



Advancing urban traffic accident forecasting through sparse spatio-temporal dynamic learning

Pengfei Cui ^a, Xiaobao Yang ^{a,*}, Mohamed Abdel-Aty ^b, Jinlei Zhang ^a, Xuedong Yan ^c

^a School of System Science, Beijing Jiaotong University, Beijing 100044, China

^b Department of Civil, Environmental Construction Engineering, University of Central Florida, Orlando, FL 32816, United States

^c School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China



ARTICLE INFO

Keywords:

Spatio-Temporal Prediction
Traffic Accident Prediction
Sparse Data
Hypergraph Learning
Self-Supervised Learning

ABSTRACT

Traffic accidents have emerged as one of the most public health safety matters, raising concerns from both the public and urban administrators. The ability to accurately predict traffic accident not only supports the governmental decision-making in advance but also enhances public confidence in safety measures. However, the efficacy of traditional spatio-temporal prediction models are compromised by the skewed distributions and sparse labeling of accident data. To this end, we propose a Sparse Spatio-Temporal Dynamic Hypergraph Learning (SST-DHL) framework that captures higher-order dependencies in sparse traffic accidents by combining hypergraph learning and self-supervised learning. The SST-DHL model incorporates a multi-view spatiotemporal convolution block to capture local correlations and semantics of traffic accidents, a cross-regional dynamic hypergraph learning model to identify global spatiotemporal dependencies, and a two-supervised self-learning paradigm to capture both local and global spatiotemporal patterns. Through experimentation on New York City and London accident datasets, we demonstrate that our proposed SST-DHL exhibits significant improvements compared to optimal baseline models at different sparsity levels. Additionally, it offers enhanced interpretability of results by elucidating complex spatio-temporal dependencies among various traffic accident instances. Our study demonstrates the effectiveness of the SST-DHL framework in accurately predicting traffic accidents, thereby enhancing public safety and trust.

1. Introduction

Traffic accidents represent a significant public health concern reported by the World Health Organization (WHO) in 2018, which caused 1.35 million fatalities annually (World Health Organization, 2019). The United Nations has proposed an ambitious target to achieving a 50 % reduction in road traffic casualties by 2030, which was reported in the *Global Plan for the Decade of Action for Road Safety 2021–2030*. Undoubtedly, accurate traffic accident prediction is crucial for both city governments to reach this objective and for citizens to enhance their confidence in public safety measures. Traffic accident prediction aims to evaluate the frequency or risk of traffic accidents in the near future using historical multi-source data sources. Numerous scholars have undertaken the task of traffic accident prediction utilizing diverse methodologies, encompassing traditional statistical methods, deep learning methods, etc. Traditional statistical models, such as the autoregressive integrated moving average (ARIMA) model (Avuglah et al., 2014),

support vector machines (SVM) model (Yu and Abdel-Aty, 2013) and Bayesian networks (BN) (De Oña et al., 2013), are often employed to model accident frequencies in different regions independently, which conflicts with real-world conditions. Recently, spatio-temporal neural networks employing deep learning methodologies have gained significant improvement in the field of traffic prediction. This is primarily because they offer comprehensive feature representation capabilities (Wang et al., 2022a), incorporating convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph neural networks (GNNs), among others. Nevertheless, understanding the complex nonlinear spatio-temporal relationships among traffic accidents remains challenging. Fig. 1 outlines the principal challenges that researchers must tackle to improve the accuracy of traffic accident prediction.

a) **Capturing dynamic global-local spatio-temporal dependencies.** The occurrence of traffic accidents has a dynamic, complicated spatio-temporal relevance. As shown in Fig. 1(a), the accident characteristics of a node and the correlation between nodes

* Corresponding author.

E-mail address: yangxb@bjtu.edu.cn (X. Yang).

evolve dynamically over time. Traffic accidents may be associated with adjacent or more distant but functionally similar areas (e.g., school zones, shopping centers, intersections, etc.). The correlations across various time steps might be local (e.g., from timestep t_1 to t_2) or global (e.g., timestep t_1 to t_{10}). Scholars employed hybrid prediction models based on CNN and RNN to capture the spatio-temporal characteristics of traffic accidents, including LSTM-CNN (Li et al., 2020), Hetero-Comvnet (Yuan et al., 2018), etc. However, these models can only capture spatial and temporal correlations within adjacent distances and continuous timestamps because of fixed kernel sizes. Inspired by the capabilities of graph neural networks to process non-Euclidean structured data by aggregating feature information from neighbors, researchers have constructed spatio-temporal graph structures for traffic prediction, such as the graph convolution-based approaches STGCN (Yu et al., 2018), and GMAN (Zheng et al., 2020) within the attentional graph information transfer scheme. However, a fixed graph structure can also restrict the model's capacity for learning, making it more challenging to extract the latent spatio-temporal correlations in traffic accidents. Moreover, traditional graph structures that only establish edges between two regions are inadequate for capturing the complex and hidden higher-order correlations inherent in traffic accidents. Similar studies employing graph neural networks for traffic prediction also indicate that, in the spatial domain, more complex relationships exist beyond distance attributes, such as OD relationships and node similarity (Young et al., 2021; Zhang et al., 2022a). Examining the accident count distribution of the region set depicted in Fig. 1(b), it becomes evident that a consistent distribution of accidents exists even when there are substantial distances between nodes, (as highlighted by the red, purple, and green circles). This further demonstrates the crucial role of interdependence and intricate relationships among various regions in influencing accident patterns, which cannot be adequately explained through simple graph representations.

