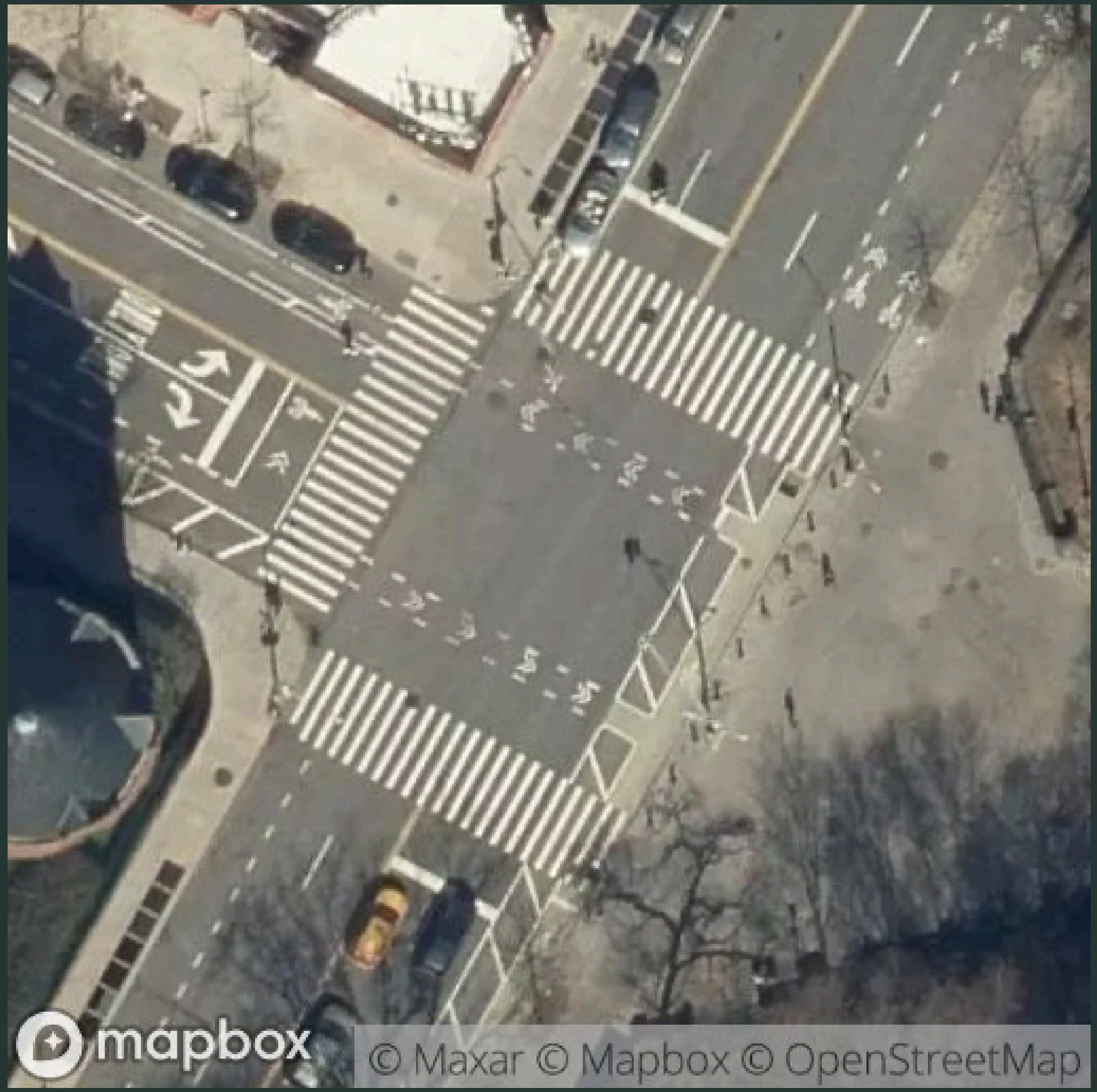
TRAFFIC VOLUME AND ACCIDENTS-BASED INTERSECTION SAFETY CLASSIFICATION

AVANI MAHAWAR, ABHINEET GULATI, JAANYA GOYAL



https://api.mapbox.com/styles/v1/mapbox/{style_id}/static/{longitude},{latitude},18,0,0/256x256@2x?
access_token=pk.eyJ1IjoiYWdwdSIsImEiOiJjbW1kemNzbWMwOGFiMm9xdWJreG04bWI2In0.SfcFrEiHm4MCuVPW8ZW6hw

Images taken from Mapbox API

PROBLEM STATEMENT

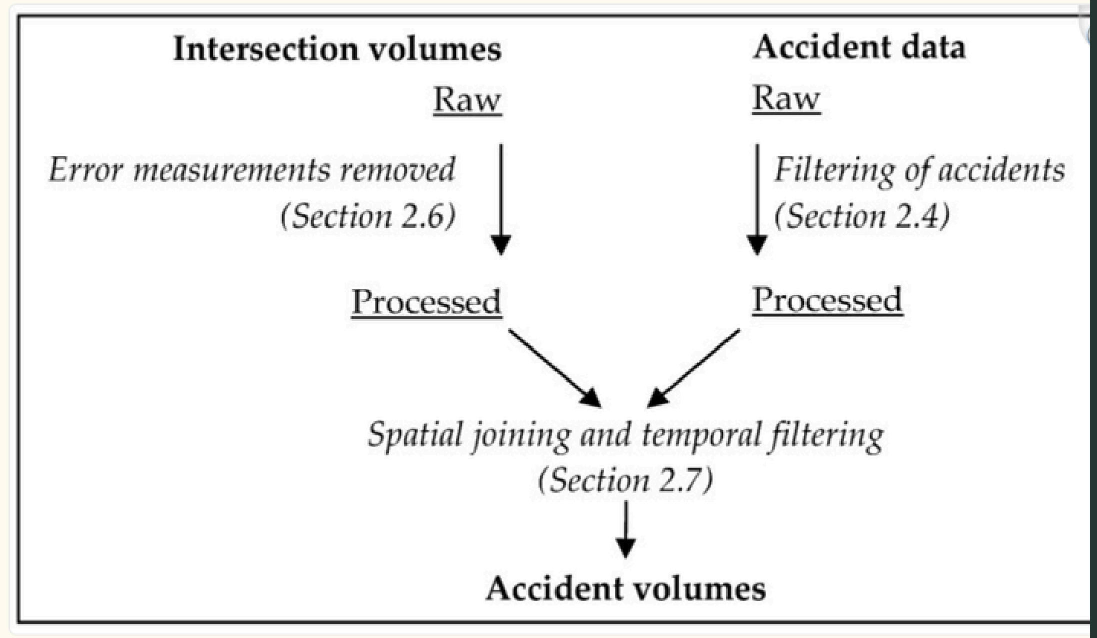
Can we proactively predict the safety of an intersection using historical crash data and traffic exposure measures, such as Annual Average Daily Traffic (AADT) and traffic density, to identify high-risk intersections before future crashes occur?

