

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

Amin Karimi Monsefi
The Ohio State University
Columbus, Ohio
karimimonsefi.1@osu.edu

Sobhan Moosavi
Lyft Inc.
San Francisco, California
smoosavi@lyft.com

Rajiv Ramnath
The Ohio State University
Columbus, Ohio
ramnath.6@osu.edu

ABSTRACT

Road construction projects maintain transportation infrastructures. These projects range from the short-term (e.g., resurfacing or fixing potholes) to the long-term (e.g., adding a shoulder or building a bridge). 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 costly and difficult to scale. An alternative is the use of computational approaches that integrate and analyze multiple types of past and present spatiotemporal data to predict location and time of future road constructions. This paper reports on such an approach, one that uses a deep-neural-network-based model to predict future constructions. Our model applies both convolutional and recurrent components on a heterogeneous dataset consisting of construction, weather, map and road-network data. We also report on how we addressed the lack of adequate publicly available data - by building a large scale dataset named “US-Constructions”, that includes 6.2 million cases of road constructions augmented by a variety of spatiotemporal attributes and road-network features, collected in the contiguous United States (US) between 2016 and 2021. Using extensive experiments on several major cities in the US, we show the applicability of our work in accurately predicting future constructions - an average f1-score of 0.85 and accuracy 82.2% - that outperform baselines. Additionally, we show how our training pipeline addresses spatial sparsity of data.

KEYWORDS

US-Constructions, Road Constructions Prediction, Dataset

ACM Reference Format:

Amin Karimi Monsefi, Sobhan Moosavi, and Rajiv Ramnath. 2022. Will there be a construction? Predicting road constructions based on heterogeneous spatiotemporal data. In *The 30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*, November 1–4, 2022, Seattle, WA, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3557915.3560943>

1 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%

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGSPATIAL '22, November 1–4, 2022, Seattle, WA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9529-8/22/11.
<https://doi.org/10.1145/3557915.3560943>

increase in just five years¹. In order to make the best use of the substantial amount of capital invested in this sector, it is crucial to determine construction sites wisely. Deciding what the next construction project is and when it is to be scheduled is traditionally done through inspection by humans using special equipment - an approach that is expensive and limited in coverage. We therefore explored an alternative for this paper, namely, the use of computational solutions to determine future constructions.

To our knowledge, determining when and where road constructions are needed via computational solutions is a relatively new, less explored research area. Existing studies have focused on image analysis, such as for detecting cracks and potholes in road surface [3, 4, 11, 43] and detecting road closures via telematics and vehicle probe data [18]. While promising, the lack of extensive input data (e.g., fine-grained satellite or street imagery at scale) has limited the applicability and extensibility of these approaches. Additionally, the data used are mostly private, which limits both independent validation and the ability to build on and extend prior work. In general, current computational solutions have been limited due to the need for data to build good models with real-world applicability.

The work presented here begins by addressing challenges with data. To this end, this paper introduces a unique dataset of 6.2 million road constructions in the United States between 2016 and 2021. Our dataset offers a variety of contextual details for each construction, including location, time, a brief human provided description, daylight and weather at the start of each construction, and several map related features that contextualize the location of a construction (e.g., if it is close to a highway junction, road-type, etc.). We also carefully describe our process for building this dataset. Thus, researchers may either directly use our dataset or mimic our approach and build their own dataset.

Next, we explore an important and useful research problem, that of “identifying future constructions from past constructions along with their spatiotemporal context (e.g., traffic, weather, and map imagery) for certain locations (represented by their geographical region hexagon – see Section 4) during specific time frames (e.g., the next 15 days). Our goal is to develop a cost-saving, coverage-enhancing approach complementary to current practices. For example, we see our approach as being used to quickly identify potential sites, which can then be evaluated by human inspections.

In our approach, we model heterogeneous spatiotemporal information using a deep-neural-network that combines *recurrent* and *convolutional* components. The convolutional component is used for extracting latent information from map imagery. The recurrent component models time-series data (e.g., traffic and weather) along with additional spatial information (e.g., features of the road-network) about a location. The output of these components are then concatenated and fed to a fully connected component which produces the final output. Our goal with this model is to predict the possibility of a construction event in the near term (specifically, the next 15

¹Visit <https://www.census.gov/construction/c30/prpdf.html> for detailed reports.

days). While this formulation best suits short-term constructions (i.e., those that take a few hours to a few days), we believe it is very useful effective, given that the majority of constructions in our data can be considered as short-term (see Section 3). Extensive experiments and results demonstrate improvements over traditional and neural-network-based machine learning baselines. On average, our proposed model outperformed the best baseline model by 3.2% in accuracy and by 2.8% in F1-score, when tested over multiple major cities in the United States. Our model also demonstrated robustness in dealing with *spatial sparsity* in tests at the state level - a real-world scenario of training data being only available for parts of a region. In summary, the main contributions of this paper are:

- **Dataset:** We introduce a new dataset of road constructions and closures for the continental United States, with about 6.2 million cases from the years 2016 to 2021. To our knowledge, this is the first public dataset that offers this type and scale of data.
- **Insights:** We glean a variety of insights by analyzing the US-Constructions dataset. We detail these insights with a view to inspiring other researchers to use our data for other 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 is capable of using heterogeneous data, and resulted in superior prediction outcomes when compared to several state-of-the-art traditional and deep-learning models.

2 RELATED WORK

We examined several previous studies on road construction issues. These studies examined a variety of topics ranging from detection of road issues (such as cracks) [13], analysis of the maintenance of roads [16, 21, 23, 37, 38], prediction and management of the costs of road construction [9, 27, 28, 30, 33], and lifecycle analysis of roads [12, 14, 15, 17, 41]. We highlight some of this work next.

Tong et al. [40] employed deep convolutional neural network (DCNN) models for finding the length of cracks in asphalt pavement from gray-scale images. This work used a dataset of 8,000 images, 7,500 images with cracks and 500 images without. They classified images to 8 different classes according to the length of the crack in centimeters, achieving an accuracy of 94.36% in this classification. In another study, Ye et al. [45] employed a convolutional neural network (CNN) model to identify potholes in asphalt pavements. They used a dataset of 400 images that were collected from different pavements under different lighting conditions. These images were cropped into 96,000 smaller images of size 256×256 pixels. Authors reported 98.95% precision to detect potholes, and their stability study suggested robustness in various real-world settings.

Automatic detection of road-closures is the topic of another group of studies [7, 34]. Cheng et al. [7] presented a high-efficiency road closure detection framework based on multi-feature fusion. Their framework had two parts, an offline road closure feature modeling part and an online identification part. For the offline modeling, they first partitioned the road-network into grids, and then extracted their road closure features of these grids and the roads within them from historical data. The online component screened out closed grid candidates based on the plunge in traffic flow. They also identified sections with road closures based on variations in turning behavior by drivers on these roads. Their framework was evaluated on three real-world datasets - from Chengdu, Shanghai, and Beijing.

With respect to the life-cycle analysis (LCA) of roads Gulotta et al. [12] applied a life-cycle approach to evaluate the energy and environmental impacts of a typical urban road in Italy. They evaluated the energy and environmental impact of various bituminous combinations. For each analyzed scenario, the contribution of each life-cycle phase to the total effects and to the energy and environmental hot-spots were identified as opportunities for improvement. Jiang et al. [17] reviewed and analyzed 94 papers that adopted LCA methods to assess the environmental effects over the life cycle of roads. Their study resulted in multiple outcomes including identifying limitations and challenges of using LCA in the environmental assessment of roads, as well as in identifying future research directions.

Our paper borrows concepts from the papers described above while tackling the different problem of predicting the possibility of a future construction event at a location. To our knowledge, this is the first study that seeks to solve a problem of this type using a purely computational approach. This work has real-world application, in that it can be used to find areas in need of maintenance. It is also cost-effective because it is based on analyzing already recorded and available information such as past road constructions, road-network features (e.g., road class data, average speed, and map annotations), weather data, and coarse-grained map images. As we show in this paper, the input data we employ is easy to collect and available to the public. This is in contrast to those studies so far, which utilized private or extensive datasets.

