



# CrashFormer: A Multimodal Architecture to Predict the Risk of Crash

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

Pouya Shiri  
Circle Cardiovascular Imaging  
Calgary, Alberta, Canada  
pouya.shiri@circlecvi.com

Ahmad Mohammadshirazi  
Ohio State University  
Columbus, Ohio, USA  
mohammadshirazi.2@osu.edu

Nastaran Karimi Monsefi  
Hamedan University of Technology  
Hamedan, Iran  
k.nastaran1998@gmail.com

Ron Davies  
Ohio State University  
Columbus, Ohio, USA  
davies.404@osu.edu

Sobhan Moosavi  
Zoox Inc.  
San Francisco, California, USA  
smoosavi@zoox.com

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

## ABSTRACT

Reducing traffic accidents is a crucial global public safety concern. Accident prediction is key to improving traffic safety, enabling proactive measures to be taken before a crash occurs, and informing safety policies, regulations, and targeted interventions. Despite numerous studies on accident prediction over the past decades, many have limitations in terms of generalizability, reproducibility, or feasibility for practical use due to input data or problem formulation. To address existing shortcomings, we propose **CrashFormer**, a multi-modal architecture that utilizes comprehensive (but relatively easy to obtain) inputs such as the history of accidents, weather information, map images, and demographic information. The model predicts the future risk of accidents on a reasonably acceptable cadence (i.e., every six hours) for a geographical location of 5.161 square kilometers. CrashFormer is composed of five components: a sequential encoder to utilize historical accidents and weather data, an image encoder to use map imagery data, a raw data encoder to utilize demographic information, a feature fusion module for aggregating the encoded features, and a classifier that accepts the aggregated data and makes predictions accordingly. Results from extensive real-world experiments in 10 major US cities show that CrashFormer outperforms state-of-the-art sequential and non-sequential models by 1.8% in F1-score on average when using “sparse” input data.

## CCS CONCEPTS

• **Computing methodologies** → **Supervised learning by classification**; • **Applied computing** → *Transportation*.

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*UrbanAI '23, November 13, 2023, Hamburg, Germany*

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0362-1/23/11...\$15.00

<https://doi.org/10.1145/3615900.3628769>

## KEYWORDS

Accident Prediction, Long Sequence Time-Series Forecasting, Transformer, Multimodal Architecture

### ACM Reference Format:

Amin Karimi Monsefi, Pouya Shiri, Ahmad Mohammadshirazi, Nastaran Karimi Monsefi, Ron Davies, Sobhan Moosavi, and Rajiv Ramnath. 2023. CrashFormer: A Multimodal Architecture to Predict the Risk of Crash. In *1st ACM SIGSPATIAL International Workshop on Advances in Urban-AI (UrbanAI '23)*, November 13, 2023, Hamburg, Germany. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3615900.3628769>

## 1 INTRODUCTION

Traffic safety is a primary global public safety concern, with an estimated 1.35 million in fatalities and 20 – 50 million in injuries yearly due to road traffic crashes. These accidents result in loss of life and damage and have significant economic costs. Predictive models can be essential in improving traffic safety by identifying high-risk areas, informing safety policies and regulations, identifying patterns and trends in accident data, and developing targeted interventions to address specific factors contributing to accidents. [24].

Given the importance of the problem, substantial research has been conducted in the field of accident risk prediction over the past few decades.

Previous studies have used various data sources to investigate and predict city-wide accident risk, including crash data, GPS data, road network attributes, land use features, and weather and population data. Examples of such studies include STCL-Net by Bao et al. [2], a crash risk prediction network by Pei Li et al. [13], unsupervised clustering and XGBoost by Shi et al. [31], Deep Accident Prediction (DAP) by Moosavi et al. [21], RiskOracle framework by Zhou et al. [42], a city-wide traffic accident risk forecasting model by Wang et al. [36], telematics data-based accident risk prediction by Hu et al. [8], and XSTNN, a Spatio-temporal Deep Learning solution by Medrano et al. [4].

The existing studies have limitations such as being only applicable in specific locations, lacking generalizability (given their input

data) [15, 30], insufficient accuracy in smaller areas [4, 37, 38], dependence on a range of data attributes that may not be available in all regions [4, 22, 27, 37], unsuitability for real-time applications [4, 21, 22, 32, 35], or using over-simplified methods [9, 14, 28]. These factors make them less practical for real-world application.

Incorporating additional information such as accident history, weather information, map images, and demographic data, is essential to a more precise prediction the risk of accidents. A more comprehensive view provides a better understanding of the various factors contributing to accidents [5, 21]. Accident history offers insight into past incidents and helps identify high-risk areas [6, 10, 33]. Weather information, such as precipitation and temperature, can indicate how the weather conditions affect the risk of an accident [21]. Map images offer a detailed view of road characteristics, and infrastructure [18, 22]. Demographic information such as population density, income levels, and age distribution help predict accident risk by identifying areas with high traffic congestion and the most at-risk groups [11]. Combining these data sources is crucial for developing effective safety measures, targeted actions, and policies to reduce accident risk [1, 12, 16, 17].

To mitigate the existing shortcomings in the literature and utilize various valuable data sources, in this paper, we introduce **CrashFormer**, a multi-modal architecture for predicting the risk of accidents for each area. CrashFormer seeks to predict the risk of accidents in different regions by using a multi-source dataset, including the history of accidents, weather information, map images, and demographic information of each area. CrashFormer consists of five components: (1) a sequential encoder, which takes time-series-based historical accident and weather data for a region, and its latent outputs contain encoded history data; (2) an image encoder, which utilizes map image data for each area to encoder characteristics of roads for a feature vector; (3) a raw feature encoder that uses demographic information of a region; (4) a feature fusion module to aggregate all the encoded latent together and (5) a classifier component that predicts the risk of accident for each location.

Our proposed approach centers on forecasting accident probabilities by utilizing historical accident and weather data presented in a sequential format. Additionally, we incorporate a comprehensive demographic dataset containing 150 attributes. This dataset was gathered through web scraping from an online portal (please refer to Section 3.3 for detailed information about the database). Furthermore, we make use of map images corresponding to the locations of accident events. To encode data from these diverse sources, we employ two advanced models. Firstly, we employ the *FEDFormer*, a transformer-based model renowned for capturing long-range relationships within sequences, to encode historical data [41]. Secondly, we utilize the Vision Attention Network (*VAN*) model to capture latent features within map images. The *VAN* model facilitates self-adaptive and long-range correlations, allowing for effective analysis of visual information when predicting accident probabilities [7].

In our experiments, we employ CrashFormer to predict accident risk (the probability of an accident) within a hexagonal area in the next 6 hours. We evaluate our proposal in 10 major cities in the United States, and the results show that our model outperforms the baselines. CrashFormer generally surpasses the next best network by 1.8% improvement in the F1-score when predicting the risk of accidents. Additionally, we observe that the proposed model shows

superior outcomes even when the training data is only available for certain parts of a city (referred to as *spatial sparsity*). This implies that the model can generalize well to new areas without much effort to tweak model parameters.