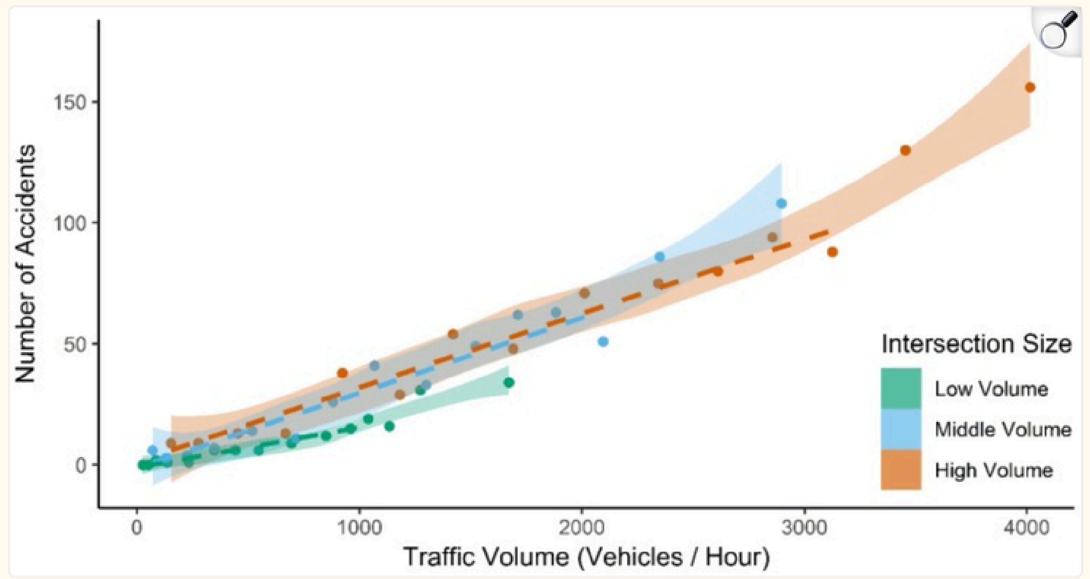
POTENTIAL APPLICATIONS

1. Transportation authorities can automatically analyze large numbers of intersections and identify hotspots before significant accidents occur.
2. Since many regions lack reliable crash statistics, we want to develop this technology so that image based safety classification can be used.
3. Authorities can increase safety protocols wherever the intersection is classified as unsafe so that lots of accidents don't need to occur.

LITERATURE REVIEW

RELATIONSHIP BETWEEN TRAFFIC VOLUME AND ACCIDENT FREQUENCY AT INTERSECTIONS





Methodology

- Analyzed the relationship between traffic volume and accident frequency at intersections.
- Used a congestion index to normalize traffic conditions across intersections for fair comparison.

Results

- Accident frequency shows a nearly linear increase with traffic volume and congestion up to congestion levels 12–13 across different intersection types.
- At highly congested conditions, crash frequency rises sharply in a non-linear manner, indicating that severe congestion significantly increases intersection crash risk.

Gap

- The study still relies heavily on historical crash data, making the approach largely reactive.
- No predictive intersection-level Safety Index is developed for early risk identification.

How will we resolve the gap?

- Our project addresses this gap by developing a proactive predictive Safety Index using crash history, AADT, and traffic density measures.

Retallack AE, Ostendorf B. Relationship Between Traffic Volume and Accident Frequency at Intersections. Int J Environ Res Public Health. 2020 Feb 21;17(4):1393. doi: 10.3390/ijerph17041393. PMID: 32098180; PMCID: PMC7068508.

MODELING OF ACCIDENTS USING SAFETY PERFORMANCE FUNCTIONS

Methodology

- Developed Safety Performance Functions (SPFs) to model the relationship between traffic exposure and accident frequency.
- Used AADT-based descriptive analysis models to predict fatality frequency, where traffic volume was identified as the major factor influencing crash occurrence
- Applied multivariate causality models incorporating variables such as traffic volume, road geometry, weather conditions, and human factors to estimate yearly fatality frequency.

Results

- Accident frequency increases with increasing traffic volume and roadway exposure.
- SPF-based models help identify high-risk road segments and support proactive safety planning.

Gaps

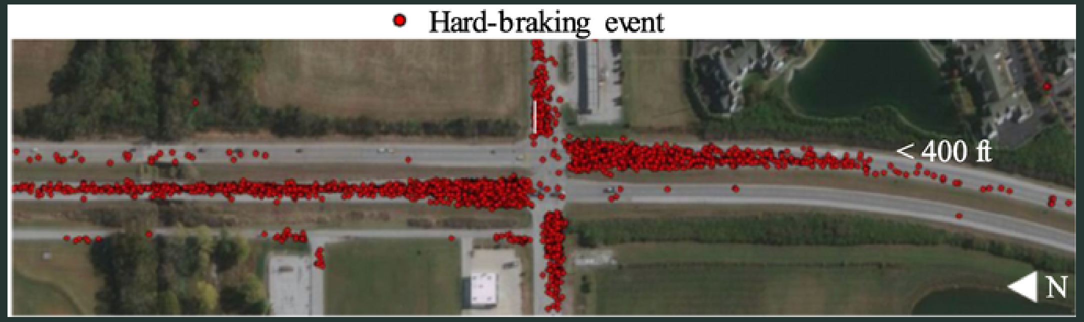
- The paper uses traditional statistical models that rely on fixed variables and historical data only.
- Visual intersection characteristics and real-time predictive safety assessment are not incorporated into the model.

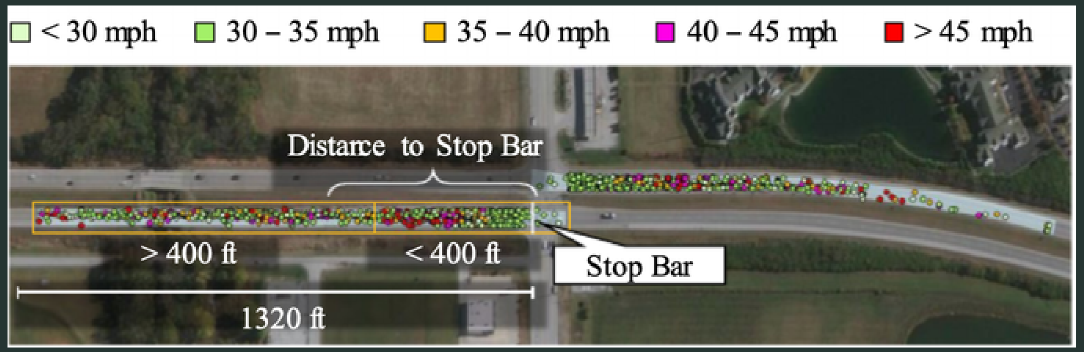
How do we resolve the gap?

- Our work extends SPF-based safety modeling by developing a proactive predictive Safety Index using crash history, AADT, and visual intersection characteristics.

Msallam, Majed & Al-Marafi, Mohammad Nour. (2019). Modeling of Accidents Using Safety Performance Functions. 10.17758/IAAST.A0614076.

A PROACTIVE APPROACH TO EVALUATING INTERSECTION SAFETY USING HARD-BRAKING DATA





Methodology

- Compared hard-braking frequency with rear-end crash frequency using statistical correlation analysis.
- Analyzed hard-braking events relative to the stop bar by categorizing events within and beyond 400 ft upstream.

Results

- Hard-braking events occurring more than 400 ft upstream showed stronger correlation with crash frequency.
- Hard-braking behavior can act as a proactive surrogate safety indicator for identifying risky intersections.