3 THE US-CONSTRUCTIONS DATASET

In this section, we describe the countrywide traffic construction dataset, which we have named *US-Constructions*, and our process for building it. The ten major steps in this process are shown in Figure 1 and described in this section. The resulting dataset contains 6.2 million road constructions that took place in the continental United States between January 2016 and December 2021, and is publicly available at <https://www.kaggle.com/datasets/sobhanmoosavi/us-road-construction-and-closures>.

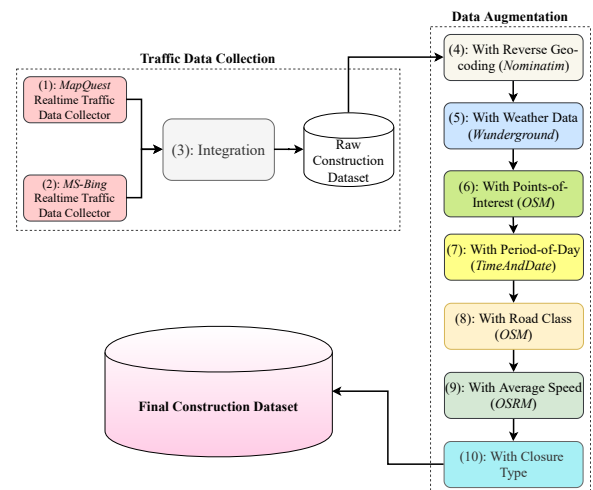


Figure 1: Process of Building US-Constructions Dataset

3.1 Collecting Raw Construction Data

We collected streaming traffic data from two real-time data providers - “MapQuest Traffic” [24] and “Microsoft Bing Maps” [5]. These providers broadcast real-time road construction and closure data collected by a variety of entities (e.g., law enforcement agencies, traffic cameras, and traffic sensors). We integrated the raw data to remove duplicates using a set of heuristic spatiotemporal filters. 0.63 million cases were pulled from MapQuest, and 5.54 million cases from BingMaps, with less than 1% overlap between the two sources.

3.2 Augmentation with Reverse Geocoding

This is the process of translating raw location data (i.e., GPS coordinates) to addresses that included elements such as *street number*, *street name*, *city*, *state*, and *zip-code*. We employed the *Nominatim* tool [29] to perform this reverse geocoding.

3.3 Augmentation with Weather Data

Weather information is a useful context for traffic constructions (in particular short-term cases). We employed *Weather Underground* API [42] to obtain the weather during each construction. Raw weather data was collected from 2,072 airport weather stations from around the United States. This raw data comes in the form of observation records, where each record consists of several attributes such as *temperature*, *humidity*, *wind-speed*, *pressure*, *precipitation* (in millimeters), and *condition*². We collected several records per day from each station. Note that each record is generated when any significant change occurs in any of the measured weather attributes.

Each construction event c was then augmented with weather data as follows. First the closest weather station s was identified. Then, of the weather observation records reported from s , we looked for the weather observation record w whose reported time was closest to the start time of c . c was then augmented with w .

3.4 Augmentation with POI Annotation

Points-of-interest (POI) are locations annotated on a map as *amenities*, *traffic signals*, *crossings*, etc. These annotations are associated with the *nodes* on a road network. A node can be associated with many POI tags; in this work, we used the 12 tags described in Table 1. We obtained these tags from the OpenStreetMap (OSM) [31] system, and for the continental United States. The applicable POI annotations for a traffic construction c are those that are located within a threshold distance τ from c . We adopted the same threshold as in [26].

3.5 Augmentation with Period of Day

Given the start time of a construction record, we used “TimeAndDate” API [39] to label it as *day* or *night*. We assign this label based on four different daylight systems, namely *Sunrise/Sunset*, *Civil Twilight*, *Nautical Twilight*, and *Astronomical Twilight*.

3.6 Augmentation with Road Class

Road class (e.g., primary, secondary, and motorway) is an important feature of the location of a construction. We used OSM to obtain this information and adopted its road classification system³ to annotate constructions. We describe our annotation process below.

²Possible values are *clear*, *snow*, *rain*, *fog*, *hail*, and *thunderstorm*.

³See <https://wiki.openstreetmap.org/wiki/Key:highway> for a list of OSM-based road types.

Table 1: Definition of Point-Of-Interest (POI) annotations based on OpenStreetMap (OSM).

Type	Description
Amenity	Refers to particular places such as restaurant, library, college, bar, etc.
Bump	Refers to speed bump or hump to reduce the speed.
Crossing	Refers to any crossing across roads for pedestrians, cyclists, etc.
Junction	Refers to any highway ramp, exit, or entrance.
No-exit	Indicates there is no possibility to travel further by any transport mode along a formal path or route.
Railway	Indicates the presence of railways.
Roundabout	Refers to a circular road junction.
Station	Refers to public transportation stations (bus, metro, etc.).
Stop	Refers to stop signs
Traffic Calming	Refers to any means for slowing down traffic speed.
Traffic Signal	Refers to traffic signals on intersections
Turning Loop	Indicates a widened area of a highway with a non-traversable island for turning around.

For a given construction c , the goal is to find the most relevant road class based on its start and end locations⁴. The OSM map data contains *nodes* and *ways*. A node is a single point defined by latitude, longitude, and a node id. A way represents a route in a road-network by an ordered list of nodes. We map location of a construction to an OSM node by using the “Nearest Service” OSRM APIs⁵ to find the S nearest nodes to the start and end location of a construction (we empirically set $S = 10$). After finding the S nearest nodes, we prune outlier nodes using an aggressive distance threshold $D = 50meters$. This results in a set \mathcal{N} of nodes nearest to the location of a construction. Next, for each node $n \in \mathcal{N}$, we find a set W_n of the OSM map “ways” that contain n , and build an aggregated set $\mathcal{W} = \bigcup_{n \in \mathcal{N}} W_n$. For each way $w \in \mathcal{W}$, we then use the OSM service “Full” to obtain road-class information⁶. Finally, we look for the most frequent road class that was returned for the ways in \mathcal{W} , and annotate the construction c with that.

3.7 Augmentation with Average Road Speed

Next, we obtain the average speed for each construction event from its start and end locations. For the about 10% of constructions that do not have an end location, we do not infer an average speed. Instead, we use the OSM “Route Service”⁷ to estimate a free-flow speed for the roads with those constructions.

3.8 Augmentation with Closure Type

Some of the constructions could result in *road closures*. We introduce a rule-based process to annotate each construction with a closure type, if there is one. To begin with, we define three cases: *road-closure*, *lane-closure*, and *no-closure*. Our process of closure-based-annotation uses the human-provided description of each construction event, and employs several *regular-expression patterns* to infer the closure type. Examples of human-provided descriptions are listed in Table 2.

⁴If the end location is not available, then we only use the start location

⁵See <http://project-osrm.org/docs/v5.5.1/api/#nearest-service>

⁶See https://wiki.openstreetmap.org/wiki/API_v0.6 and check out *GET /api/0.6/[way|relation]/#id/full*

⁷see <http://project-osrm.org/docs/v5.5.1/api/#route-service>

Table 2: Samples of human-provided description for some construction events that resulted in closure. “Source” shows which data provider was used to collect the data.

Description of Construction Event	Source
Closed due to roadwork	Bing
Closed for bridge demolition work	Bing
Right lane blocked.	Bing
Two lanes blocked	Bing
Roadway reduced to 1 lane	Bing
Lane blocked. One lane closed	Bing
Lane closed due to construction work	MapQuest
Shoulder closed on entry ramp due to construction work	MapQuest
Intermittent lane closures due to utility work	MapQuest
Three lanes closed due to construction work	MapQuest
Right lane closed due to construction work	MapQuest
One lane closed due to maintenance work	MapQuest

From manual probing of the human-provided descriptions for a large set of 10,000 randomly selected construction events we found 14 distinct patterns that represent closures. Of these patterns, four represent a *road closure* and ten of them a *lane closure*. Table 3 shows these patterns along with corresponding data source, as well as closure-type.