b) Mitigating the sparse and skewed problem of traffic accident data. As shown in Fig. 1(b), unlike conventional intensive forecasting tasks for traffic flow, predicting traffic accident presents unique challenges due to sparse and skewed distribution of data. First, traffic accident datasets exhibit significant sparsity at fine-grained locations compared to entire urban areas. Statistics findings (Moosavi et al., 2019) indicate that the density of traffic accident series is sparse in most

regions (e.g., [0,0.25]), suggesting that traffic accidents may occur in a few locations in a city at specific timestamps of the day. Deep learning techniques may encounter the zero-inflation problem, where all outcomes are predicted as zero values (Wang et al., 2018a). Several scholars have proposed data enhancement strategies to enhance the prediction performance of traffic accidents. For example, Zhou et al. (2020) proposed a prior knowledge-based data enhancement (PKDE) strategy to achieve accident data enhancement in the data processing phase. Islam et al. (2021) reconstructed and enhanced the traffic accident datasets by introducing a variational autoencoder (VAE). Some scholars have also enhanced the data richness by incorporating various sources of data (e.g., traffic flow, vehicle trajectory, weather conditions, etc.) to assist in traffic accident prediction (Li et al., 2020; Yu et al., 2021). However, rather than enhancing the model's capacity to handle sparse data, these approaches focus on enriching the data's diversity to meet the model's input requirements. The utilization of graph neural networks may strengthen the impact of skewed data distribution, primarily owing to the mechanism of local neighborhood information aggregation (Liu et al., 2021b; Zhou et al., 2022a,b). Furthermore, the adoption of various graph attention-based methods that utilize attention scores between nodes and their neighbors for aggregation, which has the potential to capturing relationships among higher similarity nodes (Zhang et al., 2019). However, they also perform poorly when the label distribution of nodes is heavily skewed, as the attentional mechanisms ignoring features of rare labels (Zhang et al., 2022b). The spatial distribution of traffic accidents is observed to be skewed across distinct regions of a city, signifying that a minority of areas account for the majority of the city's accidents. Thus, conventional methods are susceptible to regions with higher traffic accident frequency and sacrifice the forecast performance of lower-degree areas.

In light of the aforementioned limitations and challenges, we propose a framework for traffic accident prediction termed Sparse Spatio-Temporal Dynamic Hypergraph Learning (SST-DHL). The framework comprises a multi-view spatio-temporal convolutional network that initially captures the local relationships among neighboring regions and timesteps. Subsequently, the dynamic hypergraph network learning is employed to capture global cross-regional traffic accident dependencies. SST-DHL constructs a hypergraph information transmission architecture that collects each region's traffic accident occurrence patterns and

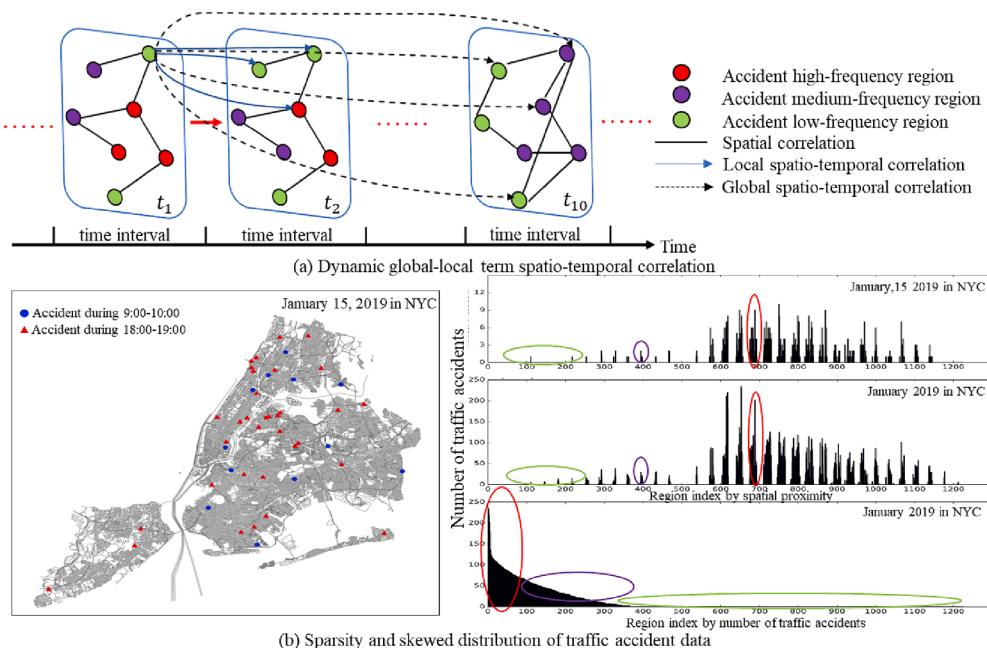


Fig. 1. Major challenges in spatio-temporal prediction of traffic accidents are depicted. Subfigure (a) illustrates the spatial and temporal graph, showcasing dynamic spatio-temporal dependencies in the network. Subfigure (b) highlights the pronounced sparsity and skewed distribution present in traffic accident data.

creates a global traffic accident representation while retaining the spatio-temporal contextual semantics. Furthermore, we employed a hypergraph contrastive learning block that facilitates collaborative supervision of local and global relational encoders, creating robust spatio-temporal representations under sparse traffic accident data. SST-DHL can effectively augment sparse data and enhance the model's capability to identify spatio-temporal traffic accident patterns across distinct locations and timestamps. Our contributions can be summarized as follows:

- i) We propose a Sparse Spatio-Temporal Dynamic Hypergraph Learning framework (SST-DHL) for traffic accident prediction. This framework addressed the problem of skewed traffic accident data distributions and sparse signals by merging hypergraph structure with self-supervised learning.
- ii) We conduct extensive experiments and rigorous analysis at different time intervals on two heterogeneous datasets. Results show that the performance of SST-DHL can obtain 7.21 %–23.09 % improvements compared to the optimal baselines. Model ablation tests also support the rationality of our suggested new framework.
- iii) The explainability of SST-DHL is thoroughly explored from the perspective of hypergraph learning. It is demonstrated that the hypergraph learning model can effectively identify the higher-order dependencies of traffic accident incidence by utilizing global dynamic hypergraph, local dynamic hypergraph, and hyperedges.