Gap

- Visual intersection characteristics and AI-based predictive safety assessment are not included.
- The study primarily focuses on hard-braking behavioral data and does not incorporate traffic exposure metrics such as AADT or traffic density.

How do we resolve the gap?

- Our work extends proactive safety analysis by integrating crash history, AADT, traffic density, and visual intersection characteristics into a predictive Safety Index.
- The proposed framework aims to identify high-risk intersections before severe crash accumulation occurs.

Hunter, Margaret & Saldivar-Carranza, Enrique & Desai, Jairaj & Mathew, Jijo & Li, Howell & Bullock, Darcy. (2021). A Proactive Approach to Evaluating Intersection Safety Using Hard-Braking Data. Journal of Big Data Analytics in Transportation. 3. 10.1007/s42421-021-00039-y.

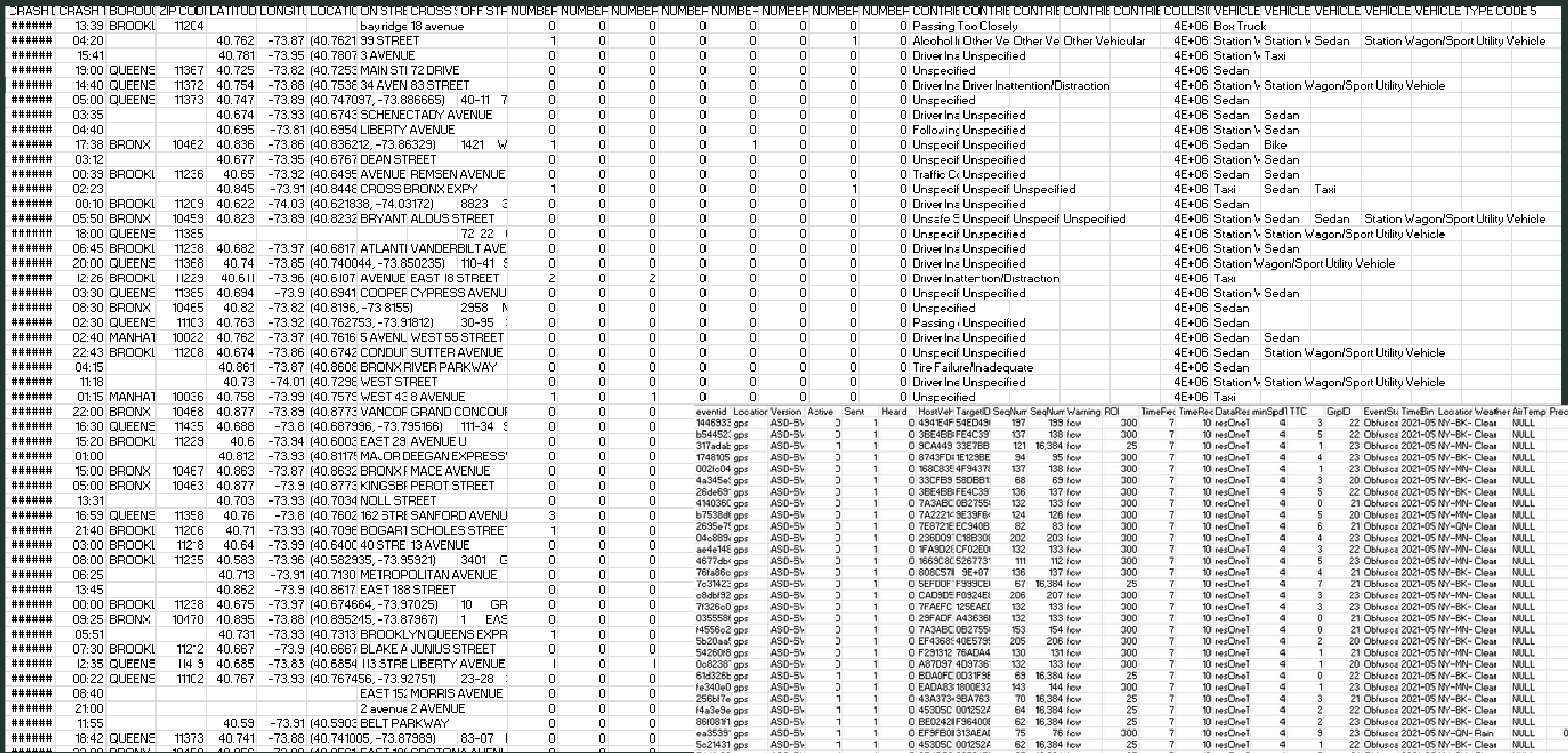
Paper	Methodology	Key Contribution	Limitation / Gap	Our Proposed Solution
Relationship Between Traffic Volume and Accident Frequency at Intersections	Used traffic volume, congestion index, and crash data to analyze accident frequency at intersections.	Showed that crash frequency increases with traffic volume and severe congestion.	Relies mainly on historical crash analysis and traditional statistical relationships. No predictive Safety Index or visual analysis.	Develop a proactive predictive Safety Index using crash data, AADT, traffic density, and visual intersection characteristics.
Modeling of Accidents Using Safety Performance Functions (SPFs)	Developed SPF models using AADT and roadway variables to predict crash/fatality frequency.	Established mathematical crash prediction models for proactive roadway safety assessment.	Limited to statistical SPF modeling and historical crash data. Does not include AI or visual intersection analysis.	Extend SPF-based modeling with predictive AI-driven safety assessment and intersection-level Safety Index generation.
A Proactive Approach to Evaluating Intersection Safety Using Hard-Braking Data	Used hard-braking events and statistical correlation with rear-end crashes for proactive safety evaluation.	Introduced surrogate safety indicators for early crash-risk identification.	Focuses mainly on behavioral probe data and excludes traffic exposure metrics and visual infrastructure characteristics.	Integrate crash history, AADT, traffic density, and visual intersection features for comprehensive proactive risk prediction.
Our Proposed Work	Combines historical crash data, AADT, traffic density, and visual intersection characteristics to develop a predictive Safety Index.	Proactive identification of high-risk intersections before severe crash accumulation occurs.		The system creates a proactive Safety Index for intersections using historical crash data, Average Annual Daily Traffic (AADT), traffic density, and normalized exposure metrics. It identifies high-risk intersections before crashes happen, employing data cleaning, normalization, and spatial alignment techniques to support preventive infrastructure planning and urban road safety management.

