Will there be a construction? Predicting road constructions based on heterogeneous spatiotemporal data

Amin Karimi Monsefi[†], Sobhan Moosavi[§], and Rajiv Ramnath[†]

†THE OHIO STATE UNIVERSITY, §LYFT INC.

Introduction

- Road constructions are essential to transportation infrastructures.

- Recently released data by the United States Census Bureau showed that the annual value of road constructions increased from **87.9 billion dollars in 2017** to **100.4 billion dollars in 2021** - an over 18% increase in just five years.

- Deciding what the next construction project is and when it is to be scheduled is traditionally done through inspection by humans using special equipment. This approach is expensive and limited in coverage.

- Determining when and where road constructions are needed via computational solutions is a relatively new, less-explored research area.

- This research reports on an approach that uses a deep neural network model to predict future constructions.

- We also introduce a unique dataset of **6.2 million road constructions** in the United States between **2016 and 2021**.

Contribution

- **Dataset:** We introduce a new dataset of road constructions and closures for the continental United States, with about **6.2 million cases from 2016 to 2021**. To our knowledge, this is the first public dataset that offers this type and scale of data.

- **Insights:** We glean various insights by analyzing the **US-Constructions dataset**. We detail these insights to inspire other researchers to use our data for different applications, especially those aimed at enhancing transportation infrastructures and their safety.

- **Model:** We present a deep neural network model to **predict short-term constructions**. Our model can use heterogeneous data and results in superior prediction outcomes compared to several state-of-the-art traditional and deep-learning models.

US-Constructions Dataset

- This dataset is a countrywide dataset of road construction and closure events, which covers **49 states of the US**. Construction events in this dataset could be any roadwork, ranging from fixing pavements to substantial projects that could take months to finish.


Traffic Data Collection / Data Augmentation
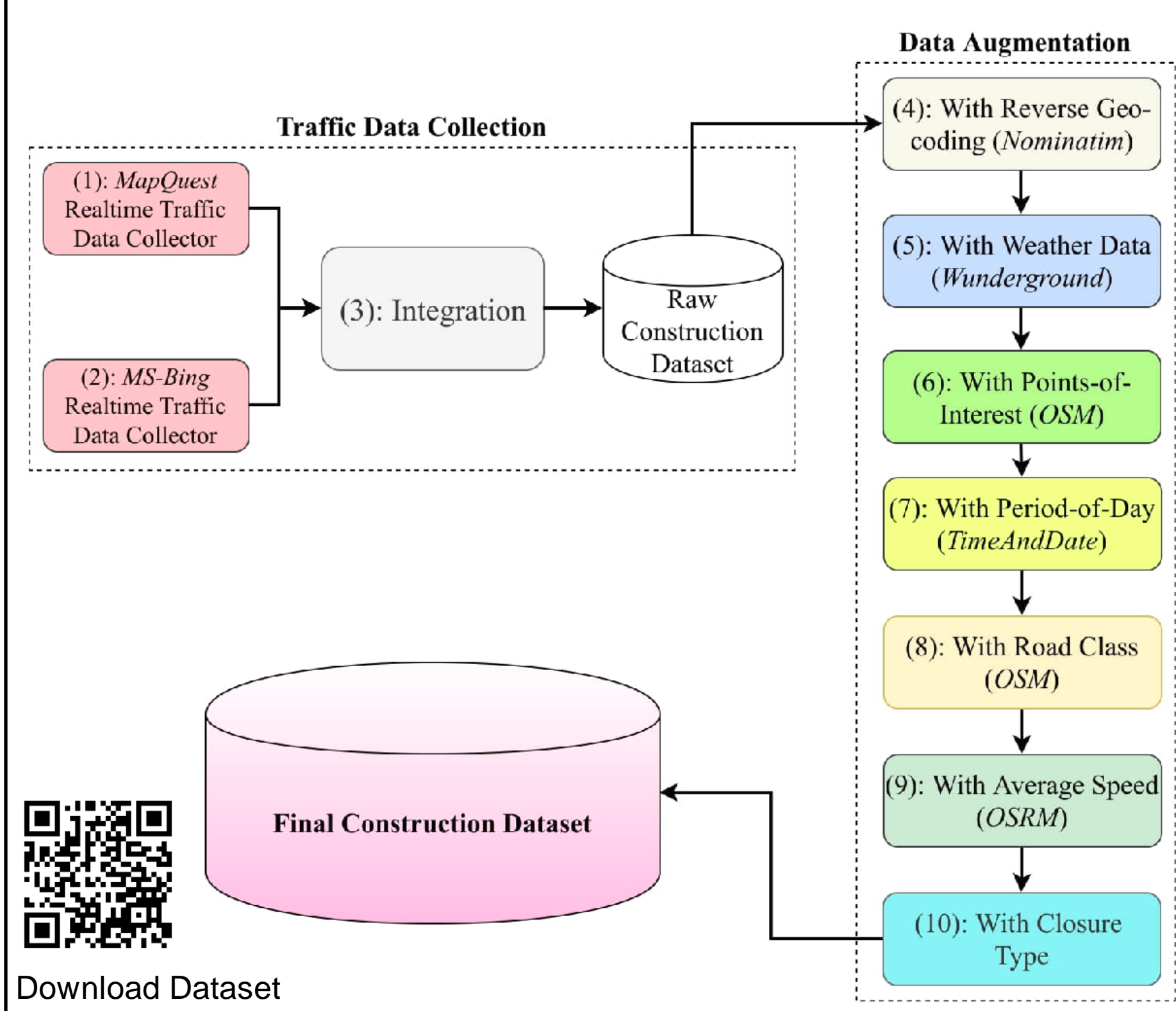
US-Constructions Dataset

