Familiarity and Cycling Experience boosts 1%-3%
  - Familiarity/not
  - Cycling Experience: fearless, confident, interested, reluctant

Improving Predictions

- Three (0.88/0.60) or Four (0.70/0.51) classes improve results and macro values

METHOD	micro/Macro-F1
Five-class (XGBoost, I>0.68)	0.66/**0.48**
Four-class (GBoost, I>0)	**0.70**/0.51
Three-class (XGBoost, I>0)	**0.88/0.60**

Important Predictive Attributes

- XGBoost:
  - Closeness centrality of the segment,
  - Presence of cycling facilities,
  - Crime rates, and
  - Slope

Predicted Map

Future Work

- Safety perceptions and route choice
  - Combine safety predictions with data from micro-mobility solutions

- Understand changes in safety perceptions due to interventions

- Safe cycling accessibility across communities

Thank You!

Vanessa Frias-Martinez
vfrias@umd.edu