Table 3: Regular expression patterns to annotate closure type for construction events

Pattern	Source	Closure Type
close* * roadwork	Bing	Road
close* * bridge	Bing	Road
hard shoulder close*	Bing	Lane
* lane* block*	Bing	Lane
* reduced * lane*	Bing	Lane
* lane* close*	Bing	Lane
close* * roadwork	MapQuest	Road
road close* *	MapQuest	Road
* lane closure*	MapQuest	Lane
hard shoulder block*	MapQuest	Lane
* reduced * lane*	MapQuest	Lane
* lane* block*	MapQuest	Lane
* lane* close*	MapQuest	Lane
* shoulder close*	MapQuest	Lane

To evaluate the effectiveness of our patterns in finding closure types, we randomly selected 1,000 construction events and annotated them with the derived patterns, which resulted in 43 road closures and 137 lane closures. After manually checking all resulting annotations and also cases without any annotations, we observed *precision* = 100% and *recall* = 100%, which showed that our set of regular expression patterns was comprehensive and accurate.

3.9 Final Dataset

The final dataset comprises 6.2 million construction records, collected between January 2016 and December 2021. Each construction record is described by 45 attributes as shown in Table 4.

We ran a meta-analysis on this dataset. Our findings from this analysis are detailed below:

Table 4: US-Constructions Dataset (details as of Dec 2021)

Total Attributes	45
Traffic Attributes (8)	id, severity, start_time, end_time, start_point, end_point, distance, and description
Address Attributes (8)	number, street, side (left/right), city, county, state, zip-code, country
Weather Attributes (10)	timestamp, temperature, wind_chill, humidity, pressure, visibility, wind_direction, wind_speed, precipitation, and condition (e.g., rain, snow, etc.)
POI Attributes (12)	All cases in Table 1
Period-of-Day (4)	Sunrise/Sunset, Civil Twilight, Nautical Twilight, and Astronomical Twilight
Other Attributes (3)	Road class (e.g., highway, primary, etc.) Average speed (e.g., 55 mph), and Closure type
Total Constructions	6.2 million
# MapQuest Constructions	634k (10%)
# Bing Constructions	5.54 million (90%)
# Reported by Both	1%

- **Top states and cities:** Figure 2 shows top states and cities, as well as top states when the number of constructions are normalized by the aggregate length of road network (in miles) for each state (data is taken from Federal Highway Administration site⁸). Note how the state distribution changes after the normalization.
- **Monthly distribution:** Figure 3(a) shows that we should expect to see more constructions start as well as finish towards the end of a year. The last four months of a year typically see more constructions finish than start, likely because of a pressure to finish ongoing work before the calendar year ends.
- **Closure distribution:** According to Figure 3(b), about 15% of constructions result in a lane closure, while about 5% resulted in a complete road closure. Also we see constructions reported by MapQuest had more closures, although the overall numbers were dominated by the lower percentages reported by BingMaps.
- **Road-class distribution:** Figure 3(c) shows that high-speed roads such as “motorway” and “primary” led in construction numbers; however “residential” roads (i.e., a low-speed road type) came third. This was an interesting observation, in that it showed that construction takes place on most types of roads.
- **Map annotation (or POI) analysis:** The distribution of constructions annotated by POI in Figure 3(d), showed that a good portion of constructions are reported near intersections. Locations near amenities and stations also make up a big share of the work.
- **Duration and daylight analysis:** As shown in Figure 3(e), over 60% of constructions only last a few hours. If we consider any construction that lasts 15 days or more to be a long-term construction, then a little over 10% of constructions are long-term. Further, Figure 3(f) shows the majority of constructions started in the “day” based on all four twilight systems. However, constructions reported by MapQuest mostly started during the “night”. In other words, a good mix of construction times were present in our dataset.

4 RESEARCH QUESTION

We define our research question in this section. Suppose we are given a set C of construction events as follows:

⁸see <https://www.fhwa.dot.gov/policyinformation/statistics/2020/hm60.cfm>

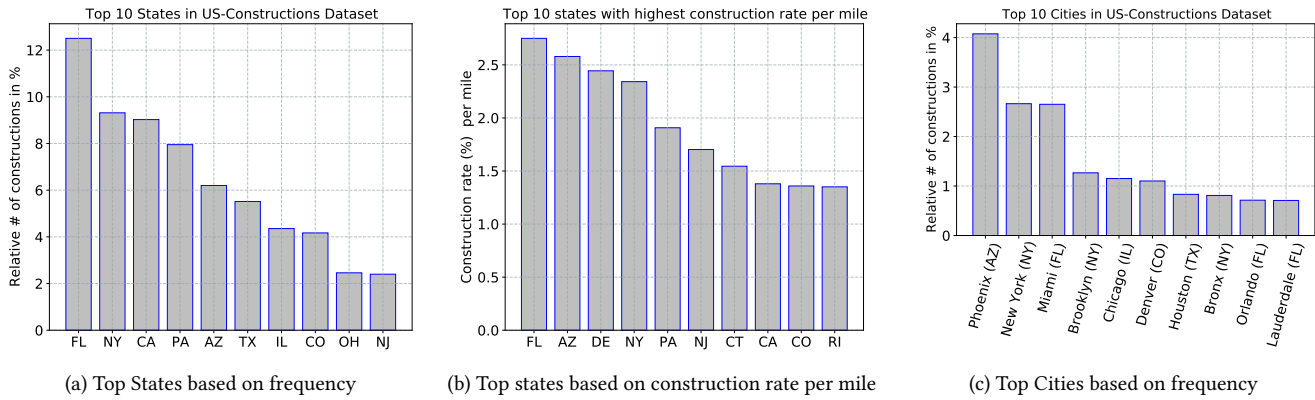


Figure 2: Construction distribution over top states (a), top states normalized by length of road network in miles (b), and top cities (c).