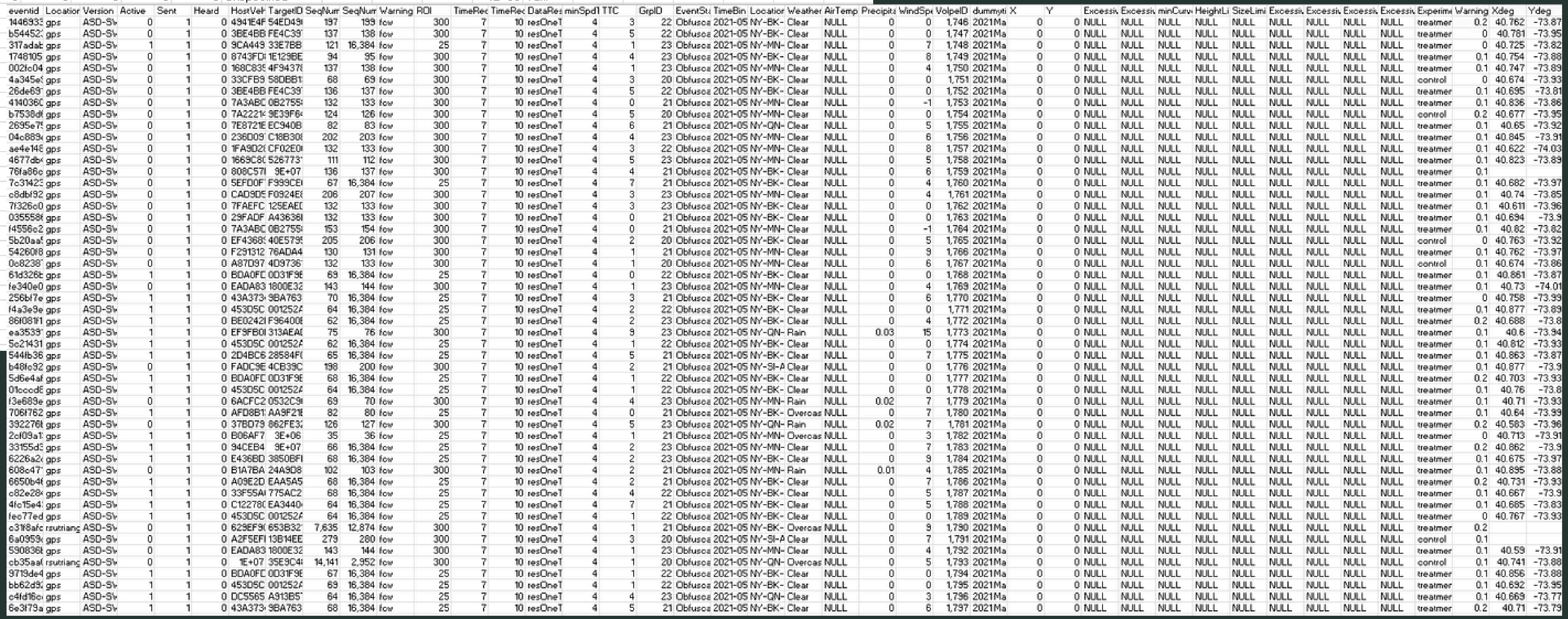
DATASET

DATASETS	**AADT**	**Crash data**
Nature	csv	csv
Features and datapoints	61 features, 174843 rows for 2019	31 features, 614098 rows
How was data collected? Any ethical concerns?	NYS GIS data, private member and private organization	Police reported traffic crashes (NYPD). Victim sensitivity, bias in policing

DATASET AND FEATURES PREPROCESSING

Data Preprocessing





data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/data_preview

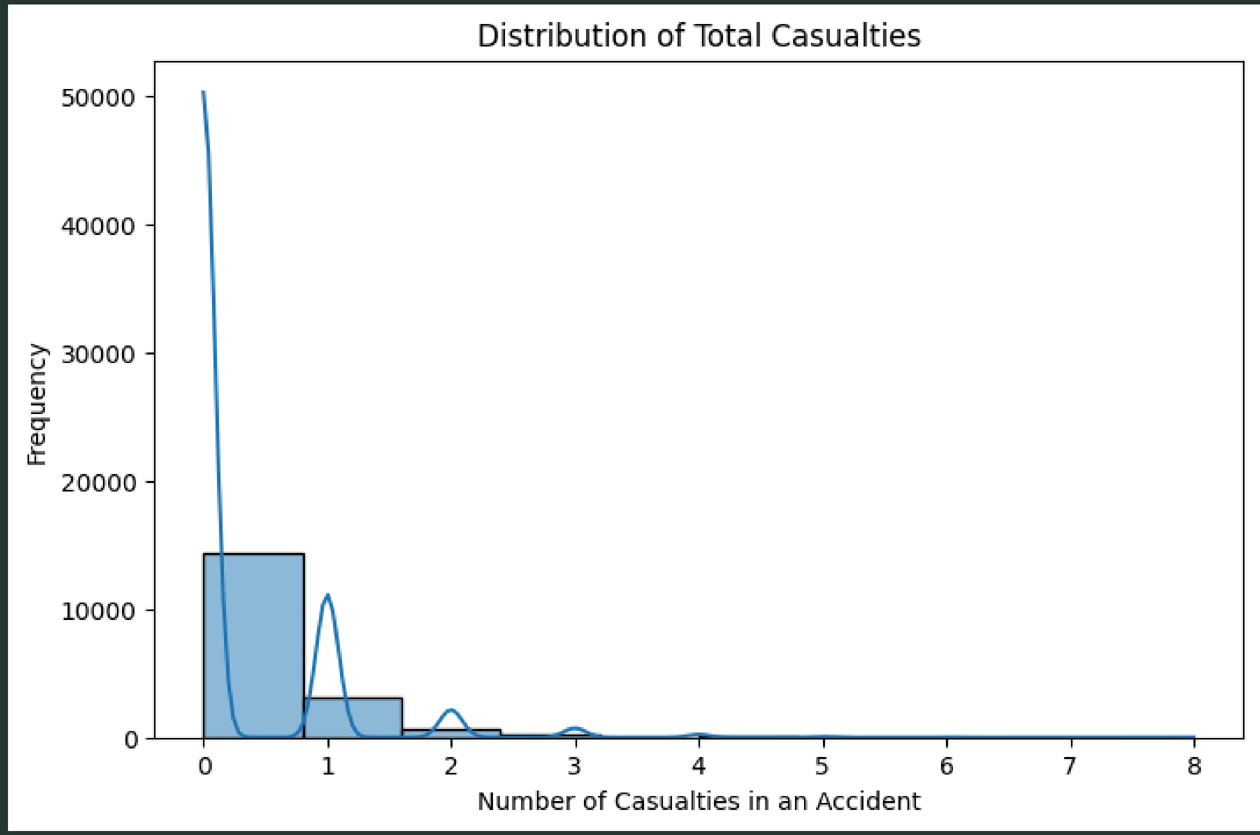
https://catalog.data.gov/dataset/nyc-cv-pilot-ie-processed-data