## 2 RELATED WORK

Numerous previous studies have utilized historical crash data to predict the probability of traffic accidents. However, our research highlights that accident records from the past are not the sole significant factors; meteorological data and demographic information also play vital roles in anticipating accident risks. In this section, we present a brief overview of significant research efforts in this domain.

Bao et al. conducted a comprehensive investigation using a spatiotemporal convolutional LSTM network (STCL-Net) to analyze the risk of accidents across a city. They employed a multi-source dataset that included crash data, taxi GPS data, road network attributes, land use features, and weather and population data specifically for Manhattan in New York City [2]. The study revealed that the predictive accuracy diminishes as the grid resolution increases. However, it is essential to note that since this study only uses data limited to a specific city, it may need further investigation to assess whether it is generalizable to locations with different accident patterns.

Shi et al. introduced a feature extraction and selection framework for accurate prediction of driver risk levels [31]. By extracting approximately 1300 features related to driving behaviors, the authors employed unsupervised clustering to estimate vehicle risk. To address the problem of class imbalance, under-sampling techniques were applied to the less risky data. Furthermore, critical features were identified through feature importance ranking using XGBoost. This study demonstrated the effectiveness of this approach in accurately predicting vehicle risk levels.

Ren et al. collected significant traffic accident data and investigated the spatiotemporal correlation within traffic-related data [27]. Their analysis revealed the non-uniform distribution of accident risks in time and space, exhibiting strong periodical temporal patterns and spatial correlations. The authors developed a Deep Neural Network (DNN) solution referred to as the Traffic Accident Risk Prediction Method (TARPML) to predict citywide crash risk. The reliance of this method on limited traffic accident data and its inability to predict risk at a fine-grained road level, are the shortcomings of this research.

Moosavi et al. proposed the Deep Accident Prediction (DAP) framework, which leverages multi-source data, including point-of-interest, weather data, and historical traffic events [21]. The DAP framework incorporates recurrent, fully connected, and embedding components, and a publicly available large-scale accident dataset was utilized to evaluate its performance. The study demonstrated the efficacy of integrating traffic information, time, and points of interest in enhancing real-time risk prediction. However, the study takes a massive area for prediction, which helps to improve accuracy, but it is not useful in the real world, and the real world needs to predict accurately for a small area.

Zhou et al. introduced the RiskOracle framework for more granular crash risk prediction, down to a minute level [42]. They developed the Differential Time-varying Graph Neural Network (DTGN),

which captures the dynamic nature of road networks and immediate changes in traffic conditions. To address biased risk predictions, region selection schemes were employed. The effectiveness of the proposed solution was verified using two real-world datasets.

Our proposed solution, CrashFormer, is an advanced time-series model designed to forecast the probability of accident risk in specific areas at any given time. This model is particularly advantageous for real-time applications due to its focus on a smaller geographic scale, specifically 5.161 square kilometers, surpassing previous studies in terms of granularity. Furthermore, our approach introduces a novel integration of real-time traffic events, weather events, map images, and demographic information, which remains unexplored in existing research. Another notable advantage of our methodology is utilizing readily available and easily accessible data from public sources, eliminating the need for extensive and complex data collection procedures in other studies.

### 3 DATASET

One of the important paper's contributions is offering a robust multi-source dataset for accident risk prediction. We gather data from multiple sources, use feature engineering to include additional useful features, and pre-process the data to better conform with the problem of concern i.e. predicting accident risks. Finally, we join and fuse the data from the different sources to create the dataset. The input datasets we use include information about accident history, weather information and events, demographic information for the accident area, and map images.

#### 3.1 Accident History

To leverage the accident history, we start with the dataset proposed by Moosavi et al. [20] and add features to it. It comprises comprehensive data on approximately 6.2 million accident records collected from Bing and MapQuest between 2016 and 2021 in the United States<sup>1</sup>. These records provide detailed information about various accident attributes, including time, location, severity, duration, length of traffic impact, and descriptive details. We enhance this dataset by adopting the hexagon-based city partitioning technique proposed by Monsefi et al. [18], employing the H3 library developed by Uber<sup>2</sup>.

Recognizing the significance of traffic patterns influenced by temporal factors, as highlighted by Badrestani et al. [1], we incorporated date and time considerations into our analysis. We introduced additional time-related features to capture the influence of the hour of the day, day of the week, and day of the month on traffic accident patterns. These features were extracted from the accident timestamps and appended to the dataset. Additionally, we determined the holiday status of the accident occurrence date and introduced an "isHoliday" flag to account for its potential impact on accident patterns.

Moreover, we considered the proximity of points of interest (POIs) to the accident locations. For this purpose, we utilized a one-hot vector of length 13, representing different types of POIs such as crossings, stations, amenities, bumps, giveways, traffic signals, no-exit points, railways, roundabouts, stops, traffic calming

measures, junctions, and turning loops. This information provides additional context regarding the surroundings of the accident sites and contributes to a more comprehensive analysis of accident risk factors.

#### 3.2 Weather events

To consider the weather conditions for the time accidents took place, we leverage a dataset collected by Moosavi et al. [19], containing fine-grained weather events recorded from 2016 to 2021, containing about 7.5 million records. This dataset includes information about the type of weather event in addition to its severity in a specific weather station. The features are weather type (rain, snow, fog, etc.), time, duration, severity, and location.

#### 3.3 Demographics

Demographics are the socioeconomic statistics describing information about the population and its characteristics. Demographic analysis studies a population based on gender, age, and race. The data include income, education, marriage, death, birth, and employment. Socioeconomic information is a type of coarse-grained data that is included in this study. This data was gathered for about 45,000 US postal areas for different ZIP Codes. The collected demographic information includes four other groups of information for the residents of a specific postal area: demographics, housing, employment and income, and education. For data collection, we used data crawling strategies on an online portal<sup>3</sup>. We collected 150 demographic features and created a public dataset<sup>4</sup>.

#### 3.4 Map Images

Textural information about a specific area where accidents occur could help the model generate more accurate predictions. In this study, we have collected map images from OpenStreetMap<sup>5</sup> [23] using the coordinates of accident events. In other words, each accident is mapped to a unique area (hexagonal region). Afterward, the coordinates of the center of the areas are used to fetch a map image and assign it to the accident event. The location's zoning has been based on Uber's Hexagonal Hierarchical Spatial Indexing<sup>6</sup> with a resolution of  $R = 7$ , which corresponds to hexagonal areas with about 2604 meters and 5161000 square meters of edge length and area, respectively. Also, the zoom level of the OpenStreetMap is chosen to be 14. We followed the method used by Monsefi et al. [18] to collect the images.

