Graph Neural Networks for Road Safety Modeling

EVALUATIONS OF SLOPE INCORPORATION IN THE GNN PREDICTION ABILITY FOR THE STATE OF MONTANA, US

- Claudio Giannini
- Ehsan Mokhtari
- Arash Bakhshaee Babaroud
- Arian Gharehmohammadzadeghashghaei

- Advanced Machine Learning Final Project
- Prof: Fabio Galasso
- Sapienza University of Rome
- 2024-2025

TASK AND MOTIVATION

Task

- Use Graph Neural Networks (GNNs) to predict traffic accidents, uncovering spatial temporal patterns to improve road safety and urban planning.
- Gathering road slope data for each edge and using as feature in training GNN
- Previous Task: Perform the predictions at an Urban Level

Motivation

Traffic accidents are a global risk, and understanding their causes is vital for safer cities. Recent studies show GNNs excel at modeling road networks, providing accurate predictions and actionable insights for safety improvements

Related Works

NEURAL INFORMATION PROCESSING SYSTEMS

- Graph Neural Networks for Road Safety Modeling: Datasets and Evaluations for Accident Analysis (Nippani et al., 2023)
- Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction (Xiaowei Gao et al., 2024)
- Investigation of the Effect of Slope and Road Surface Conditions on Traffic Accidents Occurring in Winter Months: Spatial and Machine Learning Approaches (E. Kuşkapan, M.Y. Çodur, M.A. Sahraei, 2024)

TASK AND MOTIVATION

Task

- Use Graph Neural Networks (GNNs) to predict traffic accidents, uncovering spatial temporal patterns to improve road safety and urban planning.
- Gathering road slope data for each edge and using as feature in training GNN
- Previous Task: Perform the predictions at an Urban Level
- NEW! New Task: Evaluate whether incorporating slope data enhances prediction accuracy

Motivation

Traffic accidents are a global risk, and understanding their causes is vital for safer cities. Recent studies show GNNs excel at modeling road networks, providing accurate predictions and actionable insights for safety improvements