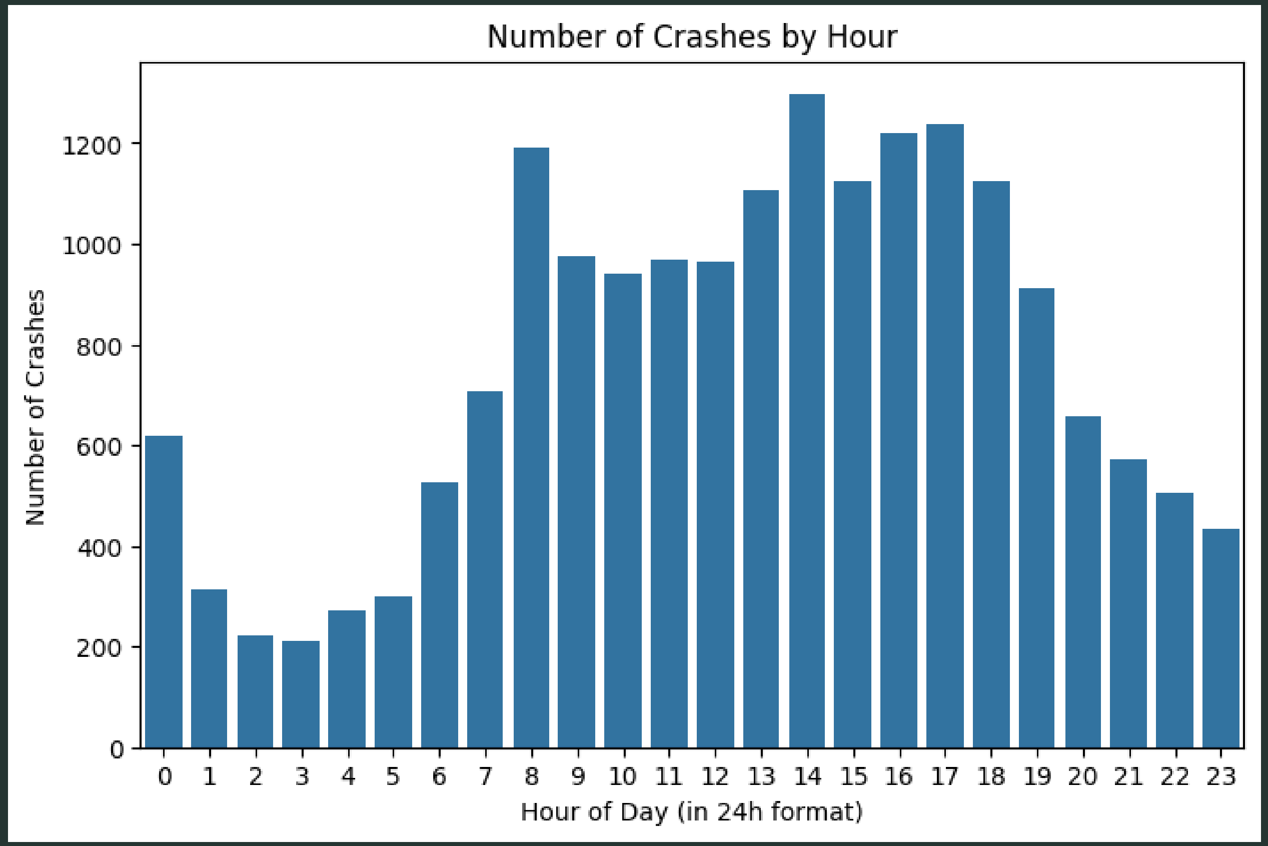
NYC Motor Vehicles Collisions Dataset

Steps:

1. Dataset Cleaning
   a. Standardizing Column Names
   b. Datetime conversion
   c. Removal of rows with missing values in critical columns
   d. Basic numeric cleaning (e.g., ensure no negative injury counts)
2. Feature Engineering: Deciding which columns are worth keeping and which aren't by plotting correlation matrix with crashes.

From 30 Columns to 6 Columns!





data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/data_preview

https://catalog.data.gov/dataset/nyc-cv-pilot-ie-processed-data

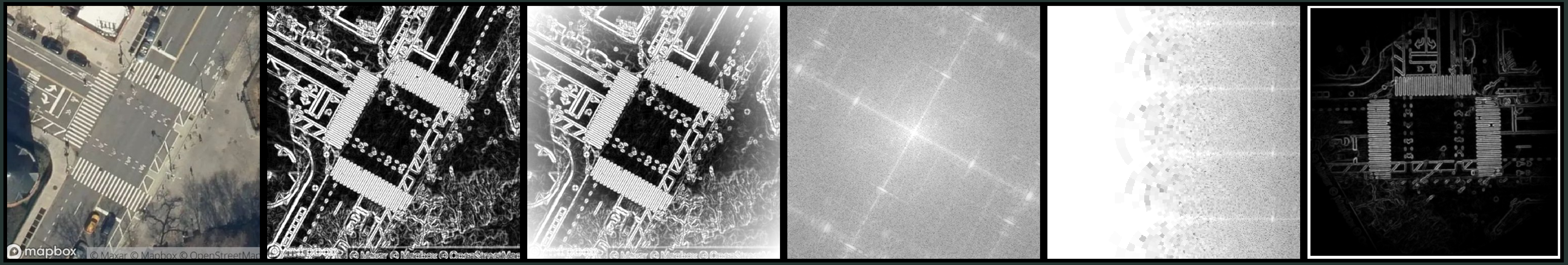
Annual Average Daily Traffic

OBJECTID	RouteNumb	RouteSignin	Qualifier	Ramp	DOT_ID	CO	FactorGroup	BeginDescrip	EndDescript	HPMSSampl	BIN	RRCrossing	NHS
3.47E+08		5	1		191661	1	30	PEEKSKILL H	WESTCHEST				
3.47E+08			0	NY1060103	272242	1	30	SR106 NB (	908 G WB (C				Y

RoadwayNar	RoadNum	Speed	Oneway	RCSTA	RCStation	GIS_ID	FROM_MEAS	TO_MEASUR	CountStatsI	CountyFIPS	StationNum	CCSTA
MILL ST	CR23	40		842026	842026	1.92E+08	0	1.32	45247619	079	2026	
NY 106 at 90		0	Y	33129	033129	2.72E+08	0	0.17	45283914	059	3129	

BeginDescription	EndDescription	AADT
PEEKSKILL HOLLOW RD	WESTCHESTER CL	2117
SR106 NB (OFF)	908 G WB (ON)	5131
VIELE AVE	EAST BAY AVE	285
POMEROY ST	END 11/41 OLAP START 11/13/4	5322
POMEROY ST	END 11/41 OLAP START 11/13/4	5322
OSCAWANA LK RD	LAKE FRONT RD	69
SO PARK AVE	SMITH ST	1391
WEST ST NB (OFF)	I-478 SB (ON)BRKLYN BTRY TNL	
OCEAN PKWY	FLATBUSH AVE	11036
STRONG AVE	NEW YORK AVE	2520
NY5	I 81	2994
END	WATERFRONT PLACE	3205
NY 26	NEWMAN RD	995
WATER ST	NY 37	697
NY60 N Main St	Shumla Rd	959

Image Preprocessing



(a) (b) (c) (d) (e) (f)

Illustration of preprocessing steps for a sample image. The plots show (a) the original MapBox image, (b) line detection based on the Scharr operator, (c) emphasizing intersection center (black one), (d) Fourier transform, (e) transformation to log-polar space to determine rotation angle, and (f) the final rotated image