Process of data building.

- **Collecting Raw Construction Data:**
  - We collected streaming traffic data from two real-time data providers, **MapQuest Traffic** and **Microsoft Bing Maps**.

Top 10 States in US-Constructions Dataset

- **Augmentation with Reverse Geocoding:**
  - Translating raw location data to addresses with elements such as **street number**, **street name**, **city**, **state**, and **zip code**.

- **Augmentation with Weather Data:**
  - Raw weather data was collected from 2072 airport weather stations around the US.
  - Several attributes such as **temperature**, **humidity**, **wind speed**, **pressure**, **precipitation**, and **condition**.

- **Augmentation with POI Annotation:**
  - Points-of-interest (POI) are locations annotated on a map as **amenities**, **traffic signals**,

Distribution of constructions (in %) near each POI
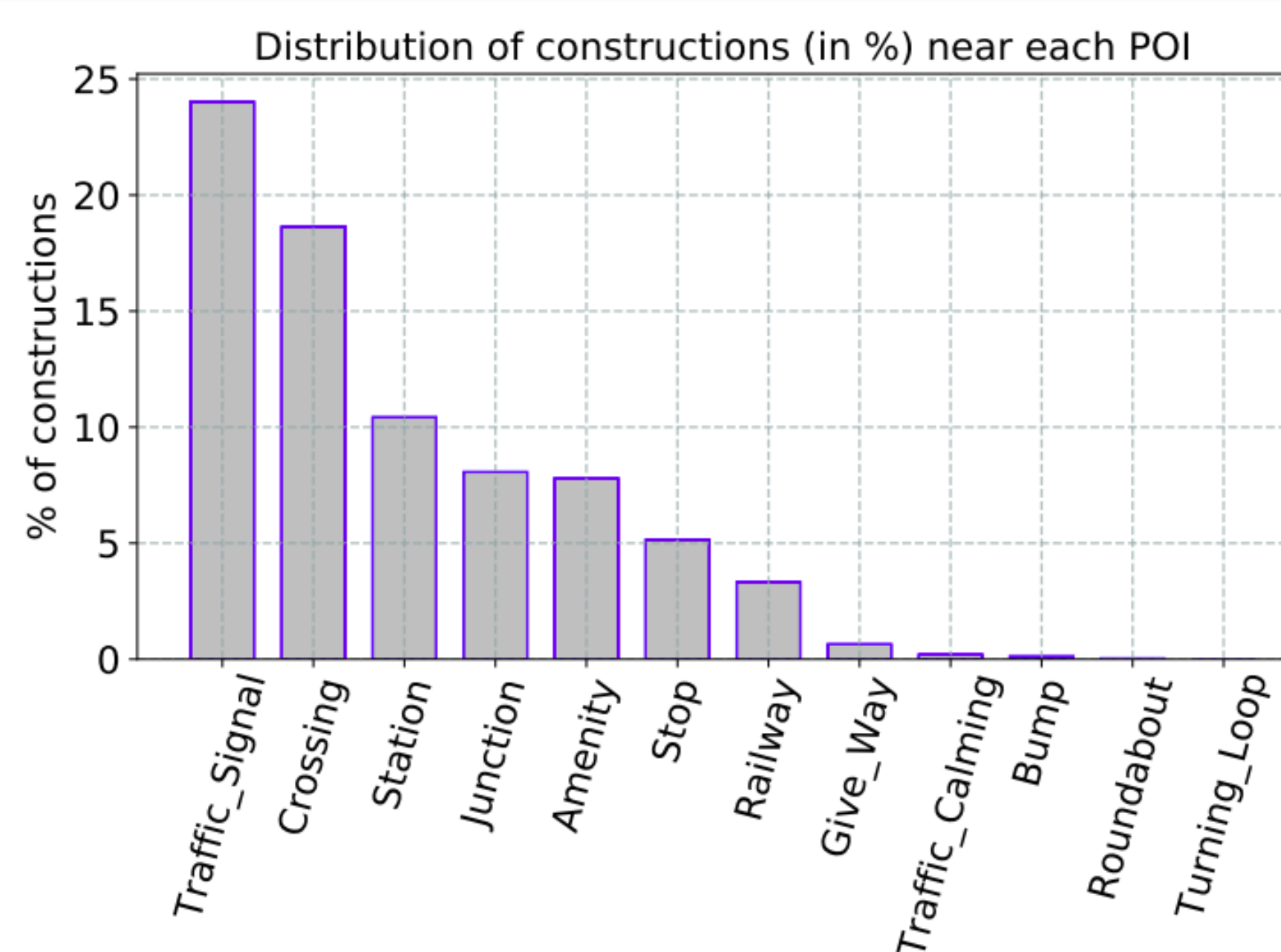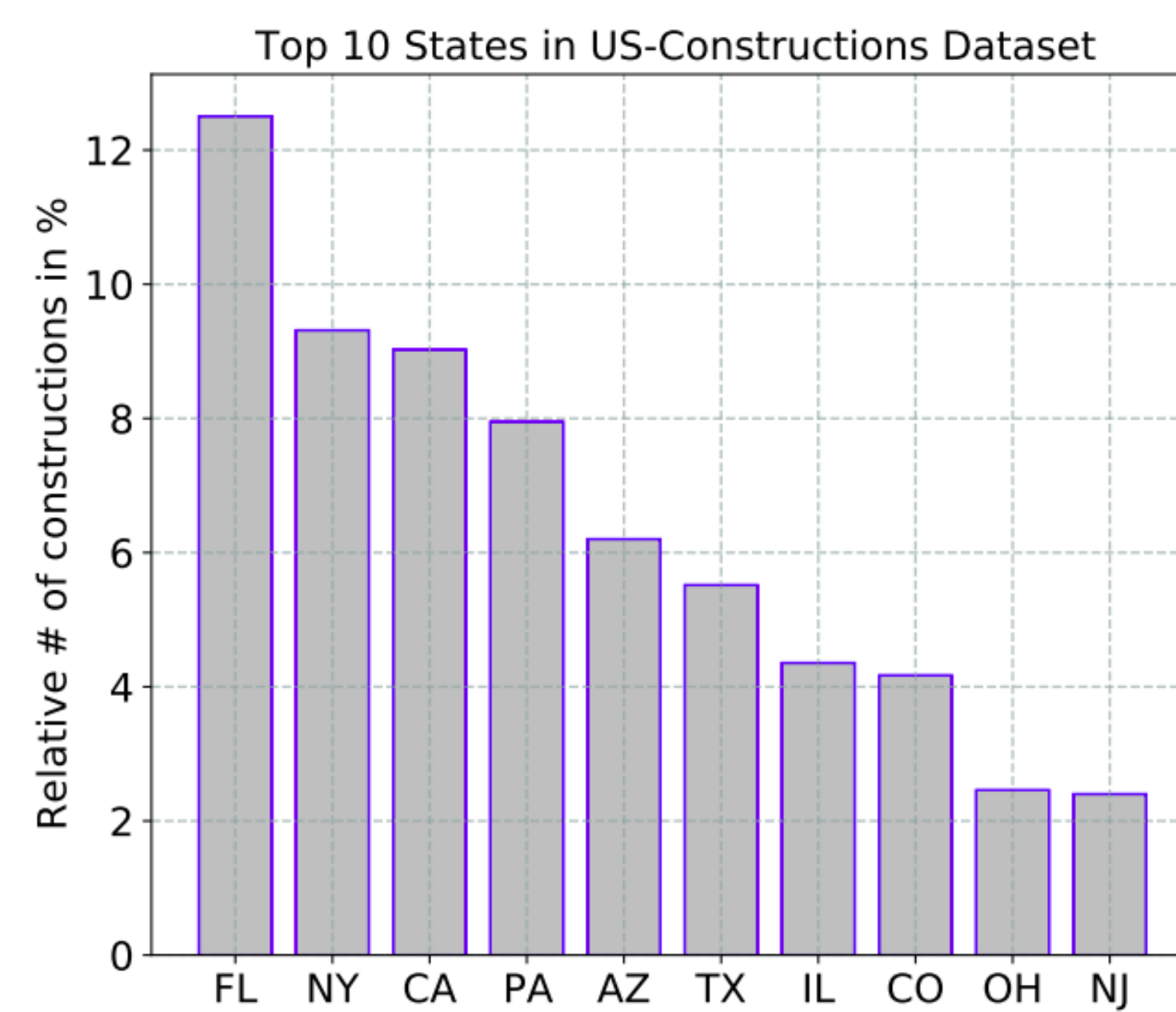
- **Augmentation with Period of Day:**
  - Based on the start time of a construction record, We assign a label to record based on four different daylight systems, namely **Sunrise/Sunset**, **Civil Twilight**, **Nautical Twilight**, and **Astronomical Twilight**.

