Investigation of US Traffic Accidents and Prediction of Accident Severity

Classification using Neural Network

Wei Zhao

Background

- Leading cause of non-natural death for U.S. citizens.
- The US suffers from the most road crash deaths among high-income countries.
- Urgent to understand the underlying mechanisms of the occurrence of traffic accidents.

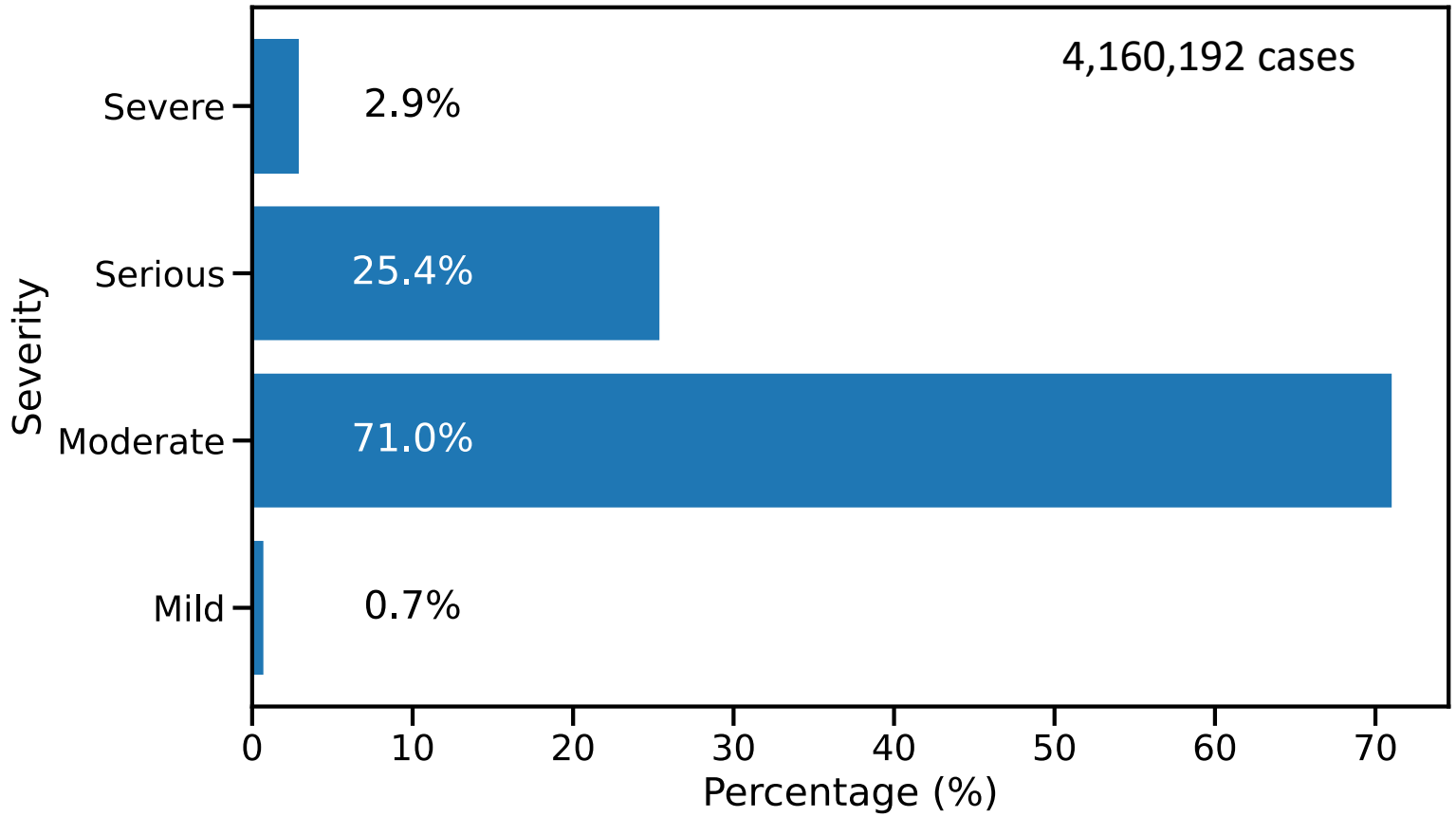
Aims:

- Investigate the occurrence of accidents.
- Build a neural network for instantaneous prediction of accident severity.