Related Works

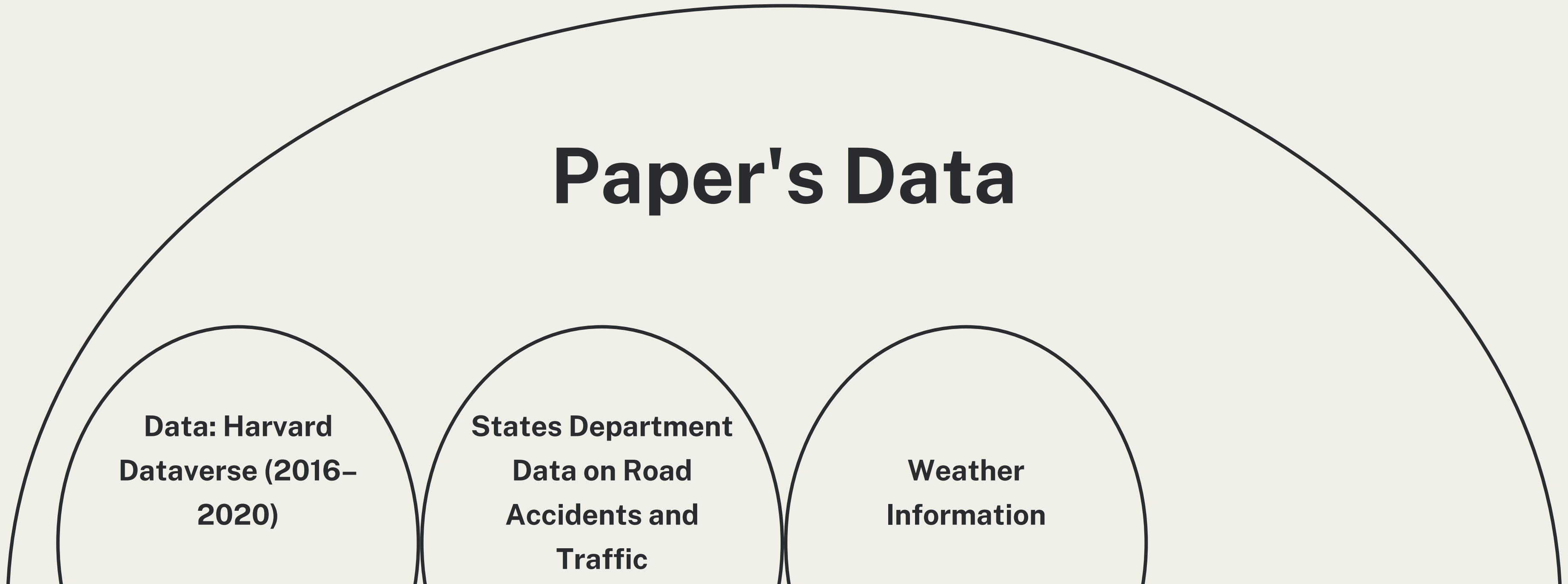
NEURAL INFORMATION PROCESSING SYSTEMS

- Graph Neural Networks for Road Safety Modeling: Datasets and Evaluations for Accident Analysis (Nippani et al., 2023)
- Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction (Xiaowei Gao et al., 2024)
- Investigation of the Effect of Slope and Road Surface Conditions on Traffic Accidents Occurring in Winter Months: Spatial and Machine Learning Approaches (E. Kuşkapan, M.Y. Çodur, M.A. Sahraei, 2024)

DATA & SLOPE INTEGRATION

“Base” Data

- Harvard Dataverse (2016–2020) for Montana (MT)
- Give the true train/val/test split: 2016-2017, 2018, 2019-2020 which in order to replicate exactly their study
- Number training examples: 40,040; 20,677; 39,222



Node-Level Features:

- Latitude, Longitude
- Node indegree and outdegree
- Betweenness centrality
- Surface temperatures: Average (tavg), Max (tmax), Min (tmin)
- Total precipitation (prcp), Average wind speed (wspd), Sea level air pressure (pres)

- Edge-Level Features:

- Binary label: One-way or not
- Multi-class label: Road type (highway, residential, etc.)
- Length of road
- Annual average daily traffic (AADT)