The subsequent sections of the present work are organized as follows: Section 2 introduces prior works relevant to this study. Section 3 provides the preliminaries and a comprehensive traffic accident occurrence patterns analysis. Section 4 outlines the proposed prediction framework, optimization procedure, and inference mechanism. The experiments and analysis of findings are provided in Section 5. Finally, Section 6 summarizes the recommendations and conclusions of this paper.

2. Related work

This Section discusses the existing related research from I) Traffic Accident Prediction, ii) Hypergraph learning, iii) and Self-supervised learning.

2.1. Traffic accident prediction

Many academics have worked extensively on traffic accident prediction issues. The methodologies for traffic accident prediction can be broadly categorized into two distinct groups: conventional statistical methods and machine learning methods such as deep learning.

In the realm of statistical methods, early researchers pay more attention to predicting future traffic accidents on a set of certain roads or specific regions. Some of them employed regression or other statistical models to examine relationships between variables and the risk of accidents, including the negative binomial model (Caliendo et al., 2007), zero-inflated Poisson regression model (Dong et al., 2014), cluster analysis methods (Dong and Chang, 2023), auto-regressive regression model (Bergel-Hayat et al., 2013), etc. Chang(2005) evaluated the performance of the negative binomial regression model and neural network in predicting the accident frequency on a national freeway in Taiwan. Such studies showed that crash risk might be related to factors such as vehicle speeds (Cui et al., 2023; Lemonakis et al., 2021), weather conditions (Ma et al., 2019; Tao et al., 2016), driver's attention (Kim et al., 2013), etc. Although previous research has identified many key injury indicators of traffic accidents, it faces challenges in capturing particular data characteristics such as periodicity, spatial autocorrelation, and heterogeneity. Machine learning based methods, including Random Forest (Xu and Luo, 2021), Support Vector

Regression (Dong et al., 2015), and K-nearest neighbor (KNN) (Lv et al., 2009), perform better in capturing nonlinear relations among traffic accidents. Lin et al. (2015) constructs a frequent pattern tree and Random Forest method to identify variables most likely to predict accidents. Further, they compare the accident prediction performances of the KNN and Bayesian networks along the same road. Results show that the frequent pattern tree-based Bayesian network performs best with an accuracy of 61.11 %. However, learning the complex relationships on the spatio-temporal scale is still challenging.

With the advent of artificial intelligence, researchers have been motivated to use deep learning approaches to solve spatio-temporal traffic accident prediction problems. Numerous prediction models are constructed to capture the temporal patterns by recurrent neural network (RNN) and their variants (Ma et al., 2021; Sameen and Pradhan, 2017) or the spatial patterns by convolution neural network (CNN) and their variants (Wang et al., 2021b; Zheng et al., 2019). Ren et al. (2018) constructed a risk prediction model for urban traffic accidents-based LSTM, which reveals the existence of complex spatio-temporal correlation among accident occurrences. Moosavi et al. (2019) constructed a deep recurrent neural network for citywide traffic accident risk prediction and created large-scale open-source accident datasets in the US. However, such models suffer hardship in capturing the spatio-temporal correlation of traffic accidents simultaneously. Further, different hybrid spatio-temporal prediction methods were employed for spatio-temporal traffic prediction (Ma et al., 2022; Wang et al., 2022b). For example, Li et al. (2020) developed an LSTM-CNN network by combining long-short term memory networks and convolutional neural network together to forecast crash risk. The anomalous urban event was predicted using a multi-view spatio-temporal framework (Huang et al., 2019). Yuan et al. (2018) proposed a Hetero-ConvLSTM framework to address the spatial heterogeneity challenge in traffic accident data. Nevertheless, models-based CNNs or RNNs can capture local spatio-temporal correlations and disregard long-range spatio-temporal correlations. Graph neural networks show excellent learning ability on non-Euclidean data, which could more adaptably construct the complex interactions between regions in traffic accident prediction tasks (Jin et al., 2023). For example, Xue et al. (2022) introduced a graph-based deep learning model to examine the spatial homogeneity of different sub-networks across 30 cities worldwide. Jin et al. (2023) created an automated dilated spatio-temporal synchronous graph model to record spatio-temporal correlation at various scales. Yu et al. (2021) employed similar methods to consider both external impacts and spatio-temporal correlations in large-scale heterogeneous data. The differential time-varying graph neural network and its variants were employed to examine the immediate temporal changes and dynamic inter-subregion correlations in traffic accidents (Zhou et al., 2022c; Zhou and Li, 2019).

Despite the aforementioned deep learning models performing well, several issues are still not well explored: I) Higher-order complex dependencies of traffic accidents have been ignored for global cross-regional traffic accident correlations. ii) The whole urban space involves sparse traffic accident data, which weakens the representation capability of traditional neural networks.