## 4 RESEARCH QUESTION

This section formally defines our research question by utilizing previously recorded traffic events, weather observations, and points of interest from [21]. Additionally, we gather map images using the method outlined in [18] and demographic information as outlined in Section 3.3. Given a set of accident events  $C$ , we define a geographic region as follows.

<sup>3</sup>see <https://www.unitedstateszipcodes.org/>

<sup>4</sup>see the final dataset <https://www.kaggle.com/datasets/aminkarimimonsefi/demographic-dataset>

<sup>5</sup>see <https://tile.openstreetmap.org;>

<sup>6</sup>see <https://www.uber.com/blog/h3/>

<sup>1</sup>download dataset from here [https://smoosavi.org/datasets/us\\_accidents](https://smoosavi.org/datasets/us_accidents)

<sup>2</sup>see <https://github.com/uber/h3>

**Definition 4.1 (Geographic Region).** A geographical region  $r$  is a hexagonal area, as defined in Uber h3 library [29]. This library provides levels 0 to 15 with different areas<sup>7</sup>. We have determined that a resolution level of 7 is reasonable for our work. This resolution has an edge size of 2.604 kilometers and an area of 5.161 square kilometers, making it a useful moderate region size for various applications. Additionally, we gathered demographic information by using US zip codes. For each region  $r$ , we identified the closest zip code and assigned that information to the region.

We use a database including high-resolution geographical images  $M$  represented as hexagonal tiles, as well as a dataset that contains the location of points of interest  $P$ , such as stop signs, amenities, and traffic lights for a specific area. Using these datasets, we can now define our research question as follows.

**Given:**

- A collection of geographical regions  $R = \{r_1, r_2, \dots, r_n\}$ .
- A set of time intervals with a fixed duration  $T = \{t_1, t_2, \dots, t_m\}$ , where  $|t| = 6 \text{ hours}$ , for  $t \in T$ .
- A database of accident 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  $r \in R$ .
- A single value  $K$  that shows length of sequenced data.

**Create:**

- A representation  $F_{rt}$  for region  $r \in R$  for a time interval  $t \in T$ , using  $P_r$ ,  $C_r$  and  $M_r$ .
- A binary label  $L_{rt}$  for  $F_{rt}$ , where a value of 1 indicates that the area has a high level of accident risk (if at least one accident was reported in that area) during  $t$  in the region  $r$ ; and 0 otherwise.

**Find:**

- A model  $\mathcal{M}$  for predicting  $L_{rt}$  using  $\langle F_{rt-K}, F_{rt-K-1}, \dots, F_{rt-1} \rangle$ . In other words, predict the label of the current time interval based on the observations from the  $K$  recent intervals.

**Objective:**

- Minimize the prediction error.

## 5 METHODOLOGY

In this section, we describe a model called CrashFormer, which is used to predict the level of accident risk. The model takes in three types of input: a feature vector representation of accident events, map image data, and demographic data.

### 5.1 Sequential Feature Vector

We aggregate all the accident events in a geographical region and average them over a  $t = 6 \text{ hours}$  time interval and then create a feature vector representation. We consider the following features for the accident events:

- **Weather:** We use a 8-dimensional vector to represent *Weather Severity*, *Precipitation*; and 6 additional indicators representing special weather events *Rain*, *Fog*, *Cold*, *Snow*, *Storm*, and *Hail*. We use the weather data previously obtained by Moosavi et al. [19].
- **POI:** We create a vector of size 13 for representing the existence of the point-of-interests (POIs, or map annotations)

within a radius of  $r$ . The vector contains *Amenity*, *Traffic Calming*, *Crossing*, *Bump*, *Junction*, *Give Way*, *No Exit*, *Railway*, *Station*, *Stop*, *Traffic Signal*, *Roundabout* and *Turning Loop*. The value for each item is 1 if any of the accidents at time  $t$  in area  $r$  has a nearby point-of-interest; otherwise, it is 0<sup>8</sup>.

- **Time-Related Contextual Information:** We construct a 4-dimensional vector for time information by aggregating all available data within a 6-hour window. From the accident timestamp, we extract the following information: *Day of Week*, *Month*, *Day of Month*, and *Is Holiday*. The use of this contextual time-related information should be useful, as it is believed that traffic patterns exhibit a cyclic variation based on hour, week, and month. To factor in the impact of work hours on traffic patterns, the hour of the day is computed from the accident timestamp, and the day of the week and month of the year are added as extra features.
- **Accident Information:** We create a 2-dimensional vector to represent the information related to the accident. We take the average of *Accident Severity*, collected from [21]. We also create a *label* indicating whether an accident occurred in the region  $r$  at time  $t$ .

For aggregating all the accidents that occurred during  $t$  in a region  $r$ , we take the value for *POI*, weather type, and accident label data. For instance, if we have several accident events during the time window  $t$  in area  $r$ , and if there is a specific POI near the location of one of the accidents, we set the corresponding *POI* to 1.

### 5.2 Map Image Representation

Map images are a valuable source of contextual information in understanding the causes of accidents. Accidents may be less frequent in areas with a sparse road network and more frequent in urban areas with a dense road network. To capture this information, each geographical area is associated with an image obtained from *OpenStreetMap* (OSM)<sup>9</sup> with a *zoom\_level* of 14. The map images used in the model are square-shaped and have a resolution of  $256 \times 256$  pixels. Each pixel represents an area of 9.555 meters. This means that the side size of the image is approximately 2.446 km, and its total area is about  $5.983 \text{ km}^2$ . The map images can encompass the entire geographical region  $r$ , with an edge length of  $2.604 \text{ km}$  and an area of  $5.161 \text{ km}^2$ .

### 5.3 Demographic Representation

The dataset of demographic data includes around 150 features. We removed the columns containing descriptions and took 144 features<sup>10</sup>. In this dataset, we have some information related to area  $r$  like *Population*, *Housing Units*, *Density*, *Gender*, *Race*, etc.