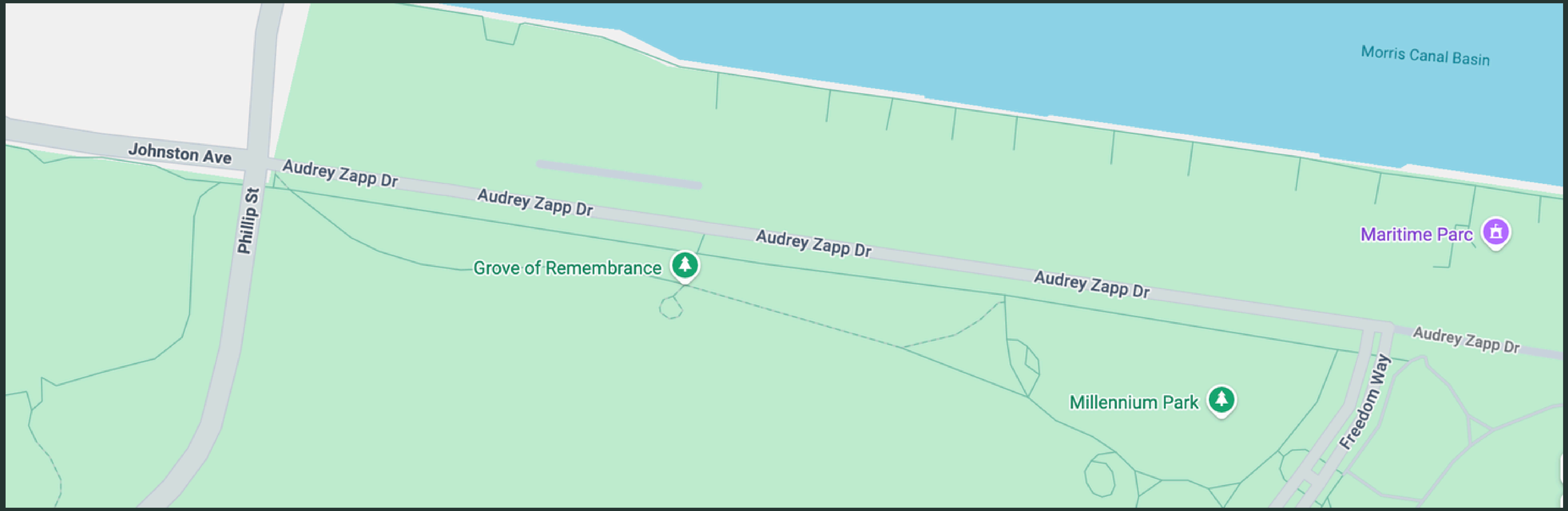
https://api.mapbox.com/styles/v1/mapbox/satellite-v9/static/{longitude},{latitude},16,0,0/256x256@2x?access_token=pk.eyJ1IjoiYWdwdSIsImEiOiJjbW1kemNzbWMwOGFiMm9xdWJreG04bWI2In0.SfcFrEiHm4MCuVPW8ZW6hw

HOW DID WE COMBINE THE DATASETS?

Step 1: Retrieve lat longs and make pairs

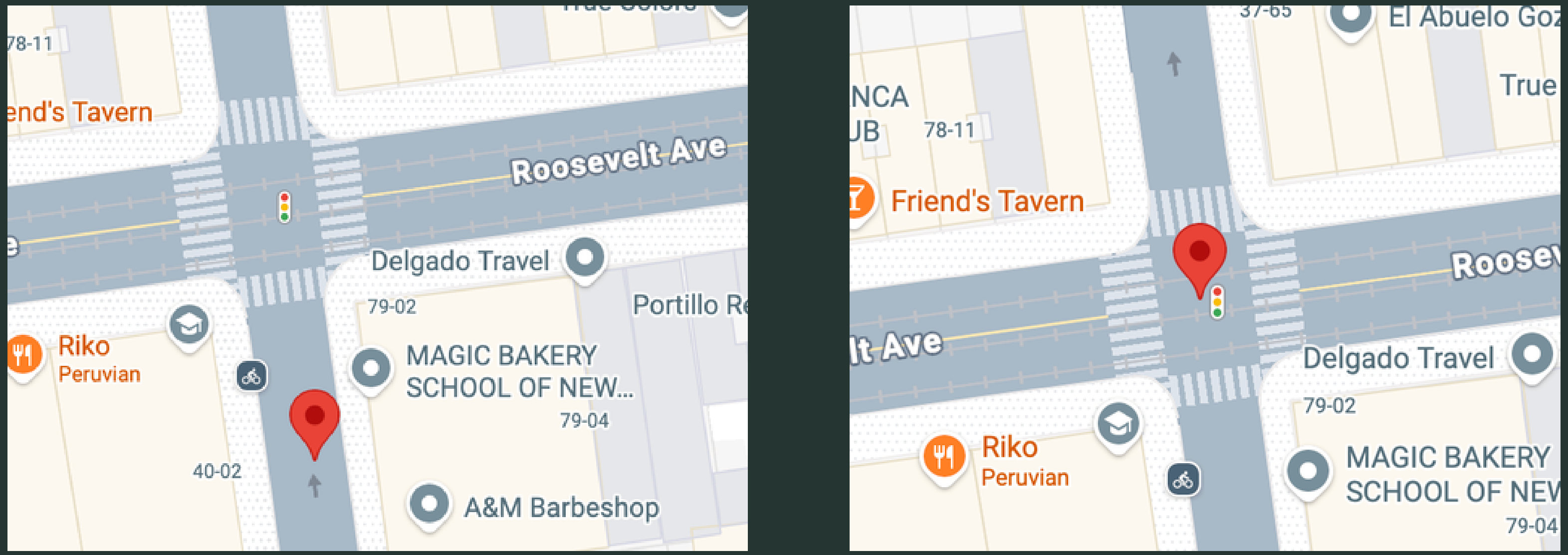
```
Processing batch 35000 to 35100
✔ SHIRLEY LA -> 40.6607705, -73.5259505
✔ HY PL -> 40.8566174, -73.1198581
✔ E FORT HILL RD -> 40.9796291, -73.8281523
```

StreetA	StreetB	LatA	LonA	LatB	LonB	AADT
100TH ST	108TH ST	40.7504913	-73.8661174	40.7371384	-73.8518019	4532
100TH ST	CENTERVILL	40.7504913	-73.8661174	40.6689291	-73.8389919	1140
QUEENS BLV	100TH ST	40.733389	-73.8707886	40.7504913	-73.8661174	6683
101ST AVE	97TH AVE	40.687271	-73.8354413	40.6966418	-73.807406	272
LIBERTY AVE	101ST AVE	40.6744049	-73.8965126	40.687271	-73.8354413	1620
SUTPHIN BLV	101ST AVE	40.70546	-73.810708	40.687271	-73.8354413	1564
101ST ST	115TH ST	40.6127043	-74.0349276	40.6714353	-73.8244881	5007