- **Augmentation with Road Class:**
  - We used OSM to obtain this information and adopted its road classification system. We define an algorithm to find road class.

Top 10 road types (excluded "road_type=None")
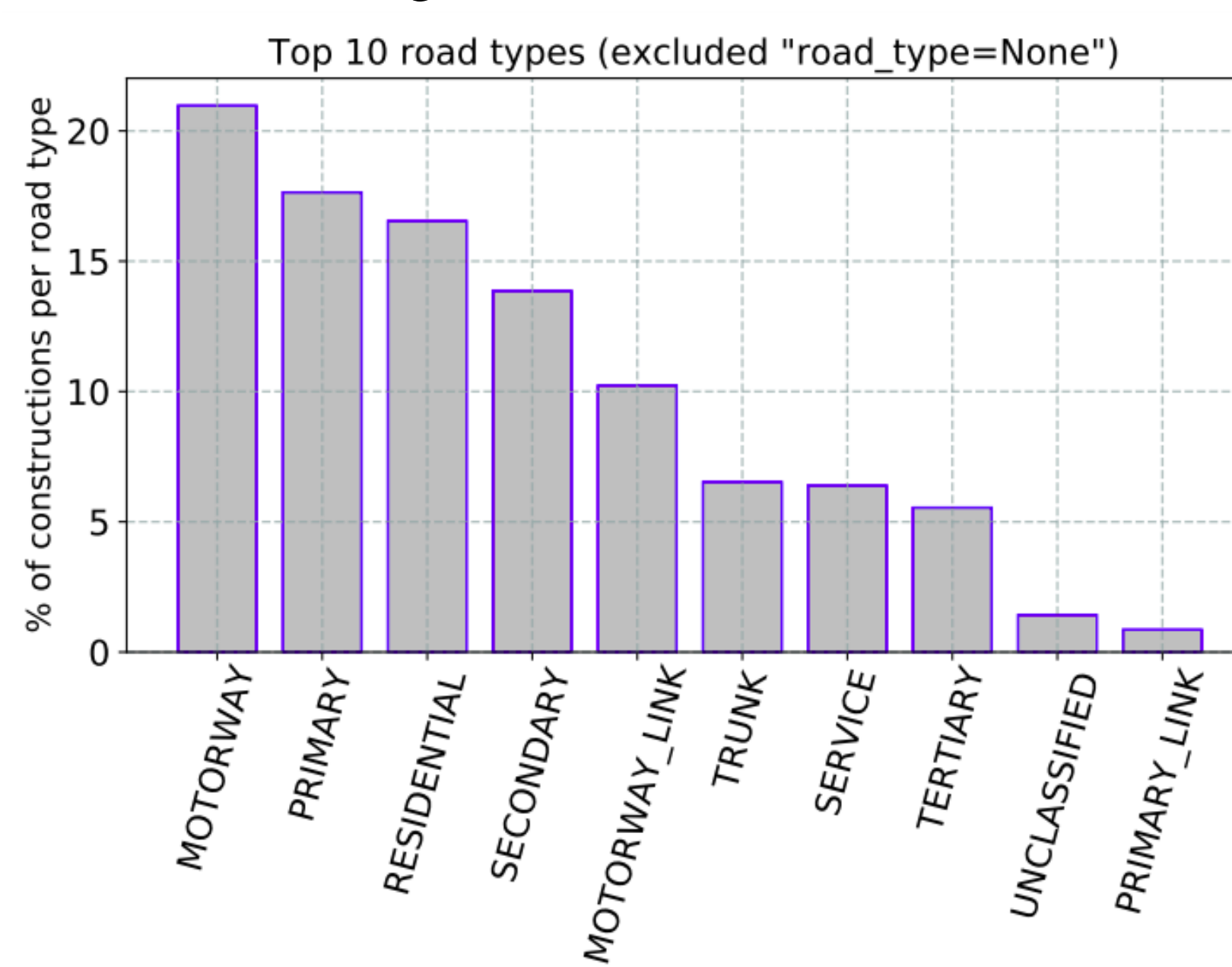
Model And Result