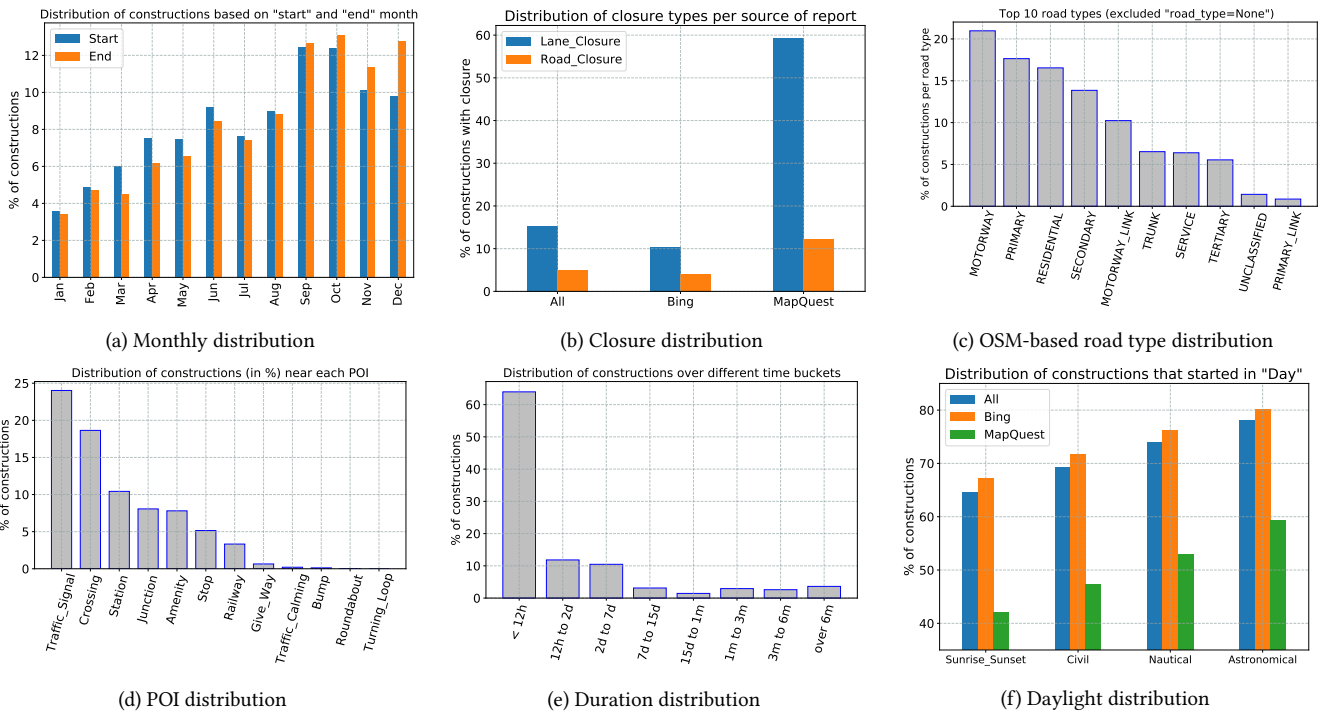


Figure 3: (a) Monthly distribution of constructions based on their start and end time; (b) Distribution of constructions resulting in road closure; (c) OSM-based road-type distribution; (d) OSM-based road annotation type (aka POI) distribution; (e) Duration distribution based on start and end time; and (f) Daylight distribution to show percentage of constructions that started in a day based on four different daylight systems.

Definition 4.1 (Construction event). We define a construction event c as $(lat, lng, start_time, end_time, desc, temp, humid, w_cond, severity, road_info, road_type)$. Here lat and lng are GPS coordinates, $start_time$ is the time of occurrence, end_time is the completion time, $desc$ is a human-provided description, $temp$ is the temperature in Fahrenheit, $humid$ is the percentage humidity, w_cond is the weather condition (e.g., rain, snow, and clear) with a $severity$ value an integer between 1 and 4 (note that all weather features are for the start of the construction), $road_info$ are additional details about the road (e.g., distance and average speed), and $road_type$ is the type of the road (e.g., motorway, primary, and secondary).

In addition to the above set of construction events, we have a database of geographical map images M consisting of hexagonal tiles (in any possible view, e.g., satellite, transit, and terrain) with sufficient resolution. Additionally, we have a dataset of points-of-interest P (e.g., amenities, traffic lights, and stop signs) for a specific zone (or region) of the map. Given these datasets, we define our research question below.

Given:

- A set of spatial regions $R = \{r_1, r_2, \dots, r_n\}$, where $r \in R$ is a hexagonal zone, according to the definition provided by Uber

h3⁹ library [35]. Here we choose a resolution level 7 which results in zones with edge size $1.2km$ and area $5.16km^2$.

- A set of fixed-length time intervals $T = \{t_1, t_2, \dots, t_m\}$, where we set $|t| = 15 \text{ days}$, for $t \in T$.
- A database of construction events $C_r = \{c_1, c_2, \dots\}$ for $r \in R$.
- A database of map image data $M_r = \{m_1, m_2, \dots\}$ for $r \in R$.
- A database of points-of-interest $P_r = \{p_1, p_2, \dots\}$ for $r \in R$.

Create:

- A representation F_{rt} for a region $r \in R$ during a time interval $t \in T$, using C_r , M_r , and P_r .
- A binary label L_{rt} for F_{rt} , where 1 indicates at least one traffic construction was reported during t in region r ; 0 otherwise.

Find:

- A model \mathcal{M} to predict L_{rt} using $\langle F_{rt_{i-10}}, F_{rt_{i-9}}, \dots, F_{rt_{i-1}} \rangle$, which means predicting the label of current time interval using observations from the last 10 time intervals.

Objective:

- Minimize the prediction error.

Note that we chose the size of regions and time intervals in order to address the sparsity of input data, while still building a viable model that could provide real-time insights. Interested researchers are encouraged to use different settings to further explore the data and the task.

5 MODEL

This section describes our construction prediction model. We start with a description of the input to the model, that is the feature vector representation of construction events and map image data. Then we describe the model.

5.1 Feature Vector Representation

We create a feature vector representation for each hexagonal geographical region r of resolution 7 during a time interval $t = 15 \text{ days}$ by aggregating all the events and averaging over them. To be precise, a construction event includes the following features:

- **Weather (14):** A vector representing two weather attributes temperature and humidity; and 12 indicators to represent special weather events `light_rain`, `moderate_rain`, `heavy_rain`, `light_snow`, `moderate_snow`, `heavy_snow`, `severe_cold`, `severe_storm`, `severe_fog`, `moderate_fog`, `hail`, and `precipitation_other`. Weather data is obtained from [25].
- **POI (15):** A vector of size 15 to represent frequency of POIs (or map annotations) within r . In addition to the cases described in Table 1, we also consider `entrance`, `give_way`, and `turning_circle`. We obtain POI data from [31].
- **Road type (25):** A one-hot vector of size 25 to show the type of the road in region r with construction. Road-type information is also extracted from [31].
- **Road information (5):** On a road segment with a reported construction, this category offers five attributes, namely road segment distance, average speed, approximate travel time, an indicate to show whether the traffic on road segment was impacted during the construction, and severity of the construction. The latter is an integer value between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the event) and 4 indicates a significant impact on traffic (i.e., long delay).

To build the aggregated view for all the construction events that occurred during t within region r , we simply average over the 59 attributes described above. Note that except for POIs, other features are different for the different constructions that we are aggregating over the region. For instance, two constructions that started at different times could have different weather attributes; or they could be associated with different road types. Thus, out of the 59 features, only 15 are constant over time.

5.2 Map Image Representation

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 (with a sparse road network), and more prevalent on a road in an urban area (with a dense road network). Figure 4 shows an example of the type of map images we use, again collected from OpenStreetMap [31]. We associate each zone with one image. To do so, we first obtain the center of our hexagon in terms of GPS coordinates, and then collect a map tile with the same center from OSM with a zoom level of 14. According to OSM¹⁰, a map tile of `zoom_level = 14` is a square image of size $256 * 256$ pixels. Each pixel of this image covers about 9.547 meters , that means the side size of the image is about $2.44km$ and its area is about $5.95km^2$. Thus, one single image can almost cover an entire zone, since each side of a zone is about $1.2km$ and its area is about $5.16km^2$.



Figure 4: Example of a map image extracted from OSM to (roughly) represent a hexagon zone (`zoom_level = 14`)

5.3 The Deep Road Construction Prediction (DRCP) Model

Our Deep Road Construction Prediction (DRCP) model is shown in Figure 5. Since our input data contains both images and time-series data we use a mix of convolutional (CNN) [44], recurrent (RNN) [22] and fully connected components.

- **CNN component:** The use of this component is to encode map image data to extract latent spatial features. The size of an input image is $256 * 256 * 3$. The *CNN* component built from these images is shown in Figure 5, and it comprises four sub-components,

⁹see <https://github.com/uber/h3>

¹⁰see https://wiki.openstreetmap.org/wiki/Zoom_levels

three convolutional blocks (with 4, 32, and 8 channels, respectively) and one *decoder* component. The decoder component contains six decoder blocks, which are convolutional layers with 8 channels in the first and 16 channels in the last three blocks. All decoder blocks include batch normalization [36] to deal with internal covariate shift, max pooling [10] for downsampling, and *ReLU* [2] as an activation function. The other sub-components (i.e., three convolutional blocks) do not leverage max-pooling, but batch normalization is used in two of them. The output of the CNN component is then converted to a vector of size 128 by using a flatten layer. Note that the activation function used in the last sub-component is *sigmoid* to properly concatenate the outputs of CNN and RNN components. The kernel size in all convolutional layers is 3×3 , and stride size is 1. It is worth noting that this design is mostly driven empirically through exploring different architectures and settings.