Dataset

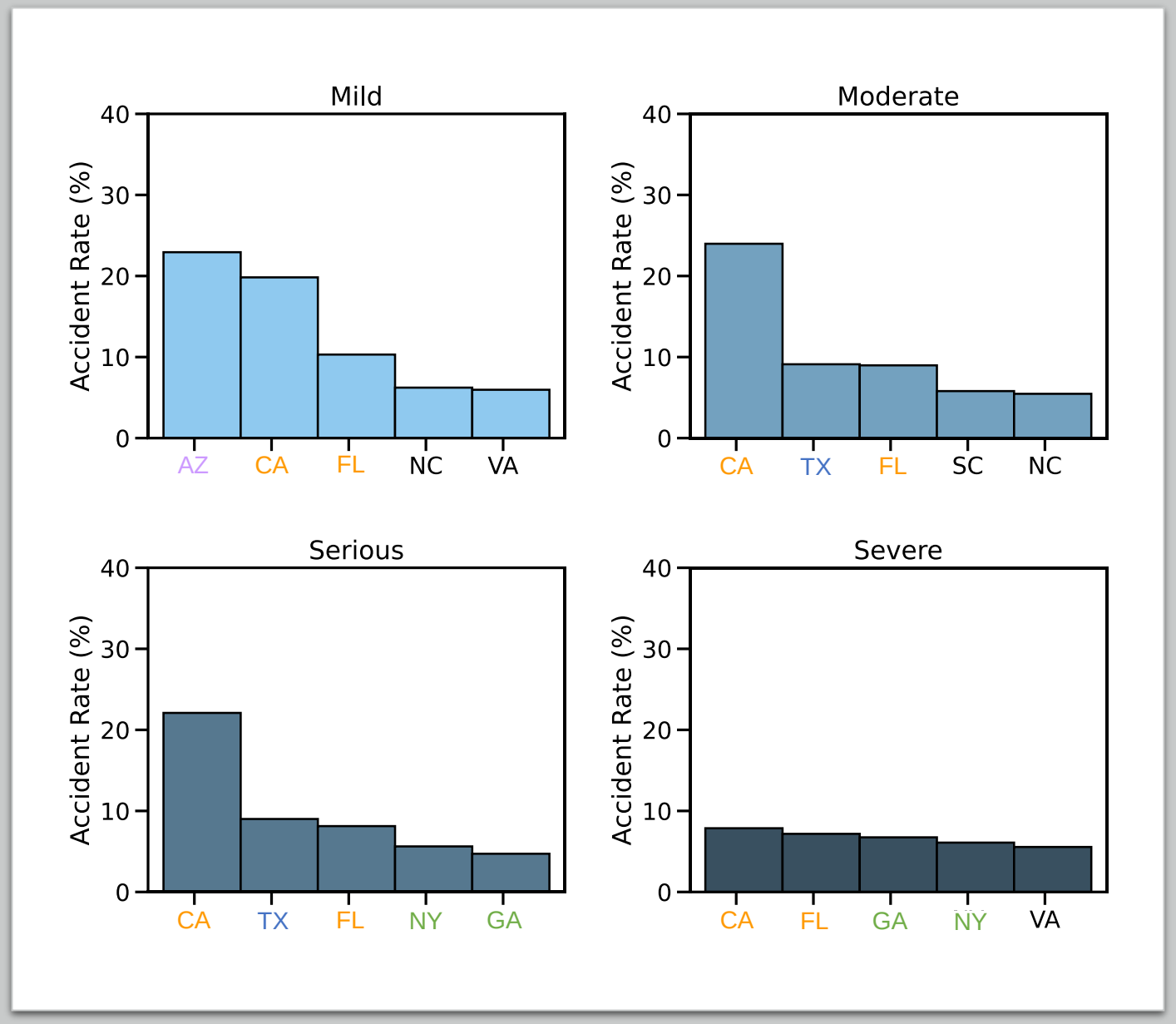
- **Features**:
  - Time
  - Locations
  - Weather conditions
  - Point of interest
- **Labels**:
  - Severity:
    - Mild, Moderate, Serious, and Severe
- **A total of ~4M cases in the clean data**

[1] Moosavi et al., 2019, arXiv preprint arXiv:1906.05409; [2] Moosavi et al., 2019, Proceedings of the 27th ACM SIGSPATIAL, ACM