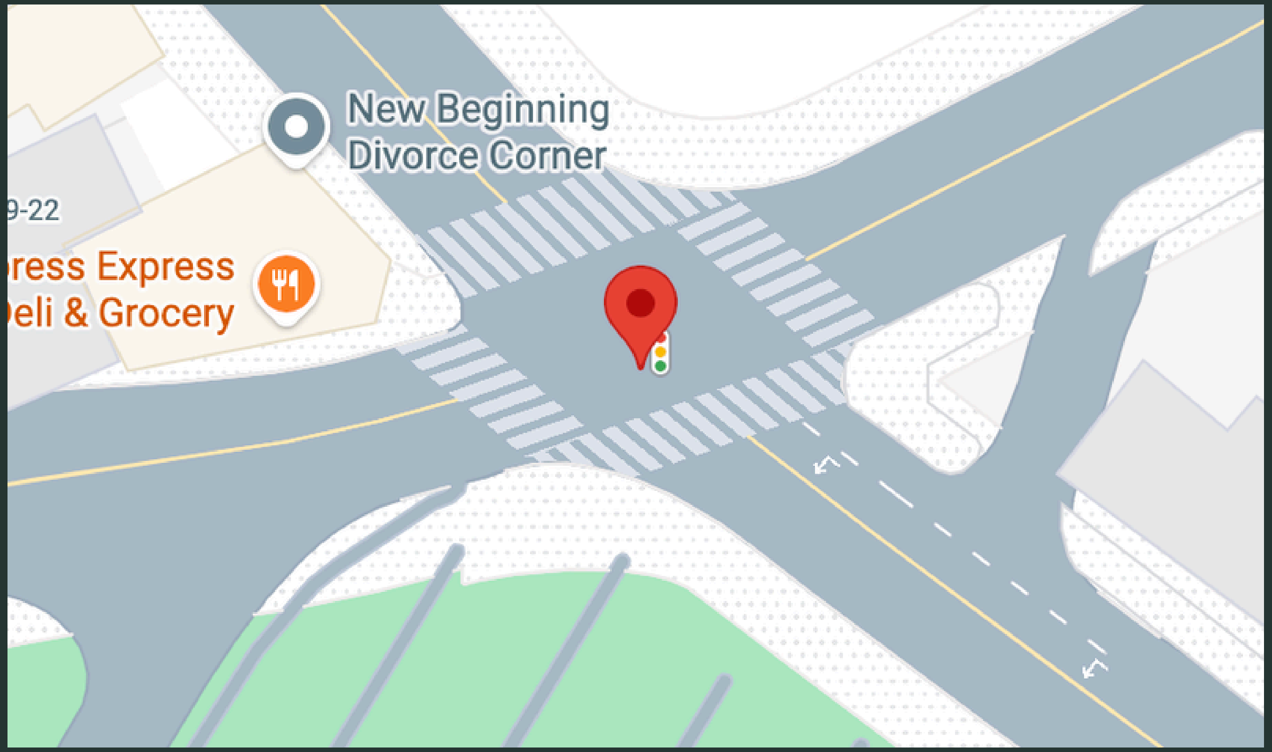


Average of AADT values for intersections with multiple values

Step 3: Retrieve intersections and map crashes



roughly 60 feet




```
df["crash_score"] = (
    10 * df["NUMBER OF PEDESTRIANS KILLED"] +
    8 * df["NUMBER OF CYCLIST KILLED"] +
    5 * df["NUMBER OF MOTORIST KILLED"] +
    5  * df["NUMBER OF PEDESTRIANS INJURED"] +
    4  * df["NUMBER OF CYCLIST INJURED"] +
    2  * df["NUMBER OF MOTORIST INJURED"]
)
```

List of intersections retrieved from (OpenStreetMap networkx) OSMNX library using python.

Build crash scores using the crash data.

Linear combination of people injured/killed.

Step 4: Match the latitude longitudes of the two files

nearest_intersection_id	intersection_lat	intersection_lon	image_pat	image_exi
2	40.78716409999999	-73.7919229	D:\nyc_rav	TRUE
4	40.7647567	-73.758627	D:\nyc_rav	TRUE
5	40.75346709999999	-73.7441643	D:\nyc_rav	TRUE
6	40.7542921	-73.743724	D:\nyc_rav	TRUE
7	40.757608600000005	-73.7477778	D:\nyc_rav	TRUE
8	40.7980478	-73.9600437	D:\nyc_rav	TRUE
9	40.7973764	-73.9605349	D:\nyc_rav	TRUE
10	40.7986542	-73.9614745	D:\nyc_rav	TRUE
11	40.79872639999999	-73.9595472	D:\nyc_rav	TRUE
12	40.79797629999999	-73.96197	D:\nyc_rav	TRUE
13	40.7993297	-73.9609777	D:\nyc_rav	TRUE
14	40.7992437	-73.9628734	D:\nyc_rav	TRUE
15	40.79856809999999	-73.9633705	D:\nyc_rav	TRUE
16	40.80042949999999	-73.9656909	D:\nyc_rav	TRUE
17	40.80110810000001	-73.9651949	D:\nyc_rav	TRUE
21	40.80096679999999	-73.9690527	D:\nyc_rav	TRUE

node_id	aadt	start_street	end_street	y	x
39076490	2290.293032	DELAVERGNI	CEDAR DRIV	40.7624294	-73.757091
39076504	2326.697674	DELAVERGNI	N LYNDONVI	40.7534671	-73.744164
42421728	2369.850574	DUANE AVE |	CR 2 | GROV	40.7980478	-73.960044
42421731	4026.5	BURNSDALE	DYER AVE | C	40.7986542	-73.961475
42421737	10734	CATHEDRAL	COLUMBUS A	40.7992437	-73.962873
42421741	3505.666666	W 106TH ST |	CONVENT AV	40.8004295	-73.965691
42421745	2538.539007	CHURCH ST	CHURCH ST	40.8013981	-73.967996
42421749	4488	ALEXANDRIA	CLINTON VL	40.8016452	-73.968574
42421769	8076.125	BURNSDALE	DYER AVE | C	40.7826383	-73.975007
42421772	3722.5	BROADWAY	CONVENT AV	40.783828	-73.977825
42421775	2619.906666	CHURCH ST	CHURCH ST	40.7845691	-73.979583
42421776	3206.190476	CLINTON VL	CLINTON VL	40.7850352	-73.980673
42421778	3858	10TH AVE	12TH AVE	40.7855915	-73.982004

Step 5: Build the labels involving crash data and AADT

Highway Safety Manual (HSM) Critical Crash Rate method to evaluate traffic safety.

Calculates a statistical baseline threshold for each intersection by comparing its number of crashes to its traffic volume exposure over 6 years.

FINAL DATASET

Unnamed: 0.	Unnamed: 0	nearest_inte	intersection_	intersection_	image_path	image_exists	crash_score	num_crashes	final_risk_sc	risk_level	geometry	index_right	AADT	dist_feet	risk_per_car	refined_risk_level
8	8	8	40.7980478	-73.960044	D:\Abhineet\	TRUE	10	7	2.970630773	High	POINT (9953	85447	5542.666666	13.98247524	1.80418571	Medium
10	10	10	40.7986542	-73.961475	D:\Abhineet\	TRUE	6	4	2.414156868	High	POINT (9949	41689	5273.333333	21.63308531	1.137800252	Low
24	24	24	40.78196059	-73.975499	D:\Abhineet\	TRUE	8	10	1.918316218	Medium	POINT (9910	14922	12902.2	5.712799269	0.620049293	Low
33	33	33	40.78501835	-73.982749	D:\Abhineet\	TRUE	109	156	3.532889697	High	POINT (9890	48437	7118.571428	19.13641692	15.31206100	High
39	39	39	40.858744	-73.930122	D:\Abhineet\	TRUE	0	15	0	Low	POINT (1003	80744	652	6.547261254	0	Low
41	41	41	40.85954499	-73.927977	D:\Abhineet\	TRUE	0	3	0	Low	POINT (1004	38974	652	1.755431639	0	Low
47	47	47	40.78462299	-73.96985	D:\Abhineet\	TRUE	12	9	3.070113457	High	POINT (9925	17661	7368.833333	8.868725120	1.628480311	Medium
61	61	61	40.7248848	-73.97517	D:\Abhineet\	TRUE	48	21	7.065239893	High	POINT (9911	85344	4094.8	4.79682991	11.72218423	High
65	65	65	40.8133809	-73.956272	D:\Abhineet\	TRUE	200	173	5.964225779	High	POINT (9963	52053	8628	9.321602586	23.18034306	High

Step 5: Build the labels involving crash data and AADT

Highway Safety Manual (HSM) Critical Crash Rate method to evaluate traffic safety.
Calculates a statistical baseline threshold for each intersection by comparing its number of crashes to its traffic volume exposure over 6 years.