2.2. Hypergraph learning

Conventional graphs can only link two nodes, ignoring the higher-order complex relationships of traffic accident occurrence and cannot mitigate the data sparsity problem. Researchers developed the hypergraph learning theory to construct higher-order linkages among multiple nodes beyond pairwise connection in conventional graphs (Wang et al., 2018b). A hypergraph is a graph in which hyperedges (generalized edges) can connect to more than two nodes. Specifically, a regular graph is a particular form of a hypergraph with a fixed order of two. In the recent three years, hypergraph representation and deep learning have been successfully combined to create entirely new methodologies, which have been widely used for the fields of social networks (Alvarez-

Rodriguez et al., 2021), classification (Hong et al., 2022), and regression tasks (Gao et al., 2022). For instance, Feng et al. (2019) provided hypergraph neural networks to encode the data representations with high-order correlation. Contisciani et al., 2022 proposed Hypergraph-MT statistical inference methods to identify communities with higher-order interactions and infer missing hyperedges. A spatio-temporal self-supervised hypergraph learning was proposed for city crime prediction (Li et al., 2022), which enables crime data enhancement and city-wide crime characterization. Since hypergraph learning has produced excellent outcomes for modeling data higher-order relationships, scholars have constructed variant models based on hypergraph structure for traffic prediction (Wang & Zhu, 2022). For example, dynamic hypergraph convolutional networks were implemented for subway ridership prediction (Wang et al., 2021a), which can capture the high-order relationships between stations and passengers' travel patterns. Wang and Zhu(2022) offered an approach to the multi-source traffic prediction problem using hypergraph theory. Results show that hypergraphs are appropriate for modeling traffic data due to the complicated topology of transportation and its dynamic temporal properties. However, traffic accident prediction tasks encounter unique challenges, including the problem of sparse data and the potential higher-order complex correlation of traffic accidents. To the best of our knowledge, fewer scholars have concentrated on hypergraph structure and simultaneous modeling for traffic accident prediction tasks.

2.3. Self-Supervised learning

Recently, it has been demonstrated that self-supervised learning (SSL) can enhance the representational capabilities of neural networks, overcoming the drawback that model training mostly depends on an adequate amount of labeled data (Liu et al., 2021a). In a self-supervised learning paradigm, the model explores supervised signals from the data by aiding the learning task necessary for predicting sparse data like traffic accidents. Existing studies demonstrate that most of the highest-performing self-supervised approaches are built on contrastive learning (Hassani & Khasahmadi, 2020; Veličković et al., 2018). Their fundamental principle is to minimize representations of nodes or graphs with irrelevant semantic information and maximize those with similar semantics. In the last three years, researchers have demonstrated that the introduction of self-supervised learning can have a significant improve traffic prediction effectiveness (Lin et al., 2020). For example, Liu et al. (2022) introduced self-supervised learning into traffic flow prediction and demonstrated that additional signals using contrast learning could effectively mitigate data sparsity and enhance spatial-temporal traffic forecast performance. Ji et al. (2022) suggested that self-supervision learning, which further mines the information from the data and the model, may improve the generality and stability of the prediction model. However, Self- supervised learning has received little attention in the field of spatio-temporal traffic accident prediction. In this paper, we introduce SSL to the urban traffic accident prediction task, which can effectively mitigate the problems of sparse traffic accident data.

3. Overview

This section begins with the preliminaries and key definitions of traffic accident prediction, followed by an exploratory examination of the spatio-temporal pattern of traffic accident occurrence.

3.1. Problem formulation

The objective of traffic accident prediction is to utilize the multi-source dataset acquired beforehand to anticipate the frequency of traffic accidents in each region of the city map for the upcoming time intervals. This problem can be conceptualized as a spatio-temporal sequence forecast task from the standpoint of machine learning.

Geographical Region. Consider a dynamic system spanning a geographical area, which can be subdivided into grids using latitude and longitude coordinates, resulting in a map detailing traffic accident. Here, each grid corresponds to a specific region, serving as the primary spatial unit for determining the incidence of traffic accidents.

Traffic Accident Data. The data of traffic accident reports are collected from multi-sources, but this paper focuses on the spatio-temporal information represented as $\langle \text{timestamp}, \text{longitude}, \text{latitude}, \text{accident category} \rangle$. The geographic location of each traffic accident report is determined by its coordinates. The traffic accident data can be transformed to a spatio-temporal tensor $X \in \mathbb{R}^{R \times T \times C}$, where R , T , C denote the number of regions, time steps (e.g., hours, days, etc.), and severity category (e.g., no apparent, light, serious, etc.), respectively. The target traffic accident data $Y \in \mathbb{R}^{R \times T}$, is calculated as the total of all accidents that fall into each of the associated severity categories. In particular, $y_{r,t} = \sum_{c=1}^C x_{r,t}(c) \times c$, where $x_{r,t}(c)$ denotes the number of accidents of type c .

Task Formulation. Based on the preceding definitions, the problem of predicting traffic accidents can be stated as Fig. 2 shows:

Given the historical spatio-temporal tensor of traffic accident features $\mathcal{X} = [x_{t-i}, x_{t-i+1}, \dots, x_t]$, we want to train a model that accurately predicts the accident frequency that will occur in the following δ timestamps in all city regions. The outcomes of the forecasting can be described as $\widehat{\mathcal{Y}} = [\widehat{y}_{t+1}, \widehat{y}_{t+2}, \dots, \widehat{y}_{t+\delta}]$, $\delta \in \mathbb{N}^+$.

3.2. Pattern analysis of traffic accidents

Before predicting traffic accidents, it is necessary to establish an appropriate data structure and conduct exploratory analysis of the spatio-temporal patterns of traffic accident occurrence. Based on the definitions in Section 3.1, this section utilizes the traffic accident data in New York City for the years 2018–2019.