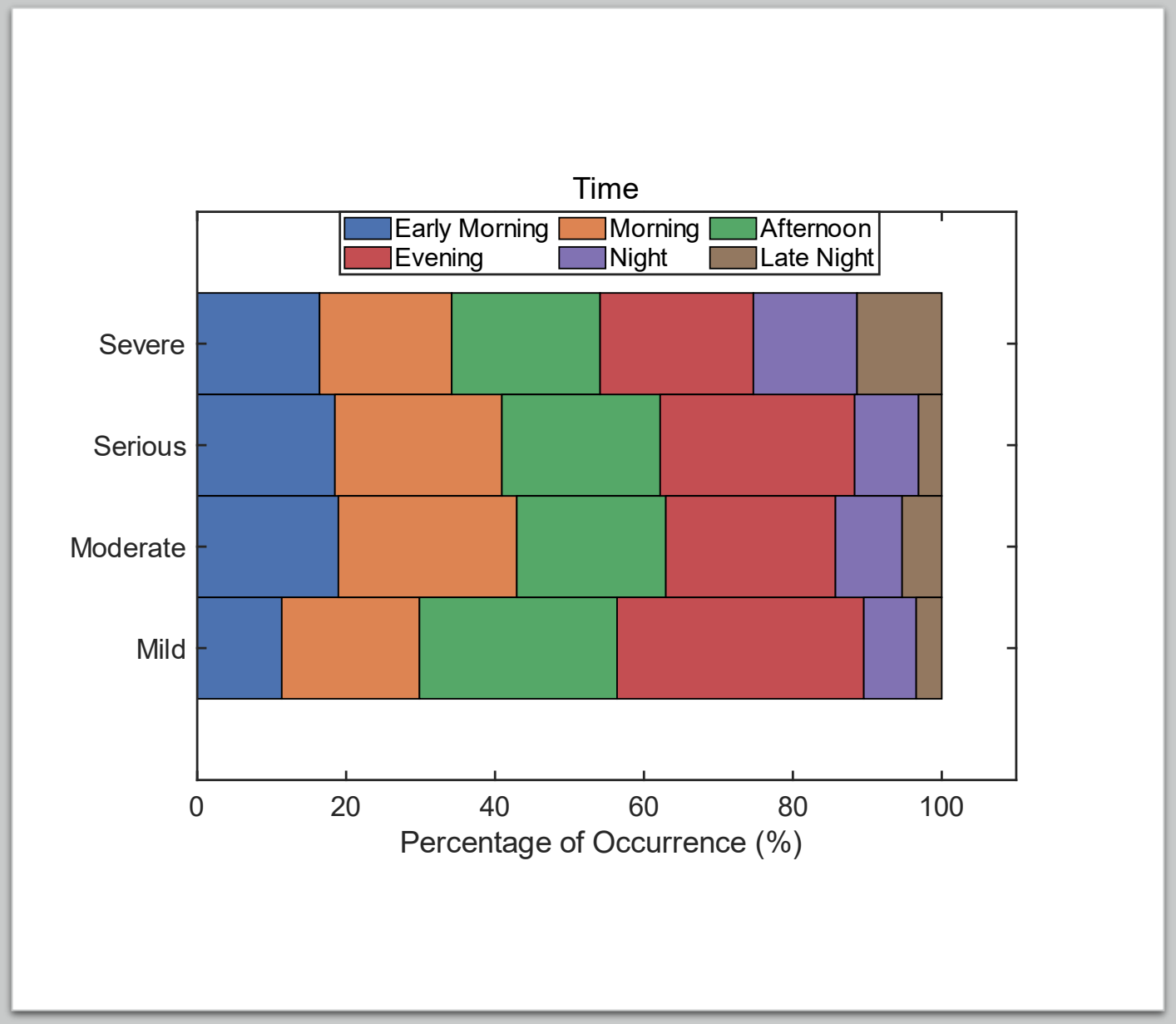
Dataset

Highly imbalanced data needs an extra attention when designing the neural network.