FINAL DATASET

Unnamed: 0.	Unnamed: 0	nearest_inte	intersection_	intersection_	image_path	image_exists	crash_score	num_crashes	final_risk_sc	risk_level	geometry	index_right	AADT	dist_feet	risk_per_car	refined_risk_level
8	8	8	40.7980478	-73.960044	D:\Abhineet\	TRUE	10	7	2.970630773	High	POINT (9953	85447	5542.666666	13.98247524	1.80418571	Medium
10	10	10	40.7986542	-73.961475	D:\Abhineet\	TRUE	6	4	2.414156868	High	POINT (9949	41689	5273.333333	21.63308531	1.137800252	Low
24	24	24	40.78196059	-73.975499	D:\Abhineet\	TRUE	8	10	1.918316218	Medium	POINT (9910	14922	12902.2	5.712799269	0.620049293	Low
33	33	33	40.78501835	-73.982749	D:\Abhineet\	TRUE	109	156	3.532889697	High	POINT (9890	48437	7118.571428	19.13641692	15.31206100	High
39	39	39	40.858744	-73.930122	D:\Abhineet\	TRUE	0	15	0	Low	POINT (1003	80744	652	6.547261254	0	Low
41	41	41	40.85954499	-73.927977	D:\Abhineet\	TRUE	0	3	0	Low	POINT (1004	38974	652	1.755431639	0	Low
47	47	47	40.78462299	-73.96985	D:\Abhineet\	TRUE	12	9	3.070113457	High	POINT (9925	17661	7368.833333	8.868725120	1.628480311	Medium
61	61	61	40.7248848	-73.97517	D:\Abhineet\	TRUE	48	21	7.065239893	High	POINT (9911	85344	4094.8	4.79682991	11.72218423	High
65	65	65	40.8133809	-73.956272	D:\Abhineet\	TRUE	200	173	5.964225779	High	POINT (9963	52053	8628	9.321602586	23.18034306	High

METHODOLOGY

EXPLORING WITH DIFFERENT NETWORKS

- ResNet 50 - Trained head, finetuned layer 4 + head
- Best recall: 0.4257
- Unfroze layer 3 and 4
- Best recall: 0.4441

- No transforms:
- Head, layer 3, layer 4
- Best recall: 0.4511

- Weighted random sampler
- Saw no improvement

OVERALL: Highest accuracy 0.495

EXPLORING WITH DIFFERENT NETWORKS

ConvNext tiny with all unfrozen layers

0.4788

- Froze stem in early blocks
- 0.4774, Accuracy improved slightly with 0.5407

- Attempted a high drop path rate
- Higher learning rate for head than stage 3
- 0.484 for recall, 0.5177 accuracy

ConvNext tiny

We trained ConvNext model with and without preprocessing the images as well.
We achieved overall higher accuracy without preprocessing the images, but this could also be due to the model learning geographical features.

PERFORMANCE METRICS AND DEPLOYABILITY OF SOLUTION

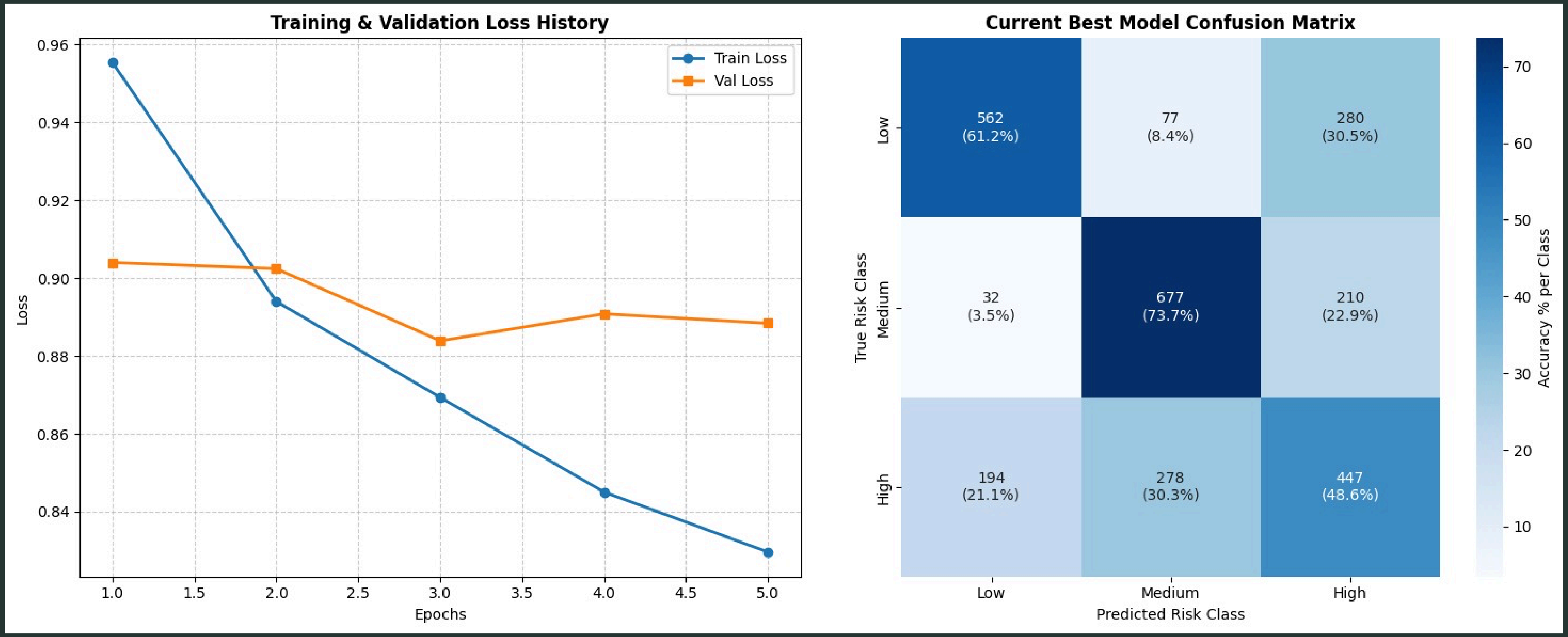
ConvNext Tiny

Final training and validation accuracy: 63%

Final test accuracy: 61%

Recall: 0.58+0.2

Number of epochs: 5



Deployability & Scalability

Can the solution be deployed at Plaksha?

- Direct deployment at Plaksha University is currently limited because the campus does not contain large-scale urban road intersections required for the proposed analysis.

Potential Real-World Deployment :

- The framework is highly suitable for deployment in dense urban traffic environments such as New York City, especially in boroughs like Manhattan, Brooklyn, and Queens, where complex intersection architectures and heavy traffic exposure exist.
- The framework can also be extended to similar high-density urban traffic systems in cities.
- Transportation authorities can use the Safety Index to identify high-risk intersections proactively and prioritise infrastructure interventions.

Challenges During Scale-Up :

- Availability and consistency of large-scale traffic and crash datasets.
- Variations in traffic patterns and road infrastructure across cities.
- Need for continuous recalibration as traffic conditions evolve.
- Increased computational and preprocessing requirements for city-scale deployment.