DATA & SLOPE INTEGRATION

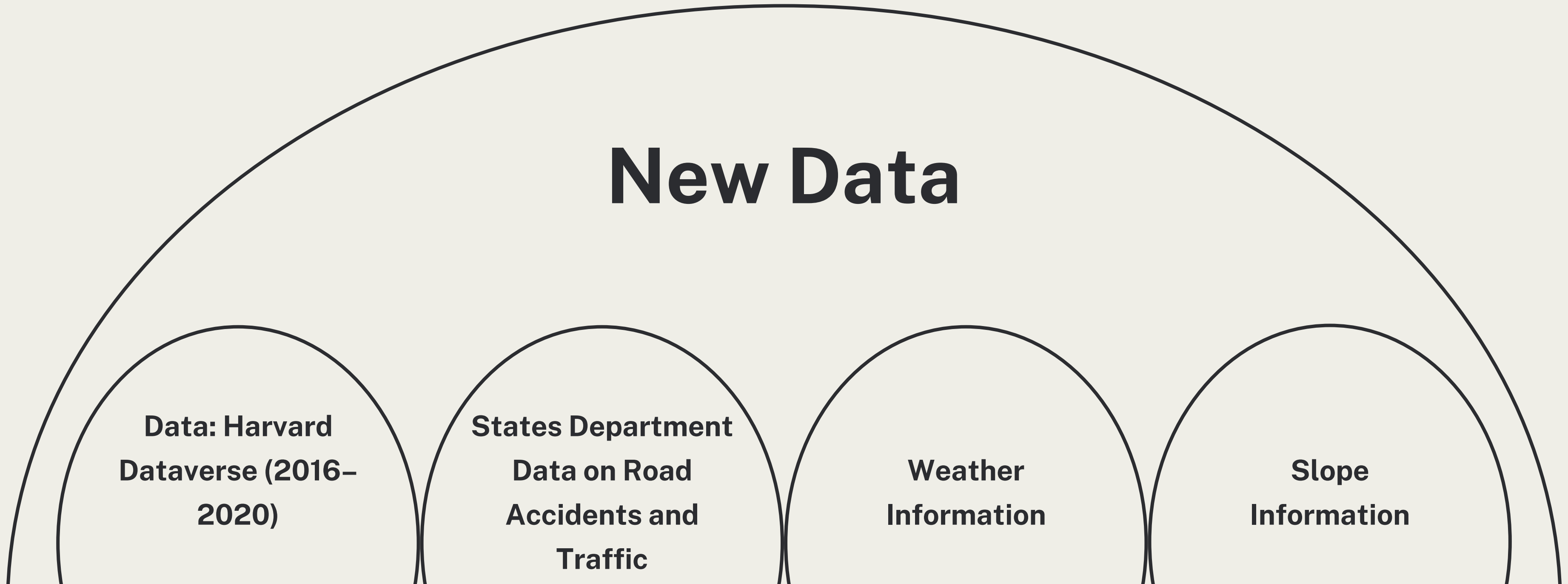
“Base” Data

- Harvard Dataverse (2016–2020) for Montana (MT)
- Give the true train/val/test split: 2016-2017, 2018, 2019-2020 which in order to replicate exactly their study
- Number training examples: 40,040; 20,677; 39,222

+

“Slope” Data

- Elevation for each node obtained through Open-Elevation API
- Slope Calculation:
  - elevation difference (target - source)/ length of edge
- Noise Reduction
- New Slope Tensor in edge_features.pt for MT



Node-Level Features:

- Latitude, Longitude
- Node indegree and outdegree
- Betweenness centrality
- Surface temperatures: Average (tavg), Max (tmax), Min (tmin)
- Total precipitation (prcp), Average wind speed (wspd), Sea level air pressure (pres)
- Elevation (m)

- Edge-Level Features:

- Binary label: One-way or not
- Multi-class label: Road type (highway, residential, etc.)
- Length of road
- Annual average daily traffic (AADT)
- Slope

SLOPE DATA AND CHALLENGES

- **<u>Cause of Noise:</u>** We used the **Open-Elevation API** for slope calculations, which, while free, struggles in mountainous terrains where elevation changes sharply over short distances. This limitation affects slope accuracy but performs better in urban areas.

- **<u>Noise Reduction:</u>** To address this, we applied thresholds to filter unrealistic slopes: values above +-0.3 were capped at +-0.05, and slopes over +-0.2 with road lengths under 50m were set to 0.

- **<u>Solution:</u>** A potential improvement could involve using the **Google Maps API** for higher precision, though it involves additional costs. These adjustments aim to balance accuracy and practicality in our study.

MODEL & APPROACH

Model

In this study, we focus on using a **Graph Convolutional Network (GCN)** with 2 graph convolutional layers to capture dependencies between nodes and edges to model traffic data.

Hyperparameters:

- Hidden Dimensionality: 256
- Learning Rate: 0.001
- Epochs: 100
- Optimizer: Adam

MODEL & APPROACH

Model

In this study, we focus on using a **Graph Convolutional Network (GCN)** with 2 graph convolutional layers to capture dependencies between nodes and edges to model traffic data.

Hyperparameters:

- Hidden Dimensionality: 256
- Learning Rate: 0.001
- Epochs: 100
- Optimizer: Adam

But How Did We Decide on These HyperParameter?

M O D E L & A P P R O A C H

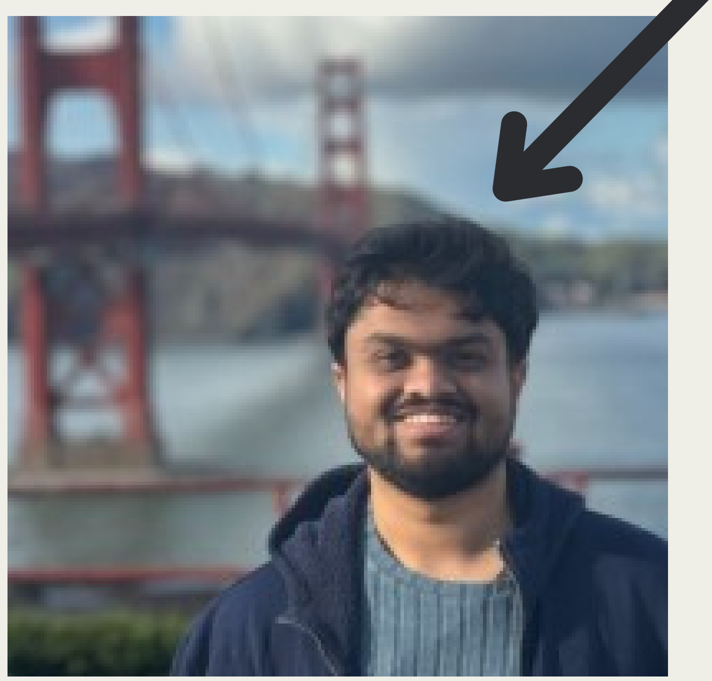
Model

In this study, we focus on using a **Graph Convolutional Network (GCN)** with 2 graph convolutional layers to capture dependencies between nodes and edges to model traffic data.