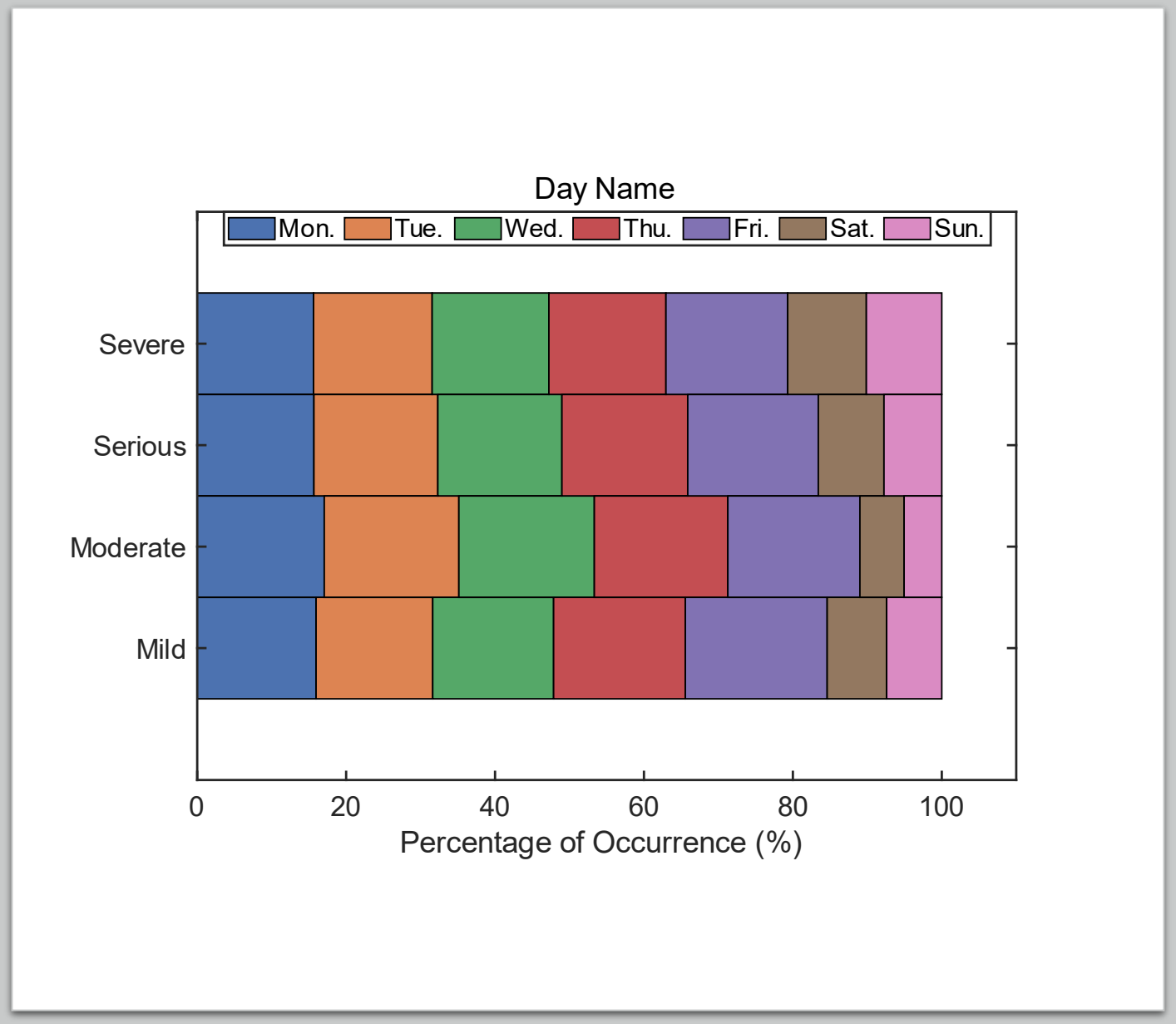
Accident Rate by State

- Most mild accidents occurred in Arizona.
- CA and FL:
  - All severity levels
- TX:
  - Moderate to serious
- GA and NY:
  - Serious to severe