- **RNN component:** To encode sequential data, we use two layers of Long Short Term Memory (*LSTM*) [46] with 59 and 45 neurons, respectively. Both layers of *LSTM* use *sigmoid* as activation function. The choice of 59 neurons in the first *LSTM* layer is to utilize sequential input of size 10×59 that represents aggregated construction event data over the past 10 time intervals. The output of this component is then converted to a vector of size 40 by using a dense layer. We employed *sigmoid* activation function for the dense layer too.
- **Fully connected component:** This component includes three layers of size 64, 16, and 1. The input to this layer is simply a concatenation of the outputs of the other two components, which is a vector of size 168. The final output of the fully connected component shows whether there will be a construction in the next 15 days (i.e., next time interval) or not.

We employed an extensive grid-search-based hyper-parameter tuning process to find the best settings (e.g., layer size and network size) for each component.

6 EXPERIMENT AND RESULTS

In this section we first describe experiment setup, then describe data, followed by baseline models description, and conclude it by presenting results and discussions¹¹.

6.1 Experiment Setup

All implementations¹² are in Python using Tensorflow [1], Keras [8], and scikit-learn [32] libraries; and experiments were run on machines at Ohio Supercomputer Center [6]. For training *DRCP* we used the *Adam* [20] optimizer. The maximum number of epochs was set to be 1000, with an early stopping policy where training stops if the loss value on the validation set does not decrease after 30 consecutive epochs. The initial learning rate was set to 10^{-4} , and if no improvements could be observed for 5 consecutive epochs, then the learning rate is reduced by a factor of 0.9 (i.e., $new\ learning\ rate = learning\ rate \times 0.9$). This reduction could potentially continue until the learning rate reaches 10^{-6} . The loss function we used was *Binary Cross Entropy* [19], since our problem is a binary classification problem and this loss function is proven to work quite well for this class of problems.

6.2 Data Description

We trained and validated our *DRCP* model, on eight states (*New York, Pennsylvania, Georgia, Texas, Colorado, Florida, Washington and Michigan*). Further, we selected nine cities (*Columbus (OH), New York City (NY), Pittsburgh (PA), Atlanta (GA), Houston (TX), Denver (CO), Miami (FL), Seattle (WA), and Detroit (MI)*). The choice of these states and cities was primarily to achieve diversity in traffic and weather conditions, population, population density, and urban characteristics (road-network, prevalence of urban versus highway roads, etc.). We split our data set into three parts. The first part was our training set with records drawn from the date range of February 2016 to the end of December 2019. The second part was our validation set, whose dates ranged from January 2020 to the end of May 2020. The last part of our data ranged from June 2020 to the end of December 2020 and was used as our as test set.

We prepared our data for use in the machine learning process as described in Section 5. Each record of data includes aggregated feature vector representation for a 15 days time interval, as well as a map image to represent the corresponding geographical zone. Note that our goal is to predict a binary label for the next interval, using data from the past 10 intervals. To mitigate label imbalance, we use *class weights* to better train different models. We empirically found the weights 1.01 and 16.01 for the classes 0 and 1, respectively. These weights were found from the use of the validation data in the *DRCP* model.

6.3 Baseline Models

We choose models described below as baselines to compare our proposal against them.

- **Logistic Regression (LR):** LR is a well-known model for binary classification tasks. To train the model we set *penalty = l2*, *random state = 0*, and used “Limited-memory Broyden Fletcher Goldfarb Shanno (lbfgs)” as the solver.
- **Random Forest (RF):** Given the nature of our input data, RF seems a natural choice to be used for the task defined in this paper. To train this model, we set *number of estimators = 10*.
- **Gradient Boosting Classifier (GB):** GB is a strong tree-based boosting model for a wide range of classification tasks, thus we choose it as a baseline in this work. We set *number of estimators = 10* to train the model.
- **Multi-layer Perceptron (MLP):** MLP is another strong choice to compare our proposal against for the construction prediction task. Here we use a hidden layer with 100 neurons, *Relu* as activation function, and *Adam* as optimizer to train the model.

Since these baseline models cannot work with sequential data, we vectorize inputs and feed them as a single vector to these models. The vectorization process makes use of both construction events data as well as map image data in a concatenated form.

6.4 Results and Discussions

In this section we define two experimentation scenarios and present results for each. The first scenario considers the case where construction data is potentially available for every zone in a region, thus training and inference can be done based on input data from all zones. The second scenario, on the other hand, assumes we only have data for some of zones in a region, but still seek to make future predictions for all zones (i.e., sparsity).

6.4.1 Scenario 1: training based on all zones. According to this scenario, the input data to train different models includes all the zones

¹¹All codes and sample data are available at <https://github.com/7Amin/DRCP>

¹²see our code on <https://github.com/7Amin/DRCP>

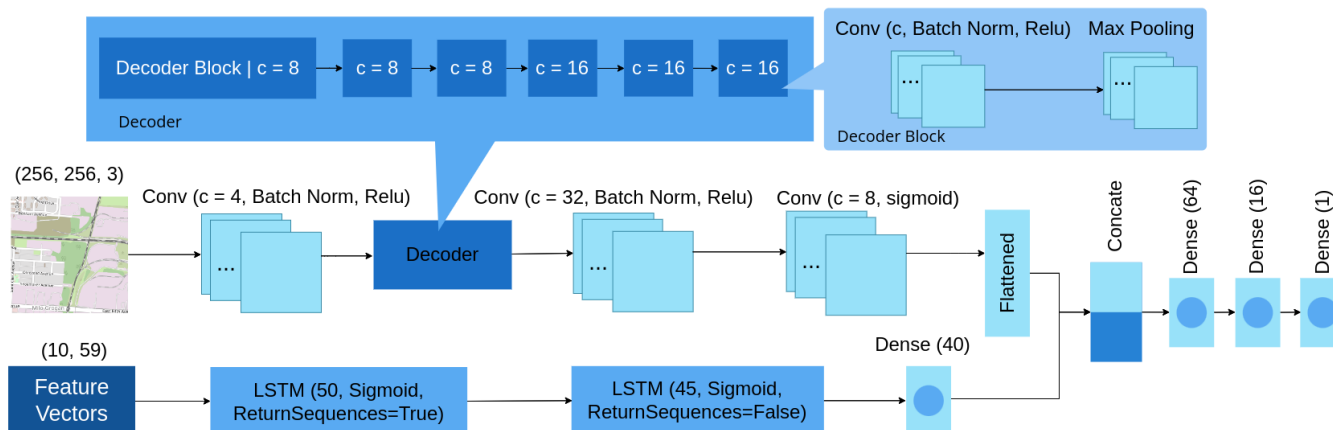


Figure 5: Deep Road Construction Prediction (DRCP) Model

that have any data to offer. Using the settings described in previous sections, we trained all the models, and Tables 5 and 6 show the results based on $f1$ -score and accuracy metrics, and for the selected cities. Our proposed model (i.e., DRCP) outperforms other models with significant margin for 8 out of 9 cities based on both f1-score and accuracy metric. On average, we observe 3.2% improvement in accuracy and 2.8% improvement in F1-score when compared to the best baseline for each city. This demonstrates the effectiveness of our proposal to make better use of spatiotemporal information to make future predictions.