**Deep Road Construction Prediction (DRCP):**
DRCP gets two types as an input:
- **An image:** The road network represented in the map tiles is also a relevant context within a spatial encoding of the context for constructions. Constructions could be less prevalent on a road located in a remote area and more prevalent on a road in an urban area.
- **Feature vector:** We create a feature vector representation for each hexagonal geographical region r of time interval t = 15 days by aggregating all the events and averaging over them.

DRCP contains three parts:
- **CNN component:** The use of this component is to encode map image data to extract latent spatial features.
- **RNN component:** To encode sequential data, we use two layers of (LSTM)
- **Fully connected component:** The input to this component is simply a concatenation of the outputs of the other two components. The final production of the fully connected component shows whether there will be construction in the next 15 days or not.

State	F1-score	Accuracy
Colorado	0.934	84.4%
Florida	0.942	85.7%
Georgia	0.942	88.1%
Michigan	0.941	88.2%
New York	0.925	84.6%
Pennsylvania	0.912	85.0%
Texas	0.938	84.7%
Washington	0.947	88.9%

Result **Houston in Texas**. A **green** hexagon shows no construction is reported or predicted during the corresponding time frame, and a **red** hexagon indicates otherwise.
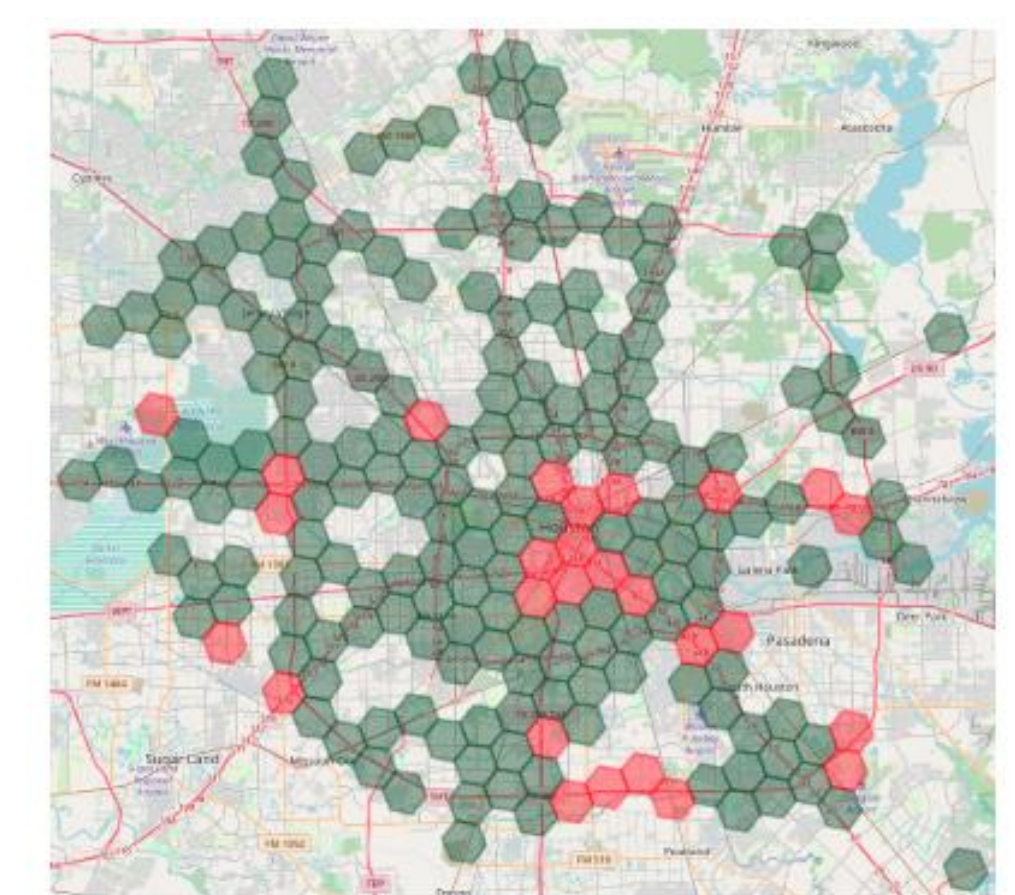
(a) Predicted / 6-1-2020 to 6-15-2020
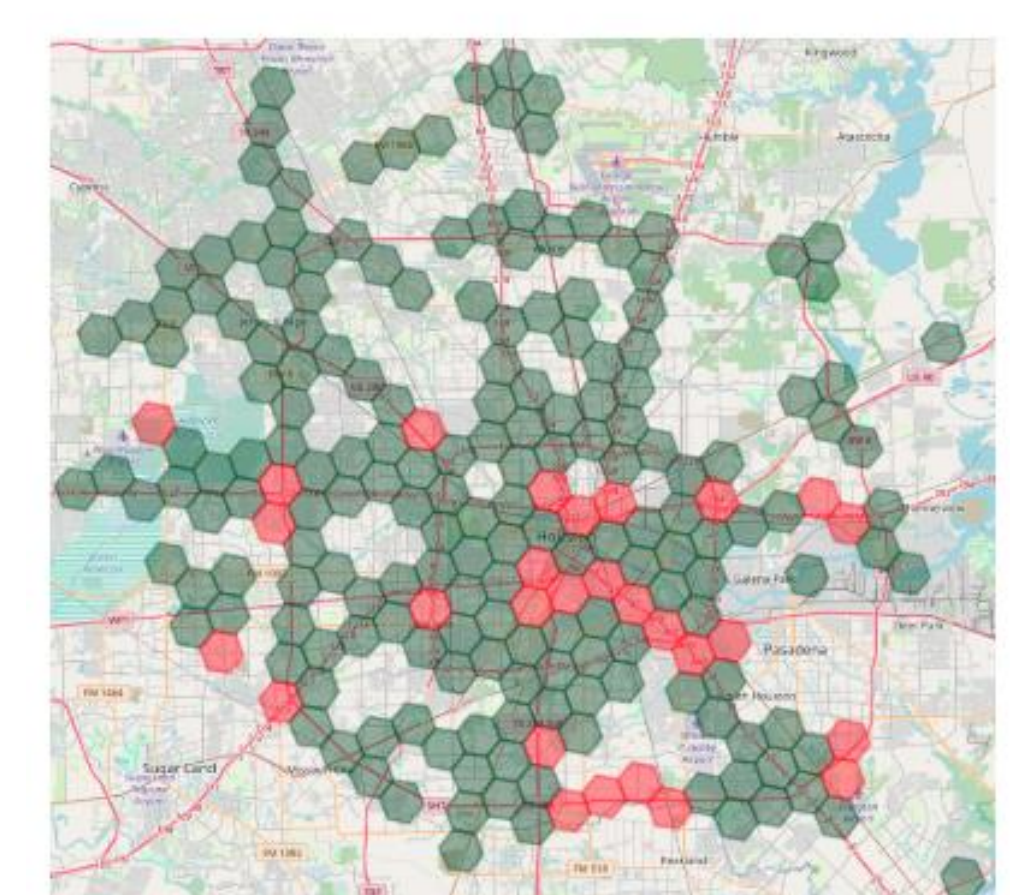
(d) Actual / 6-1-2020 to 6-15-2020

Conclusion and More

- We tackle the problem of future constructions prediction using heterogeneous spatiotemporal information.
- As directions for future research, we can extend our input data and leverage information such as traffic load and past traffic accidents.
- You can download our paper by using this QR code.