### 5.4 CrashFormer

We present a multi-modal architecture named *CrashFormer*, which leverages transformer-based architecture to forecast accident risk. This model incorporates a wide range of factors, including weather

<sup>7</sup>see <https://h3geo.org/docs/core-library/restable/>

<sup>8</sup>For more information about point-of-interests and their importance, see [18]

<sup>9</sup>see [https://wiki.openstreetmap.org/wiki/Zoom\\_levels](https://wiki.openstreetmap.org/wiki/Zoom_levels)

<sup>10</sup>for more information, see <https://www.kaggle.com/datasets/aminkarimmonsefi/demographic-dataset>

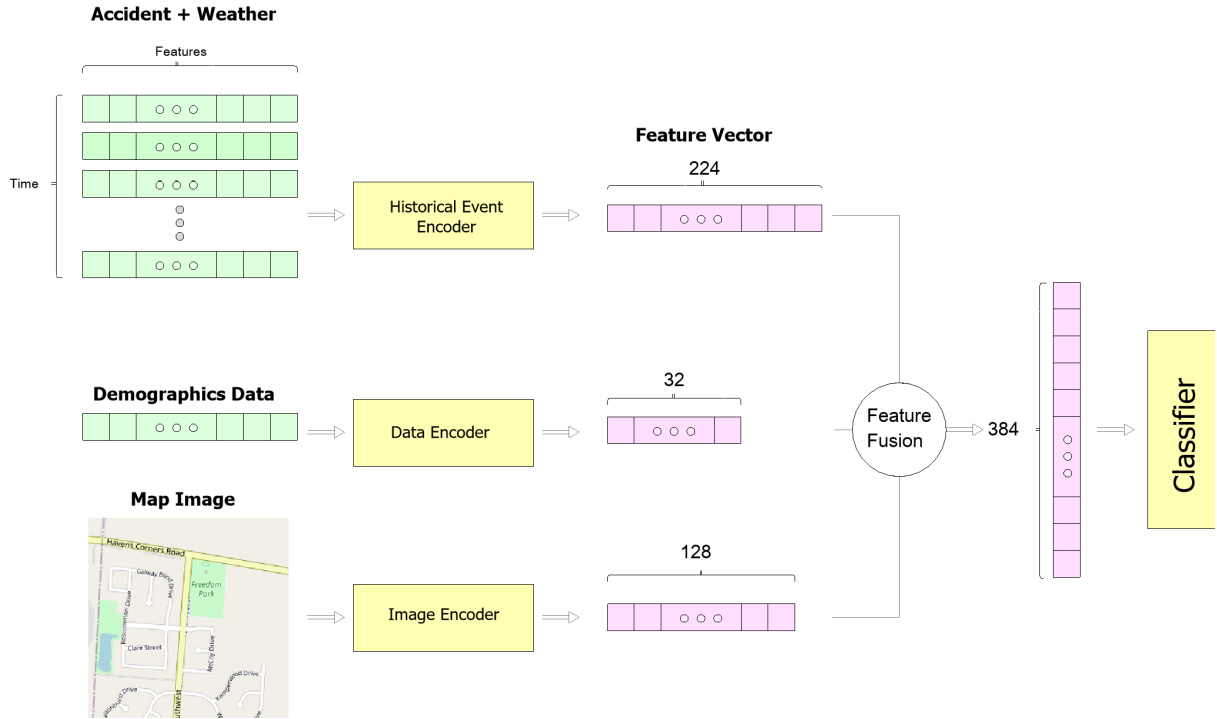


Figure 1: The architecture of CrashFormer. The sequential data based on accident and weather information, along with map images and demographic data, are each fed to a separate component, and the output feature vectors are concatenated.

Table 1: Number of hexagonal areas in each city

Houston	Miami	Los Angeles	Charlotte	Dallas	Austin	Atlanta	Phoenix	Seattle	San Diego
386	250	133	255	210	254	185	217	105	149

conditions, historical traffic data, the population density of the area, and road characteristics inferred from map images and demographic information. Building upon the findings of Najjar et al. [22], who demonstrated the feasibility of using satellite images to capture road safety similarity, we utilize map images to identify locations with similar levels of road safety based on environmental attributes, color variations (e.g., gray versus green), and the presence of specific objects (e.g., intersections, road types, roundabouts, and vegetation). By extracting visual features from map imagery, we effectively capture the characteristics and safety aspects of roads. To integrate these diverse data sources, we aggregate all relevant information within a specific location and time window, which serves as the input to the CrashFormer model. The proposed model comprises five components, each consisting of a deep neural network model. Figure 1 illustrates the architecture of CrashFormer, and the subsequent sections provide a detailed description of its constituent components.

**5.4.1 Historical Event Encoder.** We employ Frequency Enhanced Decomposed Transformer (FEDFormer), a state-of-the-art transformer-based network introduced by Zhou et al. [41], to extract the feature vector from the sequential weather and accident information. This network combines seasonal decomposition with a transformer

to capture both overall trends and detailed structures, making it a highly precise network for time-series forecasting. FEDFormer has a linear complexity to the length of the time-series sequence, making it a more efficient solution than other transformers. This network also utilizes Fourier and Wavelet blocks for capturing time-series structures by mapping data to the frequency domain. For its lower computational and memory cost and its proven accuracy on several benchmarks, we chose FEDFormer as our feature extractor for the sequential data.

The sequential data, as described in Section 5.1, is fed as input to the FEDFormer network. The input data is structured as  $(B, L_S, N_F)$ , where  $B$  represents the batch size,  $L_S$  indicates the sequence length, and  $N_F$  denotes the length of the feature vector representation, which in our case is 27. The output of this component is an encoded feature vector that encompasses valuable information about the historical records of traffic events and weather conditions. The resulting shape of the output is  $(B, 224)$ , signifying that the encoding effectively summarizes the entire historical information into a vector of size 224.

**5.4.2 Image Encoder.** We utilize VAN to extract useful road network information from map images. VAN is an attention-based model introduced by Guo et al. [7]. We used VAN – B1 for our

feature encoding. A critical attribute of attention techniques is adaptively modifying the representation based on the input characteristic. This helps us encode the map images. This module takes images with a shape of  $(B, C, H, W)$  as input, where  $B$  is the batch size,  $C$  is the number of image channels (3), and  $H$  and  $W$  are the image height and width, respectively (see Section 5.2 for details on images). The output of this component is a vector of size  $(B, 128)$ .

**5.4.3 Data Encoder.** This component is a simple multi-layer feed-forward sub-network to encode the zone-wide demographic information (explained in Section 3.3). We selected and normalized 144 demographic features and used them as input to this sub-network, which encodes the information into a small vector. The output vector size is  $(B, 28)$  ( $B$  is data batch size).

**5.4.4 Feature Fusion.** The process of consolidating latent features extracted from various sources involves amalgamating these features and generating a novel latent feature representation. This newly formed latent feature is subsequently input into the classifier layer for further analysis and classification. This approach facilitates the integration of diverse information streams, enhancing the model’s ability to make informed decisions.

**5.4.5 Classifier.** The final component of our model is a simple multi-layer feed-forward network. We combined the outputs of previous components into a single vector of size  $(B, 380)$  with a Feature fusion layer and input it into this sub-network to determine the accident risk for area  $r$  at time  $t$ . By using *softmax* as the activation function in this component, we are able to make the risk prediction.

## 6 EXPERIMENT AND RESULTS

In this section, we analyze the performance of *CrashFormer* in comparison to other leading models and baselines. We conducted our experiments using data on accidents and weather from *June 2016 to December 2021* for 10 major cities in the United States under *four* different scenarios (as outlined in subsequent sections).

### 6.1 Experiment Setup

We implemented *CrashFormer* in *Python* and used *Pytorch* [25] and *scikit – learn* [26] packages for implementing our model and baseline models. We used Ohio Supercomputer Center [3] machines for running our experiments.