Table 5: Comparing different models for selected cities for Scenario I using “F1-score”

City	LR	RF	GB	MLP	DRCP
Atlanta (GA)	0.773	0.751	0.731	0.725	0.793
Columbus (OH)	0.859	0.87	0.855	0.891	0.902
Denver (CO)	0.756	0.767	0.753	0.761	0.794
Detroit (MI)	0.774	0.786	0.743	0.728	0.795
Houston (TX)	0.891	0.895	0.892	0.884	0.908
Miami (FL)	0.856	0.843	0.845	0.868	0.845
New York City (NY)	0.895	0.898	0.884	0.906	0.942
Pittsburgh (PA)	0.798	0.794	0.782	0.768	0.832
Seattle (WA)	0.852	0.825	0.803	0.794	0.853

To visualize how our model performs, Figure 6 shows prediction results (a, b, and c) using the DRCP model and expected outcomes (d, e, and f) for Houston (TX) over three consecutive time frames. A zone represented by “green” indicates there are no road constructions reported or predicted, while a zone represented by “red” shows otherwise. It is interesting to see how the model performs over different time frames with different distributions of constructions. For instance, Figure 6-(e) shows more sparsity, while Figure 6-(f) is more dense; and in both cases our proposed model performs reasonably to make future predictions (see Figure 6-(b) and Figure 6-(c), respectively). This shows how the proper use of useful spatiotemporal information can result in satisfactory model outcomes, which is a strong indicator of applicability of this setup in the real-world. In

Table 6: Comparing different models based on Scenario I using “Accuracy” and for the selected cities

City	LR	RF	GB	MLP	DRCP
Atlanta (GA)	74.85%	72.55%	70.32%	69.44%	77.9%
Columbus (OH)	82.3%	83.19%	82.81%	86.11%	87.2%
Denver (CO)	73.56%	74.53%	73.48%	73.68%	78.2%
Detroit (MI)	74.46%	75.36%	70.9%	68.47%	78.2%
Houston (TX)	86.68%	86.8%	86.9%	86.21%	88.1%
Miami (FL)	82.86%	81.09%	82.17%	83.91%	81.6%
New York City (NY)	85.01%	83.95%	85.41%	86.33%	89.7%
Pittsburgh (PA)	76.18%	74.82%	75.53%	71.9%	77.7%
Seattle (WA)	81.39%	79.23%	77.52%	75.81%	81.43%

Section 8 we provided further visualized results of this kind for two other cities in the United States (see Figures 7 and 8).

6.4.2 Scenario II: training based on limited zones. Unlike the first scenario, here the assumption is only a limited number of zones have data to offer for training. To do so, we randomly select 60% of our zones for training, 20% for validation, and 20% for testing. To make the task even more challenging, we do state-level training and testing (as opposed to city-level as in scenario I), which results in more sparsity, and perhaps closer to a real-world setup. Using a similar training process, we individually trained the DRCP model for the selected states, and Table 7 shows the results based on “F1-score” and “Accuracy” metrics. Based on the results, we see our model performs quite reasonably when predicting on test data that include held-out zones. This is an important observation since it shows that even a partial dataset collected from a limited number of zones in a state is enough to train a generalizable model like DRCP to make inferences about future constructions for all zones in a state. Additionally, this is useful in real-world, where we may never have access to high quality data for all zones in a state or even in a city. Lastly, we note that similarities in input data (e.g., map structure and weather condition) within a state (at least for the states that we experimented with) could be another reason to justify the promising outcomes of our model in this scenario.

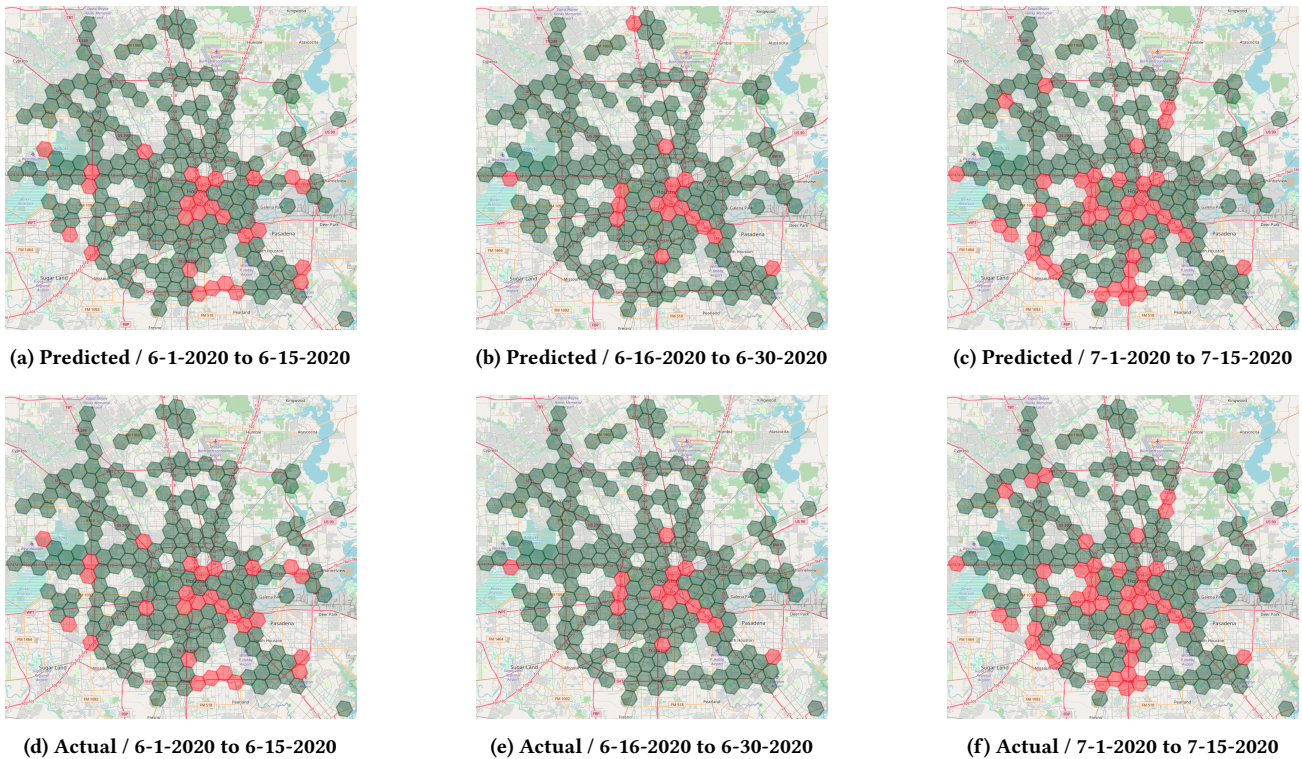


Figure 6: Example of prediction results (a to c) by the “DRCP” model along with actual labels (d to f) over three consecutive time frames for Houston in Texas. Here we make predictions for individual zones that are represented by hexagons. A “green” hexagon shows no construction is reported or predicted during the corresponding time frame, and a “red” hexagon shows otherwise.

Table 7: Prediction results using DRCP model based on Scenario II (i.e., studying spatial sparsity) for selected states

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%

7 CONCLUSION AND FUTURE WORK

In this paper we tackle the problem of future constructions prediction using heterogeneous spatiotemporal information such as past constructions, weather data, and geographical map data. To our knowledge, this is a relatively new problem space that has not been explored by the research community, maybe due to lack of comprehensive historical data about past constructions. To address this gap, this paper introduces a novel dataset of 6.2 million constructions in the United States between 2016 and 2021, that offers a variety of details around location, time, weather condition, map,

and road-network. Additionally, we formulate and solve the problem of predicting the possibility of future constructions using such data. We present a deep-neural-network-based model to efficiently utilize the heterogeneous input by combining convolutional and recurrent components in a reasonable setting. Through extensive experiments over multiple major cities and states in the United States, we show the usefulness of our proposal in a real-world setting and in comparison to several state-of-the-art baselines.

As directions for future research, we can extend our input data and leverage information such as traffic load, past traffic accidents, finer-grained weather data, demographic data for each zone, as well as other map imagery views (e.g., satellite view). In addition to data, the model that we presented in this paper can potentially be improved by jointly modeling spatial and temporal data, instead of a separate utilization. Lastly, we can extend the task that is defined in this work and tackle more advanced problems such as “construction type” prediction; that is, if a certain construction could result in a closure or not.