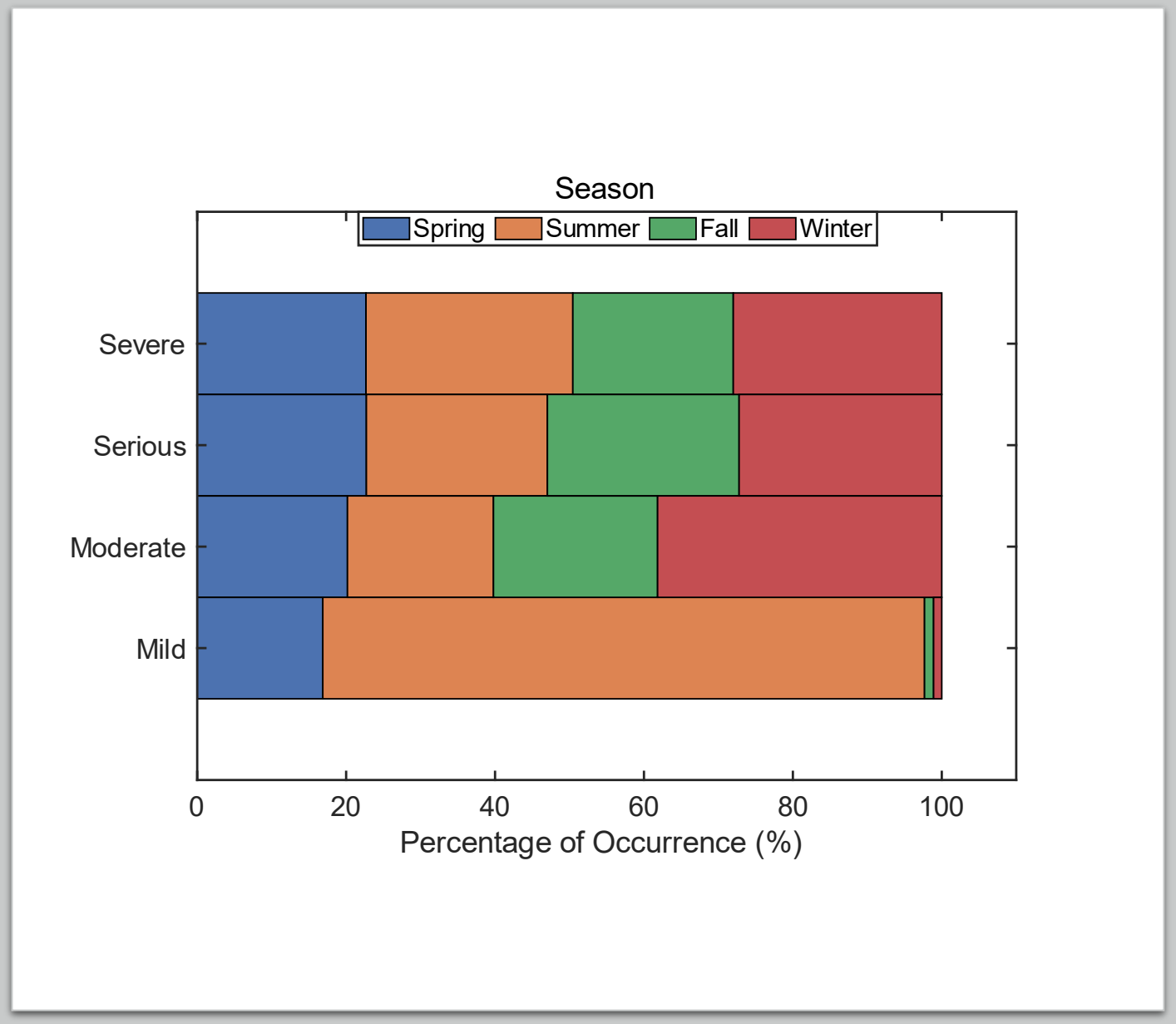
Percentage of Occurrence by Time

- Morning and evening:
  - Most frequently overall
- Night and late night:
  - Most likely severe