We chose *Adam* optimizer to train *CrashFormer*, and selected 200 as the maximum number of epochs for training. To avoid overfitting, we applied an early stopping policy, where training will stop if the value of the loss on the validation set does not decrease after 10 consecutive epochs. The initial learning rate was set to  $10^{-3}$ , 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.,  $LR_{new} = LR \times 0.9$ ). This reduction could potentially continue until the learning rate reaches  $10^{-6}$ . We used *Binary Cross Entropy* for the loss function because finding a level of accident risk for different areas can be modeled as a binary classification problem. This loss function has been proven to work effectively for this class of problems.

### 6.2 Data Description

We trained and validated our *CrashFormer* in ten cities (see table 1). The choice of these cities was primarily to achieve diversity in traffic records (accident) and weather conditions, population, population density, and urban characteristics (road network, the prevalence of urban versus highway roads, etc.).

As explained in Section 5.4, we prepared and pre-processed data to use in the machine learning process. Each data record includes aggregated feature vector representation for a 6 hours time interval for a specific area  $r$ , and we used demographic data as raw data for area  $r$ . Also, we used a map image for area  $r$  to represent the corresponding geographical area. The goal is to predict the level of accident risk, that is, a binary classification problem. We split the cities into small areas of similar size using the *H3* library. Table 1 shows the number of regions in each city. Houston in Texas, with 386 regions, is the largest city among the selected cities. In addition, splitting time into  $t = 6$  hours means that we have four samples per day. Therefore, we have 8080 time windows for each *H3* area over the entire dataset. As one could expect, the data suffers a significant class imbalance issue, where samples with label 1 (i.e., high level of accident risk) are quite rare. For a total of 17,517,440 samples, there are only 644,322 records with a label 1 (i.e., 3.67% of the data). To mitigate this issue, we weighed the classes in the loss function and empirically found weights 15.327 and 0.516 for labels 1 and 0, respectively.

### 6.3 Baseline Models

To gauge the effectiveness of our proposed solution, we carried out a thorough evaluation encompassing both Transformer-based and non-Transformer-based models. These models represent the cutting edge in time series modeling. Our aim was to assess how our approach performs in comparison to existing methods.

- **Transformer-based models**

- **Transformer:** Vaswani et al. introduced *Transformer* [34]. Recently, Transformer-based solutions have seen an increase in long-term time series forecasting use. The transformer we used for evaluation has full attention on the encoder and decoder parts. Also, we have two layers at the encoder part and one at the decoder part.
- **Informer:** The Informer model is a deep learning model that captures long-range dependencies from long sequences of time-series data [40]. We used this model with the exact settings used in [40].

- **Non-Transformer-based models**

- **DLinear:** Recently, Zeng et al. [39] introduced one-layer linear models referred to as *LTSF – Linear* and compare their models to transformer-based solutions to forecast long-term time series data. We used their model with their recommended settings to assess its performance on our problem.

Since some baseline models are not designed to work with sequential data, the input data is vectorized before feeding these models. The vectorization process uses accident and weather events data and converts the time series data into a single vector. This is necessary as the baseline models cannot process sequential data in

**Table 2: Evaluation of CrashFormer with varying length of sequences,  $F1_1$  and  $F1_0$  means  $F1\_score$  for label one and zero**

Sequence	len=4		len=8		len=12		len=16	
	$F1_1$	$F1_0$	$F1_1$	$F1_0$	$F1_1$	$F1_0$	$F1_1$	$F1_0$
Houston	<b>0.6539</b>	<b>0.9808</b>	0.6391	0.9793	0.6286	0.9787	0.5961	0.9757
Seattle	0.5058	0.9717	<b>0.5093</b>	<b>0.9721</b>	0.4996	0.9712	0.4942	0.9706
Miami	<b>0.5822</b>	<b>0.9770</b>	0.5609	0.9750	0.5377	0.9727	0.5217	0.9711
Los Angeles	<b>0.6698</b>	<b>0.9515</b>	0.6462	0.9469	0.6297	0.9435	0.6136	0.9401
Charlotte	<b>0.6167</b>	<b>0.9798</b>	0.6015	0.978	0.5935	0.9781	0.5818	0.9772
Dallas	<b>0.6156</b>	<b>0.9792</b>	0.6034	0.9782	0.5929	0.9773	0.5845	0.9767
Austin	<b>0.5547</b>	<b>0.9829</b>	0.5444	0.9823	0.5331	0.9816	0.5213	0.9808
Atlanta	0.5509	0.9820	<b>0.5526</b>	<b>0.9821</b>	0.5496	0.9819	0.5427	0.9814
Phoenix	0.4993	0.9814	<b>0.5001</b>	<b>0.9815</b>	0.4943	0.9811	0.4857	0.9805
San Diego	<b>0.5507</b>	<b>0.9802</b>	0.5496	0.9801	0.5485	0.9801	0.5451	0.9798
Average	<b>0.57996</b>	<b>0.97665</b>	0.57071	0.97555	0.56075	0.97462	0.54867	0.97339

**Table 3: Effect of map images and demographic information on the risk of accident for that area. wo/Img means without map image. wo/Demog means without Demographic**

Model	CrashFormer		CrashFormer + wo/Img		CrashFormer + wo/Demog		CrashFormer wo/Img + wo/Demog	
	$F1_1$	$F1_0$	$F1_1$	$F1_0$	$F1_1$	$F1_0$	$F1_1$	$F1_0$
Houston	<b>0.6539</b>	0.9808	0.6446	0.9812	0.6498	<b>0.9817</b>	0.6398	0.9813
Seattle	<b>0.5058</b>	<b>0.9717</b>	0.4886	0.9706	0.4941	0.9706	0.4831	0.9695
Miami	<b>0.5822</b>	0.9770	0.5627	0.9780	0.5797	<b>0.9789</b>	0.5604	0.9780
Los Angeles	<b>0.6698</b>	0.9515	0.6546	0.9646	0.6614	<b>0.9654</b>	0.6683	0.9653
Charlotte	0.6167	<b>0.9798</b>	0.6091	0.9746	<b>0.6173</b>	0.9750	0.6072	0.9742
Dallas	<b>0.6156</b>	<b>0.9792</b>	0.6141	0.9713	0.5938	0.9689	0.5912	0.9684
Austin	<b>0.5547</b>	<b>0.9829</b>	0.5370	0.9749	0.5424	0.9756	0.5523	0.9764
Atlanta	0.5509	0.9820	0.5550	0.9653	<b>0.5572</b>	0.9814	0.5594	<b>0.9822</b>
Phoenix	0.4993	<b>0.9814</b>	0.4988	0.9808	<b>0.4998</b>	0.9808	0.4938	0.9804
San Diego	<b>0.5507</b>	0.9802	0.5195	0.9804	0.5370	<b>0.9813</b>	0.5259	0.9806
Average	<b>0.57996</b>	<b>0.97665</b>	0.56858	0.97417	0.57325	0.97596	0.56819	0.97573

their original form and must be transformed into a feature vector for input.