REFERENCES

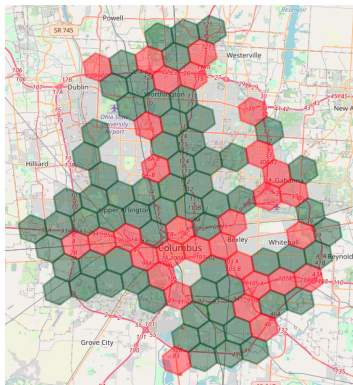
- [1] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016).
- [2] Abien Fred Agarap. 2018. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375* (2018).
- [3] Deeksha Arya, Hiroya Maeda, Sanjay Kumar Ghosh, Durga Toshniwal, and Yoshihide Sekimoto. 2021. RDD2020: An annotated image dataset for automatic

- road damage detection using deep learning. *Data in brief* 36 (2021), 107133.
- [4] Rozi Bibi, Yousaf Saeed, Asim Zeb, Taher M Ghazal, Taj Rahman, Raed A Said, Sagheer Abbas, Munir Ahmad, and Muhammad Adnan Khan. 2021. Edge AI-based automated detection and classification of road anomalies in VANET using deep learning. *Computational intelligence and neuroscience* 2021 (2021).
 - [5] Bing Map Traffic API. 2021. <https://www.bingmapsportal.com/>. Accessed: 2021-1-1.
 - [6] Ohio Supercomputer Center. 1987. Ohio Supercomputer Center. <http://osc.edu/ark:/19495/f5s1ph73>
 - [7] Guangliang Cheng, Ying Wang, Shibiao Xu, Hongzhen Wang, Shiming Xiang, and Chunhong Pan. 2017. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 55, 6 (2017), 3322–3337.
 - [8] François Chollet et al. 2015. keras.
 - [9] Uven Chong and Omar Hopkins. 2016. An international experience on the evolution of road costs during the project life cycle. *Transport Policy* 48 (2016), 60–66.
 - [10] Benjamin Graham. 2014. Fractional max-pooling. *arXiv preprint arXiv:1412.6071* (2014).
 - [11] Jinchao Guan, Xu Yang, Ling Ding, Xiaoyun Cheng, Vincent CS Lee, and Can Jin. 2021. Automated pixel-level pavement distress detection based on stereo vision and deep learning. *Automation in Construction* 129 (2021), 103788.
 - [12] TM Gulotta, M Mistretta, and FG Praticò. 2019. A life cycle scenario analysis of different pavement technologies for urban roads. *Science of the total environment* 673 (2019), 585–593.
 - [13] Dongyeob Han, Suk Bae Lee, Mihwa Song, and Jun Sang Cho. 2021. Change detection in unmanned aerial vehicle images for progress monitoring of road construction. *Buildings* 11, 4 (2021), 150.
 - [14] Umair Hasan, Andrew Whyte, and Hamad Al Jassmi. 2019. Critical review and methodological issues in integrated life-cycle analysis on road networks. *Journal of Cleaner Production* 206 (2019), 541–558.
 - [15] Endrit Hoxha, Hrefna Run Vignisdottir, Diego Maria Barbieri, F Wang, Rolf André Bohne, Terje Kristensen, and A Passer. 2021. Life cycle assessment of roads: Exploring research trends and harmonization challenges. *Science of the Total Environment* 759 (2021), 143506.
 - [16] Amjad Issa and Sameer Abu-Eisheh. 2017. Evaluation of implementation of municipal roads' maintenance plans in Palestine: A pilot case study. *International Journal of Pavement Research and Technology* 10, 5 (2017), 454–463.
 - [17] Rui Jiang and Peng Wu. 2019. Estimation of environmental impacts of roads through life cycle assessment: A critical review and future directions. *Transportation Research Part D: Transport and Environment* 77 (2019), 148–163.
 - [18] Kotaro Kataoka, Saurabh Gangwar, Karthik Yadav Mudra, and Souraj Mandal. 2018. A smartphone-based probe data platform for road management and safety in developing countries. In *2018 IEEE international conference on data mining workshops (ICDMW)*. IEEE, 612–615.
 - [19] Nikhil Ketkar. 2017. Introduction to keras. In *Deep learning with Python*. Springer, 97–111.
 - [20] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
 - [21] U Siva Rama Krishna and Chiranjeevi Tadi. 2022. Sustainable concrete pavements for low volume roads-Scientometric analysis of the literature. In *IOP Conference Series: Earth and Environmental Science*, Vol. 982. IOP Publishing, 012005.
 - [22] Chen Lei. 2021. RNN. In *Deep Learning and Practice with MindSpore*. Springer, 83–93.
 - [23] Yan Li and Chunlu Liu. 2019. Applications of multirotor drone technologies in construction management. *International Journal of Construction Management* 19, 5 (2019), 401–412.
 - [24] MapQuest Traffic API. 2021. <https://www.mapquest.com/>. Accessed: 2021-1-1.
 - [25] Sobhan Moosavi, Mohammad Hossein Samavatian, Arnab Nandi, Srinivasan Parthasarathy, and Rajiv Ramnath. 2019. Short and long-term pattern discovery over large-scale geo-spatiotemporal data. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2905–2913.
 - [26] Sobhan Moosavi, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. 2019. Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 33–42.
 - [27] Laura Moretti, Giuseppe Cantisani, and Paola Di Mascio. 2016. Management of road tunnels: Construction, maintenance and lighting costs. *Tunnelling and Underground Space Technology* 51 (2016), 84–89.
 - [28] Ali Nahvi, Jacqueline Nowak, Michael Janson, and Ryan Loos. 2021. *Data-Driven Economic Analysis of Biobased Road Preservation; A Path to Effective Low Volume Roads Management*. Technical Report.
 - [29] Nominatim Tool. 2021. <https://wiki.openstreetmap.org/wiki/Nominatim>. Accessed: 2021-1-1.
 - [30] Mbakisya Onyango, Saliha Ammour Merabti, Joseph Owino, Ignatius Fomunung, and Weidong Wu. 2018. Analysis of cost effective pavement treatment and budget optimization for arterial roads in the city of Chattanooga. *Frontiers of Structural and Civil Engineering* 12, 3 (2018), 291–299.
 - [31] OpenStreetMap contributors. 2021. Planet dump retrieved from <https://planet.osm.org>. <https://www.openstreetmap.org/>. Accessed: 2021-1-1.
 - [32] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
 - [33] K Petroutsatou, E Georgopoulos, S Lambropoulos, and JP Pantouvakis. 2012. Early cost estimating of road tunnel construction using neural networks. *Journal of construction engineering and management* 138, 6 (2012), 679–687.
 - [34] Davide Pietrobon, Andrew P Lewis, and Gavin S Heverly-Coulson. 2019. An algorithm for road closure detection from vehicle probe data. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 5, 2 (2019), 1–13.
 - [35] Kevin Sahr, Denis White, and A Jon Kimerling. 2003. Geodesic discrete global grid systems. *Cartography and Geographic Information Science* 30, 2 (2003), 121–134.
 - [36] Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry. 2018. How does batch normalization help optimization? *Advances in neural information processing systems* 31 (2018).
 - [37] Amir Shtayat, Sara Moridpour, Berthold Best, Avinash Shroff, and Divyajeetsinh Raol. 2020. A review of monitoring systems of pavement condition in paved and unpaved roads. *Journal of Traffic and Transportation Engineering (English Edition)* 7, 5 (2020), 629–638.
 - [38] V Sunitha, A Veeraragavan, Karthik K Srinivasan, and Samson Mathew. 2013. Application of Factor Analysis in Maintenance Management of Low Volume Roads. *International Journal of Pavement Research & Technology* 6, 2 (2013).
 - [39] Time And Date website. 2021. <https://www.timeanddate.com/>. Accessed: 2021-1-1.
 - [40] Zheng Tong, Jie Gao, Zhenqiang Han, and Zhenjun Wang. 2018. Recognition of asphalt pavement crack length using deep convolutional neural networks. *Road Materials and Pavement Design* 19, 6 (2018), 1334–1349.
 - [41] Giampiero Trunzo, Laura Moretti, and Antonio D'Andrea. 2019. Life cycle analysis of road construction and use. *Sustainability* 11, 2 (2019), 377.
 - [42] Weather Underground. 2014-2020. <https://www.wunderground.com/>. Accessed: 2020-1-1.
 - [43] Tian Wen, Shuo Ding, Hong Lang, Jian John Lu, Ye Yuan, Yichuan Peng, Jiang Chen, and Aidi Wang. 2022. Automated pavement distress segmentation on asphalt surfaces using a deep learning network. *International Journal of Pavement Engineering* (2022), 1–14.
 - [44] Jianxin Wu. 2017. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology, Nanjing University, China* 5, 23 (2017), 495.
 - [45] Wanli Ye, Wei Jiang, Zheng Tong, Dongdong Yuan, and Jingjing Xiao. 2019. Convolutional neural network for pothole detection in asphalt pavement. *Road Materials and Pavement Design* (2019), 1–17.
 - [46] Zheng Zhao, Weihai Chen, Xingming Wu, Peter CY Chen, and Jingmeng Liu. 2017. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems* 11, 2 (2017), 68–75.