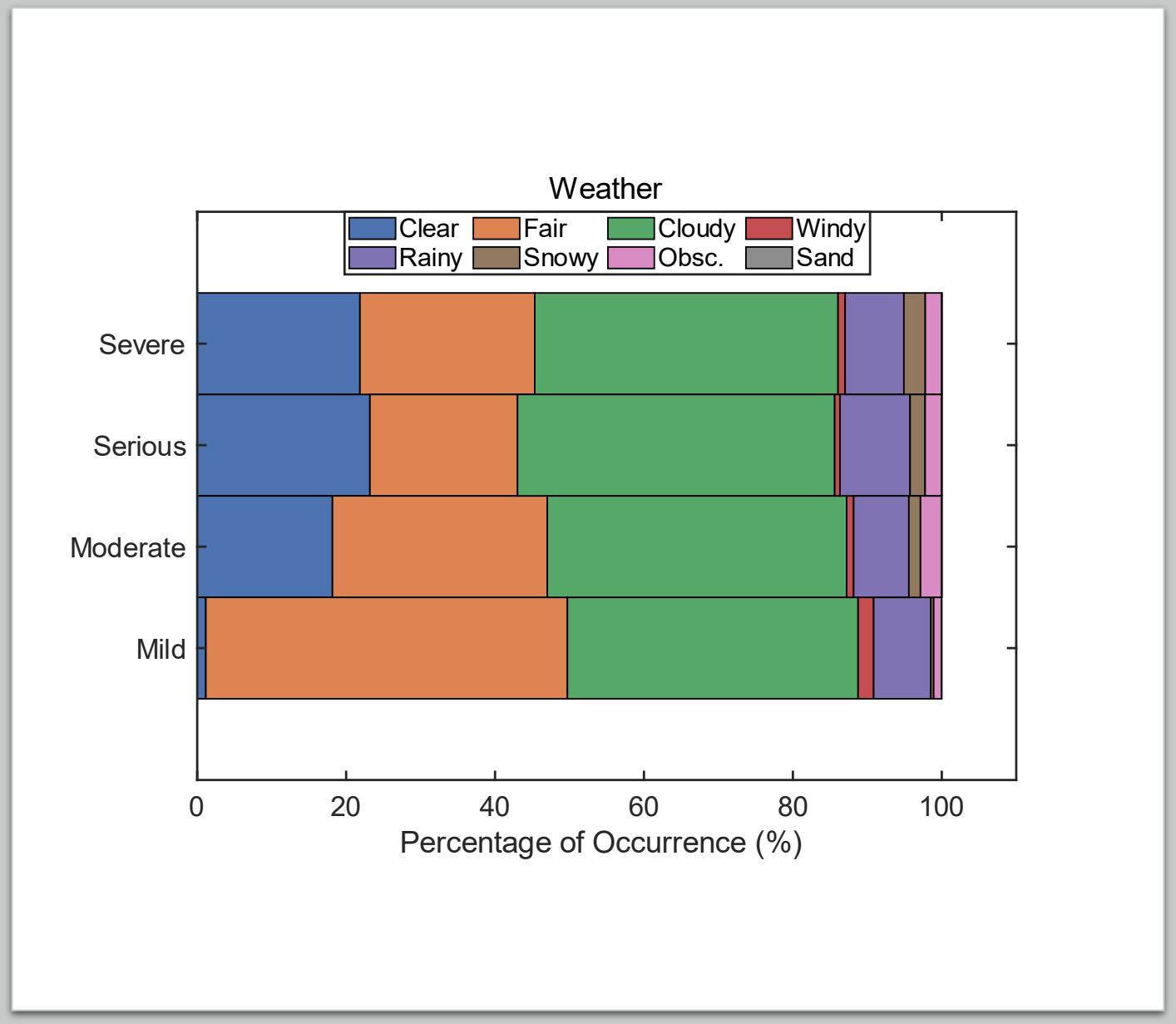
Percentage of Occurrence by Day Name

- Weekdays:
  - Most frequently
- Weekends:
  - More likely severe



Percentage of Occurrence by Season

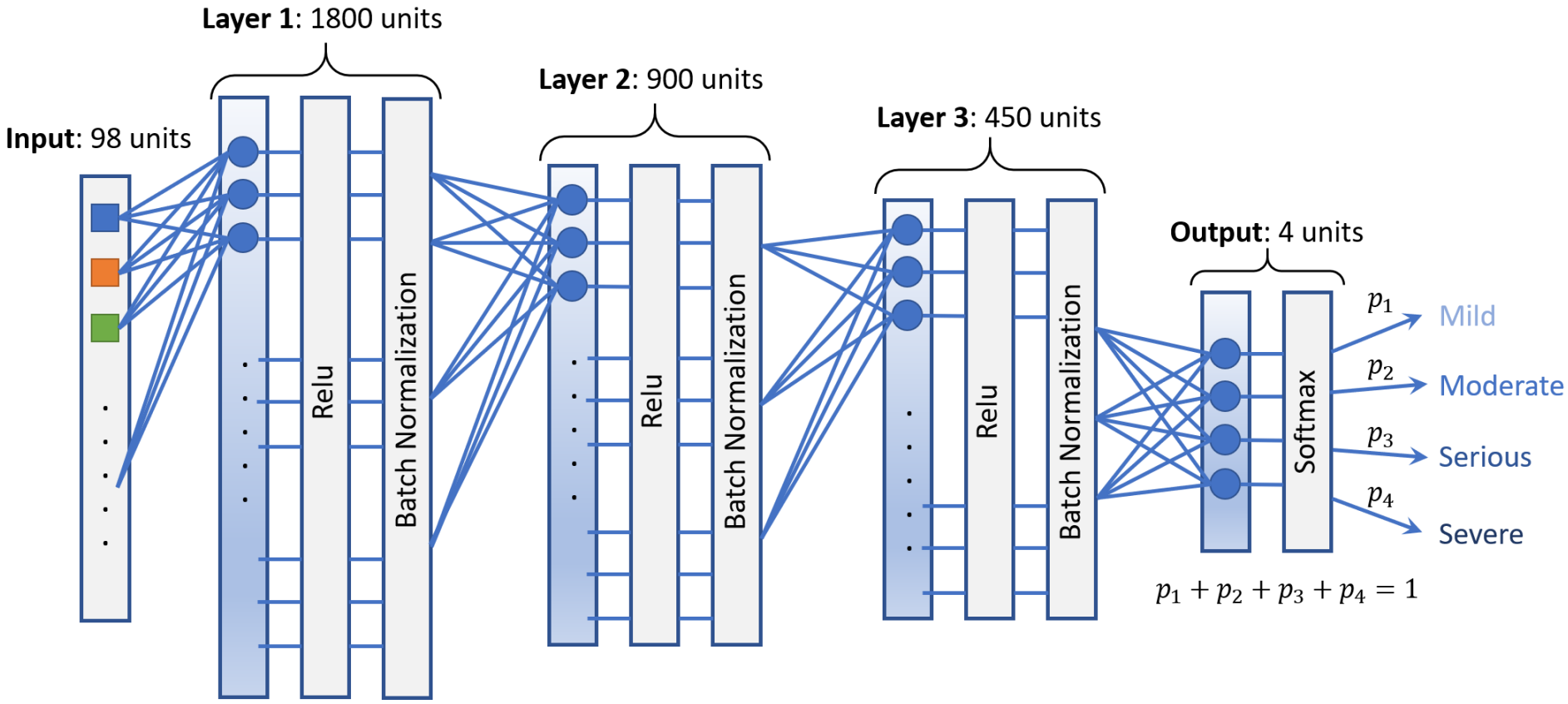
- Summer:
  - Mostly mild accidents
- Other seasons:
  - Most moderate to severe accidents in winter