The traffic accident data in New York City were initially discretized in spatially and temporally. We defined the temporal interval here as one day, resulting in a total of 730 days. The spatial resolution dimension was a uniform grid of $1500 \text{ m} \times 1500 \text{ m}$, forming a total of $35 \times 35 = 1225$ geographic areas. The original database in New York City did not explicitly classify the injury severity category for traffic accidents in New York City. In the research process, we categorized severity based on the number of injured people, dividing into four categories: Type I: 0 person injured; Type II: one person injured; Type III: two person injured; Type IV: three or more person injured. This is primarily due to the fact that the spatial and temporal distributions of accidents with different severity categories differ greatly from one another. As a result, this work differs from earlier prediction tasks by taking into account the predictions for various accident categories independently.

After discretization, we obtain the tensor $X \in \mathbb{R}^{R \times T \times C}$, whose ele-

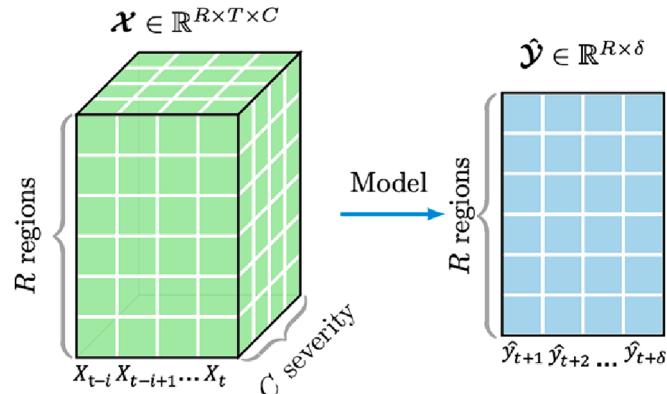


Fig. 2. Schematic diagram of the spatio-temporal traffic accident prediction task.

ments $x_{r,t,c}$ represents the count of traffic accidents occurring in region r , time stamp t and severity category c . Subsequently, we conduct an analysis of the spatial and temporal patterns of traffic accidents in New York City based on tensor X .

Spatial Distribution of Traffic Accidents. We draw the heat map of traffic accident frequency in New York City in 2019 in Fig. 3, aiming to investigate the potential relationship between accident frequency and the spatial location of regions. Fig. 3 illustrates that traffic accident incidence is non-uniformly distributed and is strongly related to a region's geographic location. Specifically major commercial and business districts tend to be the main locations of traffic accidents. In addition, the accident-prone areas are not the same for different traffic accident categories. Consequently, it is vital to consider the geographical distribution variability of different accident categories.

Temporal Pattern of Traffic Accidents. In addition to detecting changes in the geographical dimension of traffic accidents, observing their temporal evolution pattern is essential. Thus, we present the time series of traffic accidents as shown in Fig. 4. The distribution indicates the variation of the number of accidents (y-axis) of different accident severity categories with time (365 days). The time series of 2018 and 2019 are plotted in each subplot, where the green area indicates that the difference between the number of accidents occurring in 2018 and 2019 on the same day is similar. It is evident that different accident categories exhibit distinct time series trends. In contrast, for the same accident severity category, two-year accident time series are correlated. Therefore, it is essential to consider that the model needs to capture the time correlation over longer distances when predicting.

Spatio-Temporal Pattern of Traffic Accidents. The previous section explored the spatial distribution and time series of traffic accidents, respectively. However, real-world traffic accident occurrences coincide with simultaneous temporal and spatial dynamics. The pattern of traffic accident progression under simultaneous temporal and geographical changes can be discovered using the Hovmöller diagram (Martius et al., 2006), which compresses space into one dimension while representing temporal evolution in the other. Fig. 5 shows the daily accident occurrence anomalies averaged from January 2019 to December 2019, with the y-axis being time (from January 2019 to January 2020), increasing

from top to bottom) and the x-axis being longitude Fig. 5(a) and latitude Fig. 5(b), respectively. The darker blue color corresponds to a higher than average accident frequency (i.e., high accident frequency time zone), while the lighter the color, the lower the accident frequency. Notably, the spatially concentrated accident area is concentrated between the range $74^{\circ}\text{W} - 73.8^{\circ}\text{W}$ and $40.7^{\circ}\text{N} - 40.8^{\circ}\text{N}$. This is mainly because this latitude and longitude range covers two of the busiest and most expensive boroughs in New York City, Manhattan, and Brooklyn. Moreover, in terms of the temporal dimension, accidents occur most intensively from June to August, due to the summer season.

3.3. Spatio-Temporal correlation of traffic accidents

Based on the observed spatio-temporal evolution pattern of traffic accidents, we further explore the intrinsic spatio-temporal correlation of accident occurrence. Based on definitions from Section 3.1, we have traffic accident spatio-temporal vectors $X(r_i; t_j)$ for each accident severity category c , where the spatial location $R : i = 1, \dots, I$ and time steps $T : j = 1, \dots, J$. To measure the spatio-temporal correlation of traffic accidents, we initially establish the empirical mean and empirical covariance as defining metrics. The empirical spatial mean $\hat{\mu}_{x,r}(r_i)$ of position r_i and the empirical temporal mean $\hat{\mu}_{x,t}(t_j)$ of time t_j can be obtained as follows:

$$\begin{aligned}\hat{\mu}_{x,r}(r_i) &= \frac{1}{T} \sum_{j=1}^T X(r_i; t_j) \\ \hat{\mu}_{x,t}(t_j) &= \frac{1}{R} \sum_{i=1}^R X(r_i; t_j)\end{aligned}\quad (1)$$

where $X(r_i; t_j)$ represents the traffic accident frequency occurring in region r_i and time t_j . $\hat{\mu}_{x,r}(r_i)$ is the mean of traffic accidents across all time steps, and $\hat{\mu}_{x,t}(t_j)$ is the mean of traffic accidents in all regions at a time t_j . Therefore, we can use the lag covariance to determine the spatial correlation between spatial locations r_m and r_n by:

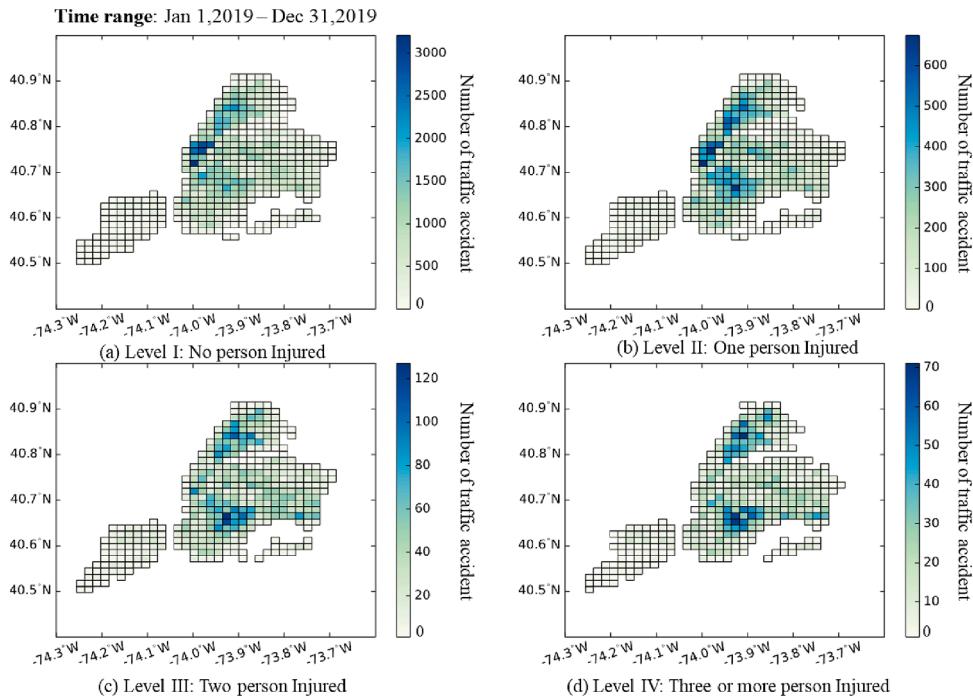


Fig. 3. The heatmap of traffic accident frequency in New York City in 2019 with $1500\text{ m} \times 1500\text{ m}$ spatial resolution. Deeper shades of blue indicate higher frequencies of traffic accidents.

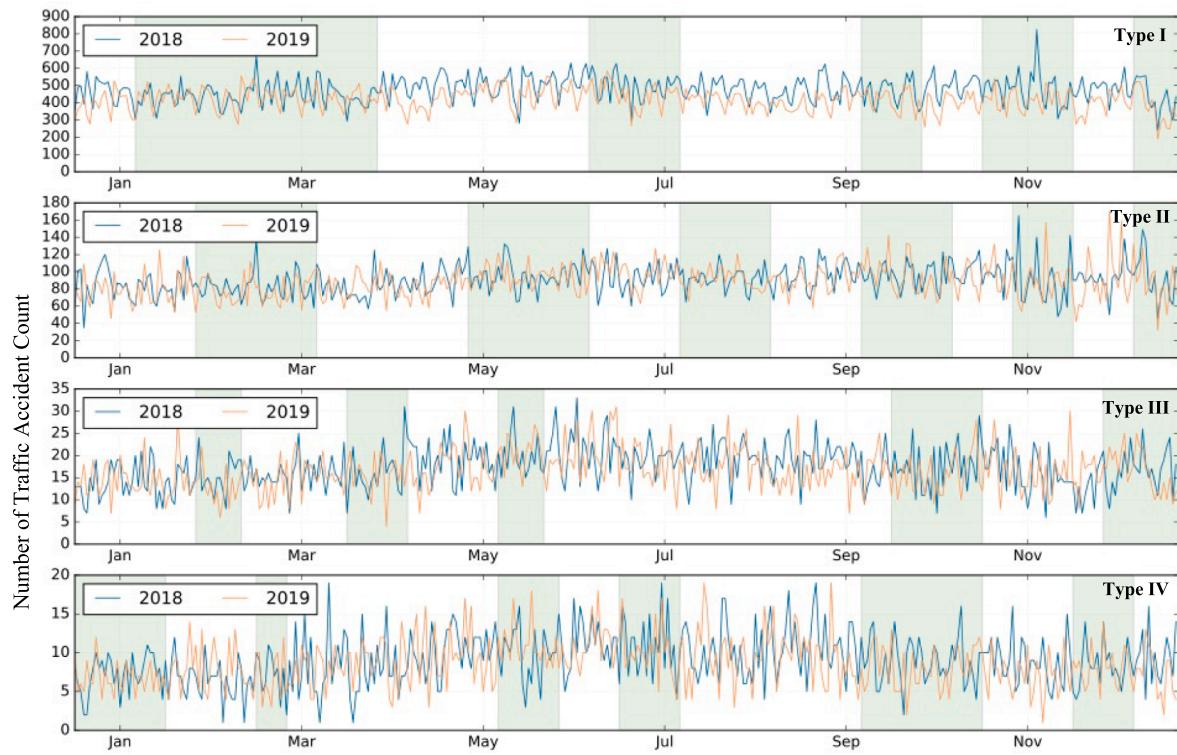


Fig. 4. Traffic accident daily time series chart in New York City in 2018–2019.