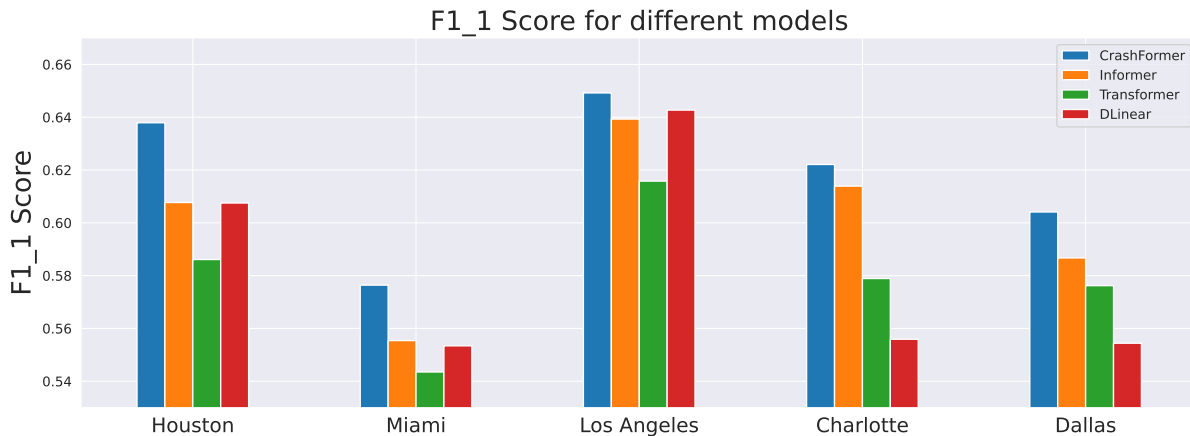
## 6.4 Results and Discussions

In this section, we conduct four experiments to evaluate the performance and generalization abilities of our proposed model, CrashFormer. The first experiment focuses on investigating the impact of hyperparameter selection on the model’s performance. We carefully tune the hyperparameters and analyze their effect on the prediction results. In the second experiment, we explore the enhancement in performance achieved by incorporating map images and demographic information as additional features. By integrating these data sources, we aim to capture more comprehensive contextual information that can improve the accuracy of accident risk predictions. The third and fourth experiments involve different data-splitting approaches to further assess CrashFormer’s performance. In the third experiment, we employ a training set consisting of data up until December 30, 2020, while using the remaining data as the test set. This scenario assumes the complete availability of accident data for all areas and allows us to evaluate the model’s performance under ideal conditions. In the fourth experiment, we simulate a real-world scenario of spatial sparsity by randomly selecting certain areas

within a city for the training set while using the remaining data as the test set. This setup reflects the challenge of making accurate predictions for an entire city when only limited data from specific areas are available. By examining CrashFormer’s performance in this context, we gain insights into its ability to generalize effectively and make reliable predictions with limited data.

**6.4.1 Experiment 1.** In Experiment I, the primary objective is to identify the optimal value for the hyperparameters of CrashFormer, with a particular focus on the length of sequential data, denoted as  $L_S$ . This experiment plays a crucial role in determining the configuration that yields the highest prediction performance.

Table 2 presents the experimental results for various choices of  $L_S$ , specifically ranging from 4 to 16. With  $L_S = 4$ , the data is divided into four consecutive time windows, each spanning a duration of 6 hours. Consequently, CrashFormer incorporates information from the past 24 hours to predict the risk of accidents for the subsequent 6-hour period. The evaluation metric employed in this experiment is the  $F1\_Score$ , which takes into account both label one (high accident risk) and label zero (low accident risk) to account for the class imbalance present in the dataset. The  $F1_1$  and  $F1_0$  scores represent the  $F1\_Score$  calculated specifically for label one and label zero, respectively.



**Figure 2: Comparing CrashFormer to baselines.  $F1_1$  denotes  $F1\_score$  for label one (high accident risk).**

To conduct this experiment, the dataset is partitioned based on regions (or areas) and time intervals. The training set comprises 70% of the data, while 10% is allocated for validation, and the remaining 20% is utilized as the test set.

Analyzing the results presented in Table 2, it is evident that in the majority of cases, utilizing data from four previous consecutive time windows ( $L_S = 4$ ) yields superior prediction performance compared to the other choices. Across all ten cities, averaging the results shows that selecting  $L_S = 4$  leads to a substantial improvement in the  $F1_1$  score. Specifically, compared to  $L_S = 8$ ,  $L_S = 12$ , and  $L_S = 16$  configurations, adopting  $L_S = 4$  demonstrates enhancements of 1.62%, 3.42%, and 5.70% in the  $F1_1$  score, respectively.

These findings provide strong evidence that CrashFormer achieves its best prediction performance when incorporating data from the past 24 hours (represented by  $L_S = 4$ ). By leveraging this time window, the model effectively captures the relevant temporal patterns and dependencies necessary for accurate accident risk prediction. As a result, for the subsequent experiments, the choice of  $L_S = 4$  is adopted as the standard setting.

**6.4.2 Experiment II.** In Experiment II, the aim is to assess the impact of incorporating map images and demographic data, in addition to sequential data, on the performance of CrashFormer. This experiment serves as an ablation study to understand the individual contributions and combined effect of these additional data sources.

To conduct the experiment, the model’s input is modified by excluding map images, demographic data, or both. The predictive performance of CrashFormer is then evaluated under each configuration. Similar to Experiment I, the data is split based on regions (or areas) and time intervals using the same strategy.

Table 3 presents the results obtained from this experiment, showcasing the impact of each data source on CrashFormer’s performance. Across most cities, the inclusion of both demographic and map images yields improvements in the model’s prediction performance. Averaging the results obtained from all ten cities, it is observed that excluding map images leads to a reduction of 2.0% in the  $F1_1$  score, excluding demographic data results in a 1.17%

reduction, and excluding both (utilizing only sequential data) leads to a 2.07% reduction.

These findings highlight the importance of incorporating map images and demographic information in CrashFormer to enhance its predictive capabilities. The addition of map images, in particular, has a more pronounced impact on the model’s predictions compared to demographic data. This suggests that spatial information captured through map imagery plays a significant role in identifying accident risk, enabling CrashFormer to better understand the geographical context and identify potential accident-prone areas.

The observed reductions in the  $F1_1$  score when excluding specific data sources underscore the competitive advantage of CrashFormer in leveraging a comprehensive set of inputs. By incorporating map images and demographic data alongside sequential information, CrashFormer demonstrates its ability to capture diverse factors that contribute to accident risk, resulting in improved prediction performance.