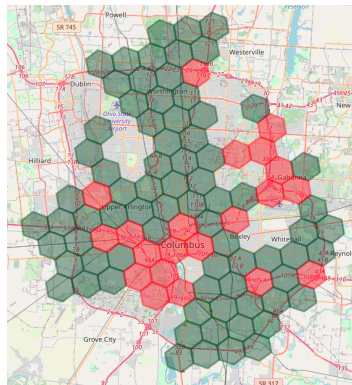
8 APPENDIX

8.1 Visualizing road construction predictions

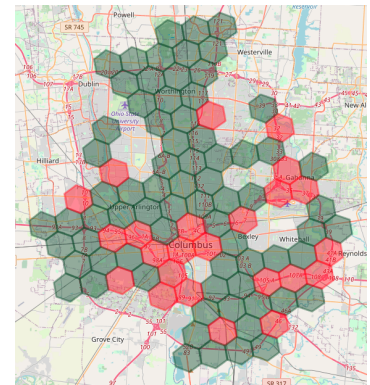
In this section we visualize road construction prediction results for Columbus (in Ohio) and Miami (in Florida) by Figures 7 and 8, respectively. Settings and details are the same as what we earlier described in Section 6.4.1. Similar to what we presented before, we can see how our proposed model performs future construction prediction in two different cities. It is worth noting again that some of the chosen time frames are sparse, given the number of reported constructions during those. However, our proposed model is capable of making solid predictions even when there is sparsity.



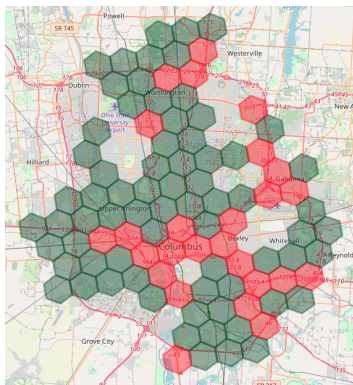
(a) Predicted / 6-1-2020 to 6-15-2020



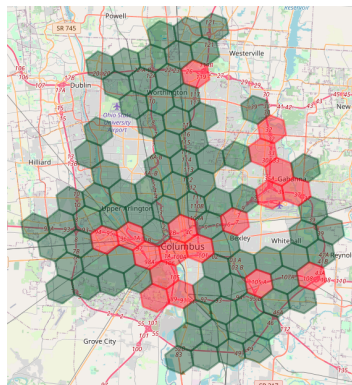
(b) Predicted / 6-16-2020 to 6-30-2020



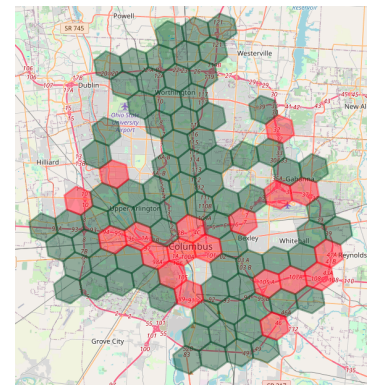
(c) Predicted / 7-1-2020 to 7-15-2020



(d) Actual / 6-1-2020 to 6-15-2020

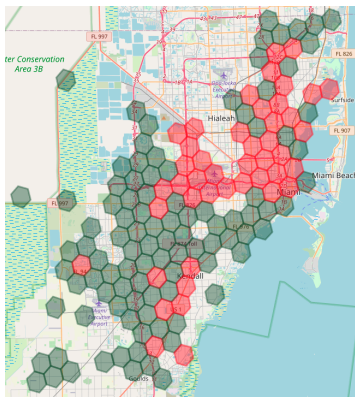


(e) Actual / 6-16-2020 to 6-30-2020

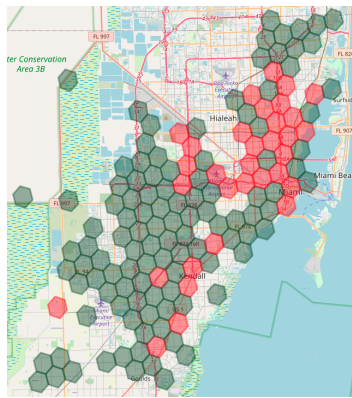


(f) Actual / 7-1-2020 to 7-15-2020

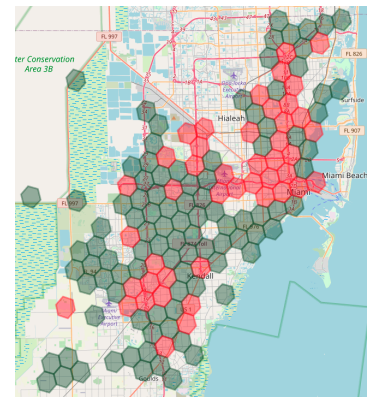
Figure 7: Example of prediction results (a to c) by the “DRCP” model along with actual labels (d to f) over three consecutive time frames for Columbus in Ohio. Here we make predictions for individual zones that are represented by hexagons. A “green” hexagon shows no construction is reported or predicted during the corresponding time frame, and a “red” hexagon shows otherwise.



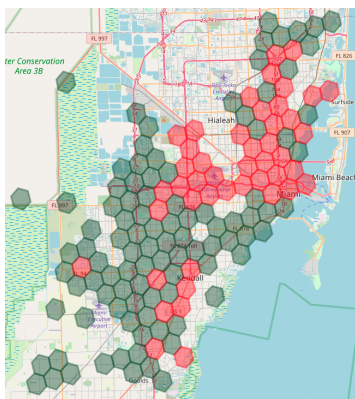
(a) Predicted / 6-1-2020 to 6-15-2020



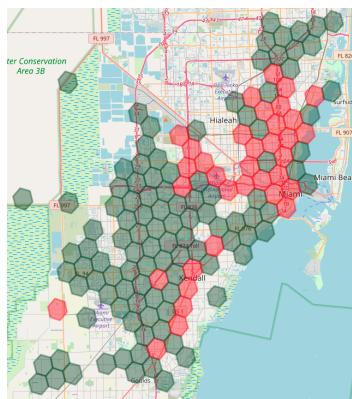
(b) Predicted / 6-16-2020 to 6-30-2020



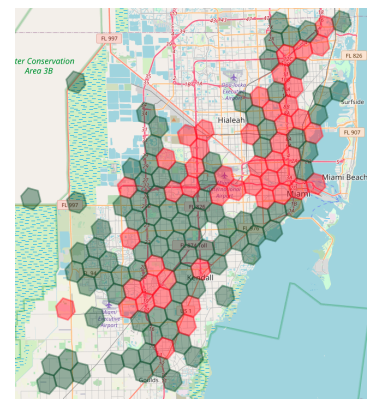
(c) Predicted / 7-1-2020 to 7-15-2020



(d) Actual / 6-1-2020 to 6-15-2020



(e) Actual / 6-16-2020 to 6-30-2020



(f) Actual / 7-1-2020 to 7-15-2020

Figure 8: Example of prediction results (a to c) by the “DRCP” model along with actual labels (d to f) over three consecutive time frames for Miami in Florida. Here we make predictions for individual zones that are represented by hexagons. A “green” hexagon shows no construction is reported or predicted during the corresponding time frame, and a “red” hexagon shows otherwise.