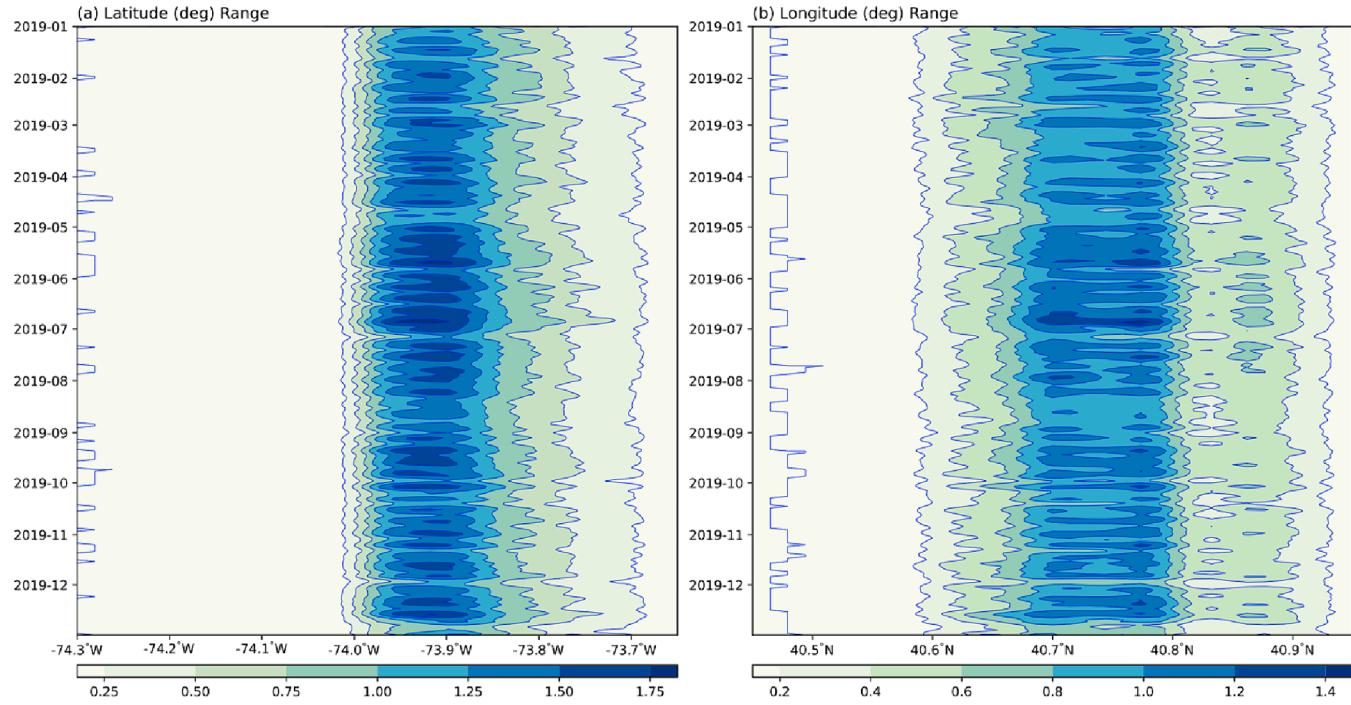


Fig. 5. Counter plots for both the longitude (left) and latitude (right) coordinates for the traffic accident in New York City. The color denotes the traffic accident count anomaly.

$$\hat{C}_x^{(\tau)}(r_m, r_n) = \frac{1}{T - \tau} \sum_{j=\tau+1}^T (X(r_m; t_j) - \hat{\mu}_{x,r}(r_m))(X(r_n; t_j - \tau) - \hat{\mu}_{x,r}(r_n)) \quad (2)$$

where $\hat{C}_x^{(\tau)}(r_m, r_n)$ denotes the spatial empirical $lag - \tau$ covariance between any two regions r_m and r_n . Further, we consider empirical spatio-

temporal covariograms to measure the combined spatio-temporal dependence across various spatial and temporal lags. In this case, we assume that the first moment (mean) depends only on the spatial and temporal lag differences, and the second moment (covariance) depends only on the first moment. The empirical spatial-temporal covariance of the spatial lag h and the temporal lag τ is then calculated as follows:

$$\hat{C}_x(h; \tau) = \frac{1}{|N_r(h)|} \frac{1}{|N_t(\tau)|} \sum_{r_i, r_k \in N_s(h)} \sum_{t_j, t_l \in N_t(\tau)} (X(r_m; t_j) - \hat{\mu}_{z,r}(r_m))(X(r_n; t_l) - \hat{\mu}_{z,r}(r_n)) \quad (3)$$

Thus, the higher value of $\hat{C}_x(h; \tau)$ denotes the higher spatio-temporal correlation of two regions with Euclidean distance h and time interval τ . $N_s(h)$ denotes a pair of regions within some distance h . $N_t(\tau)$ refers to a pair of time points separated by a temporal lag of τ . And $|N(\cdot)|$ represents the number of items included in $N(\cdot)$.

Fig. 6 illustrates the spatio-temporal correlation of traffic accidents in New York City in 2018–2019. The horizontal axis represents the Euclidean distance between two observed samples, while the vertical axis reflects the passage of time between them. A deeper color signifies a stronger spatio-temporal association. It can be seen that the spatial correlation of accidents is lower at shorter (within 10 km) and longer (beyond 20 km) distances from a spatial perspective. Notably, when traffic accidents occur within a distance of 10–20 km and a time interval of 250 days, the correlation is notably higher. This underscores the importance of considering historical accident occurrence data at appropriate scales to account for cyclicity, proximity, trends, and spatial correlation effectively.