**6.4.3 Experiment III.** In Experiment III, the objective is to evaluate the performance of CrashFormer in predicting accident risk after a certain date, assuming complete availability of data for all areas. The dataset is split into training and test sets based on time, with the training set consisting of data from June 21, 2016, to Dec 30, 2020, and the test set containing the remaining data.

To assess the effectiveness of CrashFormer in comparison to state-of-the-art models for time series prediction, several baselines are utilized. These baselines include the Informer and DLinear. Notably, the input to all baselines in this experiment is limited to sequential data only, without the inclusion of map images or demographic information.

Figure 2 presents the results of this experiment, specifically focusing on the  $F1_1$  scores of CrashFormer compared to the different baselines for the five selected cities: Houston, Miami, Los Angeles, Charlotte, and Dallas. The comparison demonstrates the superiority of CrashFormer over the top-performing baseline models in terms of predictive accuracy.

For Houston, CrashFormer outperforms the best baseline model by 4.73% in terms of  $F1_1$  score. Similarly, for Miami, CrashFormer



achieves a 3.64% improvement over the top baseline. In the case of Los Angeles and Dallas, CrashFormer surpasses the best baseline by 1.00% and 2.73%, respectively.

The exceptional performance of CrashFormer can be attributed to several factors. Firstly, the utilization of a more advanced model architecture, namely FEDFormer, contributes to its superior predictive capabilities. FEDFormer incorporates both sequential data and additional map images and demographic information, enabling the model to capture a broader range of features and contextual factors that influence accident risk.

The inclusion of demographic and map data in CrashFormer proves to be advantageous, as it provides valuable insights into the underlying factors influencing accident occurrence. By incorporating demographic information, CrashFormer can consider population density, age distribution, and other relevant demographic factors that may impact accident risk. The integration of map images allows the model to analyze the spatial characteristics of an area, identifying road network structures, intersections, and other geographical elements that contribute to accident proneness.

**6.4.4 Experiment IV.** In Experiment IV, the objective is to evaluate the impact of spatial sparsity on the prediction performance of CrashFormer. The experiment focuses on the city of Houston, which has the highest number of areas or hexagonal zones in the dataset. The dataset is divided into training, validation, and testing sets using a similar strategy as described in Section 6.4.1.

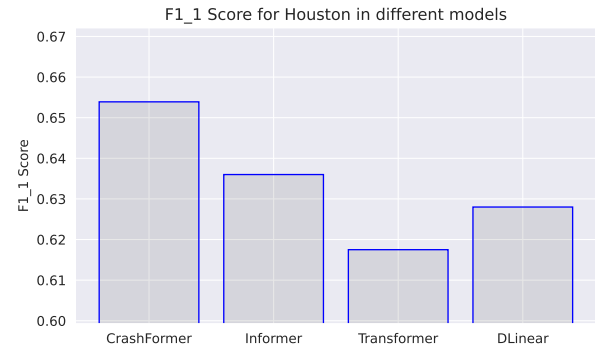
For this experiment, the performance of CrashFormer is compared to the baselines in terms of  $F1\_1$  score. The baselines employ sequential data as input, while CrashFormer utilizes data from all three sources: sequential data, map images, and demographic data. The purpose is to assess how CrashFormer’s comprehensive approach performs when faced with spatial sparsity.

Figure 3 presents the results of this experiment, highlighting the superiority of CrashFormer over the baselines. With an  $F1\_1$  score of 0.6539, CrashFormer outperforms all other baselines for the city of Houston. Specifically, CrashFormer achieves a 2.737% improvement in  $F1\_1$  score compared to the best-performing baseline model, Informer.

The enhanced performance of CrashFormer can be attributed to multiple factors. Firstly, the inclusion of demographic data and map images provides valuable additional information for modeling the spatial characteristics of accident risk. The incorporation of demographic data allows CrashFormer to consider the demographic profile of different areas, such as population density, age distribution, and other socio-economic factors, which can influence accident occurrence.

Furthermore, the integration of map images enables CrashFormer to analyze the road network structures, intersections, and other geographical features that contribute to accident proneness. This spatial awareness allows CrashFormer to capture the spatial dependencies and patterns in accident data, resulting in improved prediction accuracy even in sparsely sampled areas.

Moreover, the use of the FEDFormer model architecture further enhances CrashFormer’s performance. FEDFormer effectively integrates the sequential, map, and demographic data, enabling the model to capture complex temporal and spatial relationships for accurate accident risk prediction. The combination of these factors



**Figure 3: Comparing *CrashFormer*’ with the baselines to test the impact of spatial sparsity on accident prediction in Houston (TX)**

allows CrashFormer to generalize well to new areas and make better predictions for areas where the network has not been previously trained.

The results of Experiment IV demonstrate the superior generalization and predictive capabilities of CrashFormer when faced with spatially sparse data. By surpassing the performance of the baselines, CrashFormer showcases its ability to leverage comprehensive data sources and advanced modeling techniques to make accurate predictions in areas where limited data is available.

## 7 SUMMARY

In this study, we have presented *CrashFormer*, a transformer-based network for predicting traffic accidents by leveraging multiple data sources. Our approach combines historical accident and weather data, proximity to points of interest, demographic information, and map images to enhance the accuracy of accident risk prediction. Through rigorous experiments using real-world data from ten major cities in the United States, we have demonstrated the effectiveness of *CrashFormer* compared to state-of-the-art models. The integration of diverse data sources and the utilization of transformer-based modeling techniques have shown significant improvements in accident risk prediction.

Furthermore, the deployment and evaluation of *CrashFormer* in real-world settings would be a valuable direction for future research. Conducting large-scale experiments and assessing the model’s performance in different cities or regions can provide insights into its practical applicability and scalability.

In conclusion, *CrashFormer* offers a comprehensive and effective approach to predicting accident risk by leveraging multiple data sources and advanced modeling techniques. The future work outlined above aims to further enhance the model’s predictive capabilities and validate its performance in real-world scenarios, ultimately contributing to improved traffic safety and accident prevention efforts.

## REFERENCES

- [1] Ebrahim Badrestani, Behnam Bahrak, Ali Elahi, Adib Faramarzi, Pouria Golshanrad, Amin Karimi Monsefi, Hamid Mahini, and Armin Zirak. 2019. Real-time travel time estimation using matrix factorization. *arXiv preprint arXiv:1912.00455* (2019).