Hyperparameters:

- Hidden Dimensionality: 256
- Learning Rate: 0.001
- Epochs: 100
- Optimizer: Adam

He Said So!



PhD A. Nippani

“

Number of layers	2	3	4
	85.2±0.1	84.9±0.3	84.4±0.4
Hidden dimensionality	128	256	512
	84.5±0.4	**85.2**±0.1	84.5±0.5
Learning rate	$1e^{-2}$	$1e^{-3}$	$1e^{-4}$
	85.0±0.7	**85.2**±0.1	84.0±0.5
Epochs	50	100	200
	84.0±0.2	**85.2**±0.1	**85.2**±0.3

“

ANALYSIS OF RESULTS

- **ROC-AUC:** 82.62% (Baseline) vs. 83.22% (with Slope)
  - +60bps with Slope, meaning that it enhances the model's ability to distinguish between high and low-risk edges
- **Recall:** 52.01% (Baseline) vs. 47.14% (with Slope)
  - Slope leads to a slight drop in capturing true positives (accidents).
- **Precision:** 5.26% (Baseline) vs. 8.13% (with Slope)
  - Fewer false positives, making the model more reliable for predicting accidents

Conclusion:

- For the benchmark paper's analysis, the model's performance is evaluated using ROC-AUC, and with the addition of slope, we have observed a slight improvement in this metric, indicating enhanced ranking ability of accident-prone and non-accident-prone road segments. However, adding slope also improves precision, reducing false positives, but this comes at the cost of a trade-off with recall, as fewer true positive accidents are captured.

Metric	**Train**	**Validation**	**Test**
ROC-AUC	82.99%	82.73%	82.62%
Recall	55.46%	55.96%	52.01%
Precision	8.29%	6.46%	5.26%

Table 1: Baseline Results (Without Slope)

Metric	**Train**	**Validation**	**Test**
ROC-AUC	83.31%	83.03%	83.22%
Recall	44.35%	46.79%	47.14%
Precision	11.89%	8.96%	8.13%

Table 2 : Results With Slope

FINAL CONCLUSION

Comparison of Results: Adding slope as a feature improved the model's performance slightly, increasing ROC-AUC from 82.62% to 83.22% and significantly enhancing precision (from 5.26% to 8.13%). However, recall decreased slightly, indicating a trade-off between identifying true positives and avoiding false positives. These mixed effects suggest that further optimization of the model are necessary to improve overall prediction accuracy and balance precision and recall more effectively.

Future Improvements:
- Use more precise elevation data sources like Google Maps API to improve slope calculations.
- Experiment with categorical encoding for slope (e.g., flat, mild, steep) to reduce noise and better represent inclines.
- Test the model in diverse states

Final Thoughts: While slope integration demonstrates clear potential, refining data quality and feature representation will be critical for achieving more robust and actionable predictions in future projects.

REFERENCES

- Nippani, A., Li, D., Ju, H., Koutsopoulos, H. N., & Zhang, H. R. (2023). Graph Neural Networks for Road Safety Modeling: Datasets and Evaluations for Accident Analysis. Presented at NeurIPS 2023. Northeastern University, Boston. Retrieved from https://neurips.cc/virtual/2023/poster/73579

- Gao, X., Jiang, X., Haworth, J., Zhuang, D., Wang, S., Chen, H., & Law, S. (2024). Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. SpaceTimeLab, University College London (UCL), Peking University (PKU), MIT, University of Florida, and The Bartlett Centre for Advanced Spatial Analysis, UCL.

- Kuşkapan, E.; Çodur, M.Y.; Sahraei, M.A. Investigation of the Effect of Slope and Road Surface Conditions on Traffic Accidents Occurring in Winter Months: Spatial and Machine Learning Approaches. Appl. Sci. 2024, 14, 11629. https://doi.org/10.3390/app142411629

Thank you!