Percentage of Occurrence by Weather

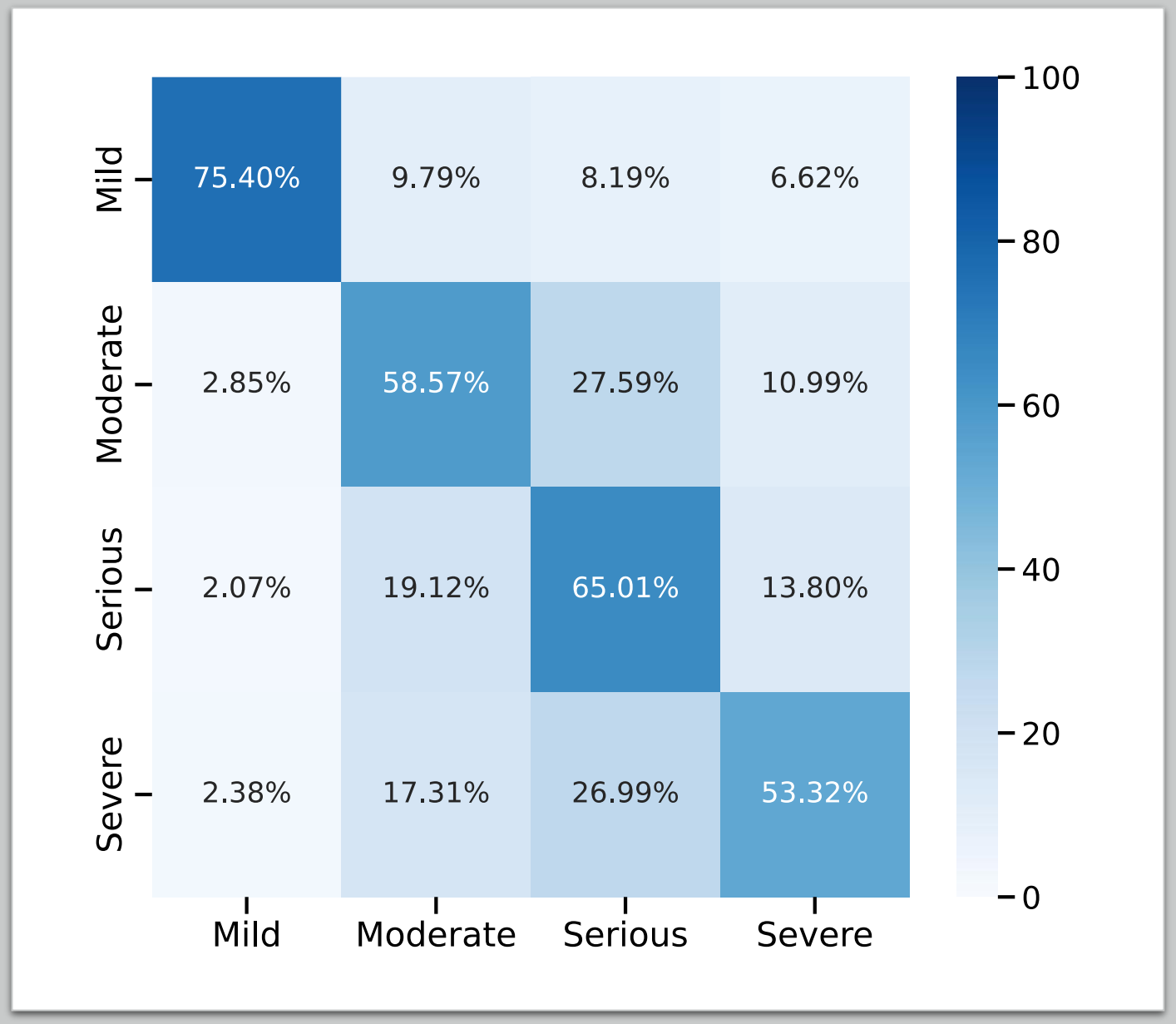
- Mostly occur in nice weather
  - Clear, fair, cloudy
- Windy/Rainy/Snowy/Obscure
  - Mostly serious