- [2] Jie Bao, Pan Liu, and Satish V Ukkusuri. 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accident Analysis & Prevention* 122 (2019), 239–254.
- [3] Ohio Supercomputer Center. 1987. Ohio Supercomputer Center. <http://osc.edu/ark:/19495/f5s1ph73>
- [4] Rodrigo de Medrano and José L Aznarte. 2021. A New Spatio-Temporal Neural Network Approach for Traffic Accident Forecasting. *Applied Artificial Intelligence* 35, 10 (2021), 782–801.
- [5] Daniel Eisenberg. 2004. The mixed effects of precipitation on traffic crashes. *Accident analysis & prevention* 36, 4 (2004), 637–647.
- [6] Bahar Farahani and Amin Karimi Monsefi. 2023. Smart and collaborative industrial IoT: A federated learning and data space approach. *Digital Communications and Networks* (2023). <https://doi.org/10.1016/j.dcan.2023.01.022>
- [7] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. 2022. Visual attention network. *arXiv preprint arXiv:2202.09741* (2022).
- [8] Xianbiao Hu, Xiaoyu Zhu, Yu-Luen Ma, Yi-Chang Chiu, and Qing Tang. 2019. Advancing usage-based insurance—a contextual driving risk modelling and analysis approach. *IET Intelligent Transport Systems* 13, 3 (2019), 453–460.
- [9] Chukwutoo C Ihueze and Uchendu O Onwurah. 2018. Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. *Accident analysis & prevention* 112 (2018), 21–29.
- [10] David Jaroszewski and Tom McNamara. 2014. The influence of rainfall on road accidents in urban areas: A weather radar approach. *Travel behaviour and society* 1, 1 (2014), 15–21.
- [11] Montaña Jiménez-Espada, Juan Miguel Vega Naranjo, and Francisco Manuel Martínez García. 2022. Identification of Mobility Patterns in Rural Areas of Low Demographic Density through Stated Preference Surveys. *Applied Sciences* 12, 19 (2022), 10034.
- [12] Kirill K Lebedev et al. 2022. ND<sub>100</sub> Estimates of Demographic Damage from Road Accidents and Opportunities for Loss Management. *Economics of Contemporary Russia* (2022).
- [13] Pei Li, Mohamed Abdel-Aty, and Jinghui Yuan. 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis & Prevention* 135 (2020), 105371.
- [14] Da-Jie Lin, Mu-Yen Chen, Hsiu-Sen Chiang, and Pradip Kumar Sharma. 2021. Intelligent Traffic Accident Prediction Model for Internet of Vehicles With Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems* (2021).
- [15] Lei Lin, Qian Wang, and Adel W Sadek. 2015. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. *Transportation Research Part C: Emerging Technologies* 55 (2015), 444–459.
- [16] Huiwen Liu, Weihua Zhang, Yajun Wang, Changsheng Wang, and Dan He. 2022. A study on the characteristics of traffic accidents of novice drivers. In *2nd International Conference on Internet of Things and Smart City (IoTSC 2022)*, Vol. 12249. SPIE, 1003–1011.
- [17] Sonu Mathew, Srinivas S Pulugurtha, and Sarvani Duvvuri. 2022. Exploring the effect of road network, demographic, and land use characteristics on teen crash frequency using geographically weighted negative binomial regression. *Accident Analysis & Prevention* 168 (2022), 106615.
- [18] Amin Karimi Monsefi, Sobhan Moosavi, and Rajiv Ramnath. 2022. Will there be a construction? Predicting road constructions based on heterogeneous spatiotemporal data. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*. 1–4.
- [19] 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.
- [20] Sobhan Moosavi, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. 2019. A countrywide traffic accident dataset. *arXiv preprint arXiv:1906.05409* (2019).
- [21] Sobhan Moosavi, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. 2019. Accident Risk Prediction based on Heterogeneous Sparse Data. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM. <https://doi.org/10.1145/3347146.3359078>
- [22] Alameen Najjar, Shun'ichi Kaneko, and Yoshikazu Miyana. 2017. Combining satellite imagery and open data to map road safety. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.
- [23] OpenStreetMap contributors. 2021. Planet dump retrieved from <https://planet.osm.org>. <https://www.openstreetmap.org/>. Accessed: 2021-1-1.
- [24] World Health Organization. 2015. *Global status report on road safety 2015*. World Health Organization.
- [25] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc., 8024–8035. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [26] 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>
- [27] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu, and Jinzhi Lei. 2018. A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. <https://doi.org/10.1109/itsc.2018.8569437>
- [28] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu, and Jinzhi Lei. 2018. A deep learning approach to the citywide traffic accident risk prediction. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 3346–3351.
- [29] Kevin Sahr, Denis White, and A Jon Kimerling. 2003. Geodesic discrete global grid systems. *Cartography and Geographic Information Science* 30, 2 (2003), 121–134.
- [30] Qiangqiang Shanguan, Ting Fu, Junhua Wang, Tianyang Luo, et al. 2021. An integrated methodology for real-time driving risk status prediction using naturalistic driving data. *Accident Analysis & Prevention* 156 (2021), 106122.
- [31] Xiupeng Shi, Yiik Dieuw Wong, Michael Zhi-Feng Li, Chandrasekar Palanisamy, and Chen Chai. 2019. A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accident Analysis and Prevention* 129 (Aug. 2019), 170–179. <https://doi.org/10.1016/j.aap.2019.05.005>
- [32] Yunzhi Shi, Raj Biswas, Mehdi Noori, Michael Kilberry, John Oram, Joe Mays, Sachin Kharude, Dinesh Rao, and Xin Chen. 2021. Predicting Road Accident Risk Using Geospatial Data and Machine Learning (Demo Paper). In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. 512–515.
- [33] JD Tamerius, X Zhou, R Mantilla, and T Greenfield-Huitt. 2016. Precipitation effects on motor vehicle crashes vary by space, time, and environmental conditions. *Weather, Climate, and Society* 8, 4 (2016), 399–407.
- [34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [35] Dhanya Viswanath, K Preethi, R Nandini, and R Bhuvaneshwari. 2021. A Road Accident Prediction Model Using Data Mining Techniques. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 1618–1623.
- [36] Dongning Wang, Balaji Gopalakrishnan, Sardar Afra, and Adam R Wiseman. 2018. A cnn model for measuring driver risk using synthetic images from accelerometer data. In *Proceedings of the 1st SIAM Workshop on Artificial Intelligence in Insurance*.
- [37] Zhuoning Yuan, Xun Zhou, and Tianbao Yang. 2018. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 984–992.
- [38] Zhuoning Yuan, Xun Zhou, Tianbao Yang, James Tamerius, and Ricardo Mantilla. 2017. Predicting traffic accidents through heterogeneous urban data: A case study. In *Proceedings of the 6th international workshop on urban computing (UrbComp 2017)*, Halifax, NS, Canada, Vol. 14. 10.
- [39] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2022. Are Transformers Effective for Time Series Forecasting? *arXiv preprint arXiv:2205.13504* (2022).
- [40] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11106–11115.
- [41] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. *arXiv preprint arXiv:2201.12740* (2022).
- [42] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Hengchang Liu. 2020. RiskOracle: A Minute-Level Citywide Traffic Accident Forecasting Framework. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 01 (April 2020), 1258–1265. <https://doi.org/10.1609/aaai.v34i01.5480>