Neural Network Architecture



- Batch size = 1024
- Learning rate = 1e-6
- Number of epochs = 50
- Monitor='val_loss'
- Early stopping: Patience of 10 epochs
- Weighting for class imbalance

[1] Holden et al., 2006, Science ; [2] Cai et al., 2018, PLoS one



Confusion Matrix & Recall Scores

	Recall
Mild	0.75
Moderate	0.59
Serious	0.65
Severe	0.53
Average	0.63

Summary and Future work

- The occurrence of accidents depends on locations.
  - CA and FL have top accident rates with all severities.
- The occurrence of accidents depends on time, name of day, season, and weather.
- Built a neural network for instantaneous prediction of accident severity.

Future work may include:

- NLP for accident description to enhance the predictive power of the neural network
- Instantaneous prediction of the traffic delay and distance affected by accidents.

Email

weizhao1001
@gmail.com

Thank you !

GitHub

github.com/
weizhao-BME

LinkedIn

linkedin.com/in/
weizhao-bme/

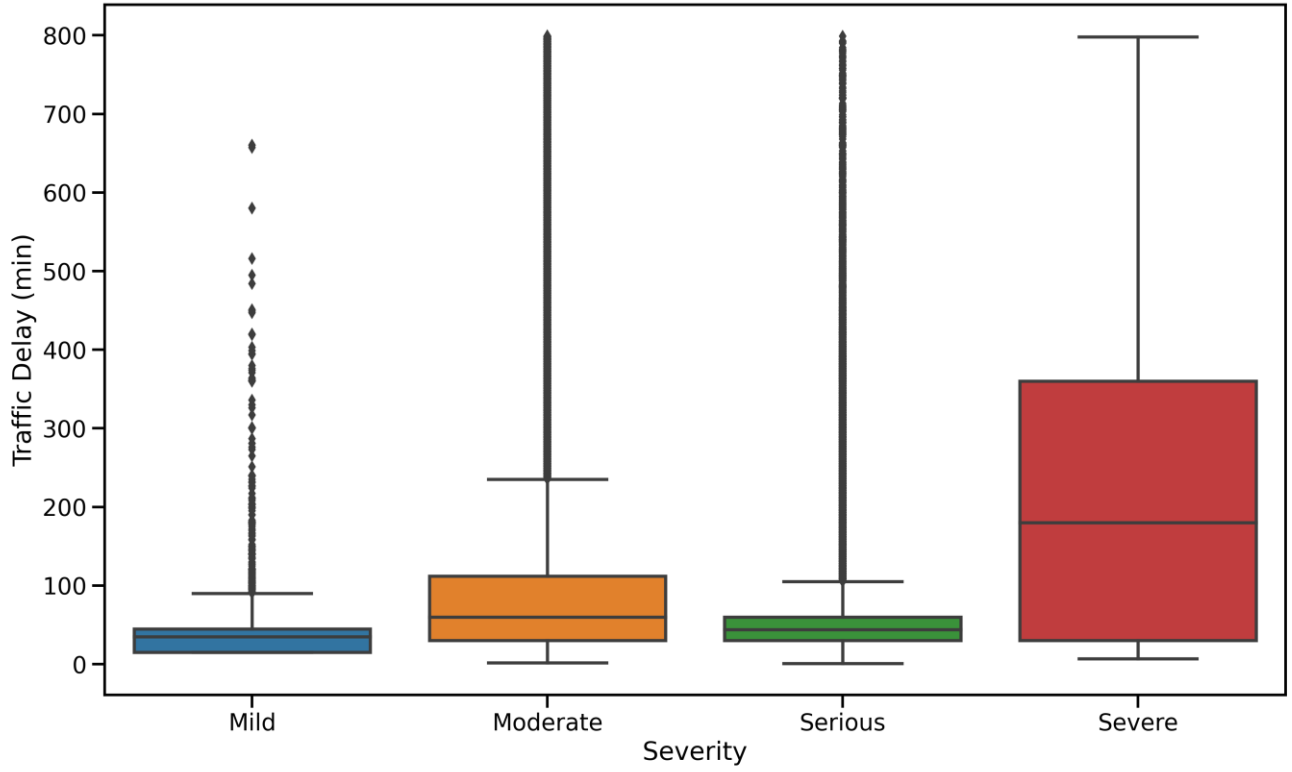
Blog

weizhao-bme.medium.com

Google Scholar

scholar.google.co
m/citations?user
=vlOb7iIAAAAJ&
hl=en

Appendix – Traffic Time Delay



Appendix – Distance Affected by Traffic