After conducting an in-depth exploratory spatio-temporal analysis of traffic accidents in New York City, we identified several key challenges in predicting the traffic accident frequency: i) Complex spatial dependence between local and global distances. ii) Consideration of proximity, periodicity, and correlation within the time dimension. iii) The essential sparse characteristics of traffic accident data.

4. Methods

In this section, we delve into the specifics of the proposed traffic accident prediction model called Sparse Spatio-Temporal Dynamic Hypergraph Learning (SST-DHL). As shown in **Fig. 7**, the SST-DHL mainly consists of three essential components:

i) The multi-view spatio-temporal convolutional encoder provides an initialized representation to detect the local traffic accident correlation within the adjacent regions and continuous timestamps. ii) A hybrid dynamic hypergraph network incorporating higher-order interactions

across all regions, facilitating an examination of the global spatio-temporal dependence of traffic accident occurrence. iii) The two-stage self-supervised learning was employed to enhance the traffic accident pattern representation on sparse data, thereby facilitating interaction between local and global spatio-temporal patterns of traffic accidents. The details and functions of each model would be described separately below:

4.1. Traffic accident embedding layer

To generate an initial representation of traffic accident occurrences, we first develop an embedding layer. This layer involves constructing an initialized embedding $e_c \in \mathbb{R}^d$ for each accident type c . For the spatio-temporal traffic accident tensor $X \in \mathbb{R}^{R \times T \times C}$, we formulate the initial representation $e_{r,t,c} \in \mathbb{R}^d$ as follows:

$$e_{r,t,c} = \frac{x_{r,t,c} - E(X)}{\sqrt{\text{Var}(X)}} e_c \quad (4)$$

where the Z-Score normalization is employed for the initial tensor $x_{r,t,c}$, where $E(\cdot)$ and $\text{Var}(\cdot)$ represents the mean and variance of tensor X . With the traffic accident coding layer, we obtain a richer representation of the spatio-temporal tensor of traffic accidents, denoted by $E \in \mathbb{R}^{R \times T \times C \times d}$.

4.2. Multi-View Spatio-Temporal convolution encoder

Traffic accidents frequently exhibit complex patterns within metropolitan areas in real life. Therefore, we first employ a multi-view spatio-temporal convolutional encoder to capture the local traffic accident dependencies among adjacent regions and timestamps.

Spatial Traffic Accident Pattern Encoding. To map local spatial relationships of traffic accidents, we create a spatial convolutional network to accomplish this, which is formulated as follows:

$$H_{t,c}^{(R)} = \sigma(\delta(W^{(R)} * E_t + b^{(R)}) + E_{t,c}) \quad (5)$$

where $E_{t,c} \in \mathbb{R}^{R \times d}$, $H_{t,c}^{(R)} \in \mathbb{R}^{R \times d}$, represent the initial feature representation of the spatio-temporal tensor of traffic accidents and the output

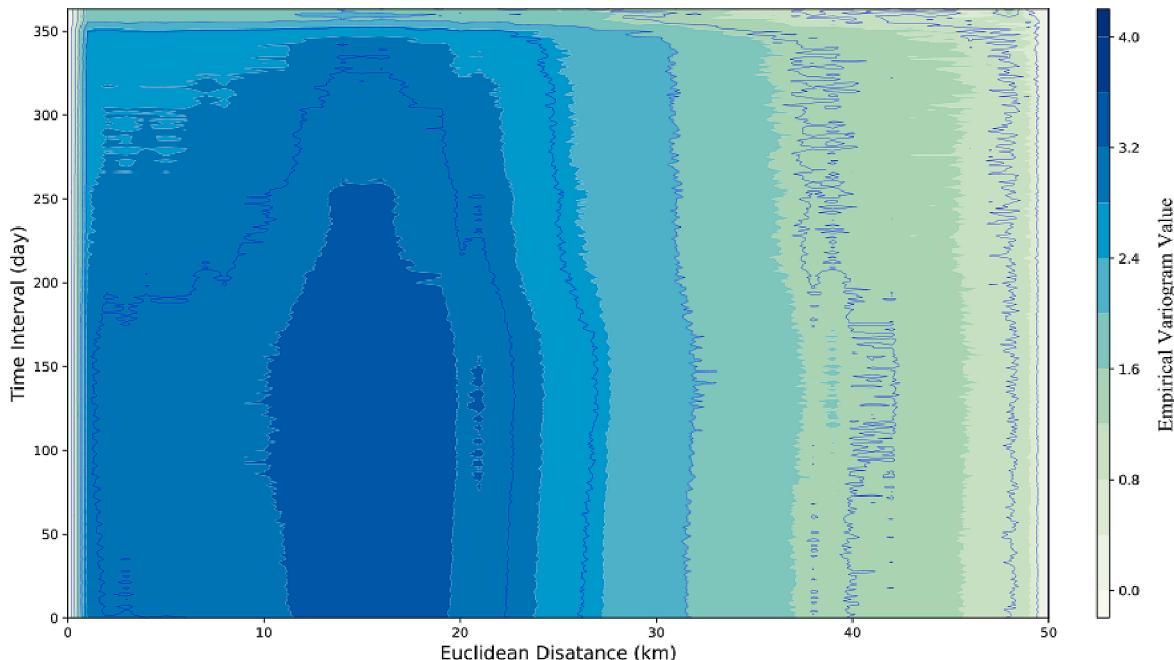


Fig. 6. Contour plot of the spatio-temporal correlation of traffic accidents in New York City (2018–2019). The Euclidean separation and time gap between the two grids are represented by the horizontal and vertical axes, respectively. The color denotes the intensity of spatio-temporal correlation.

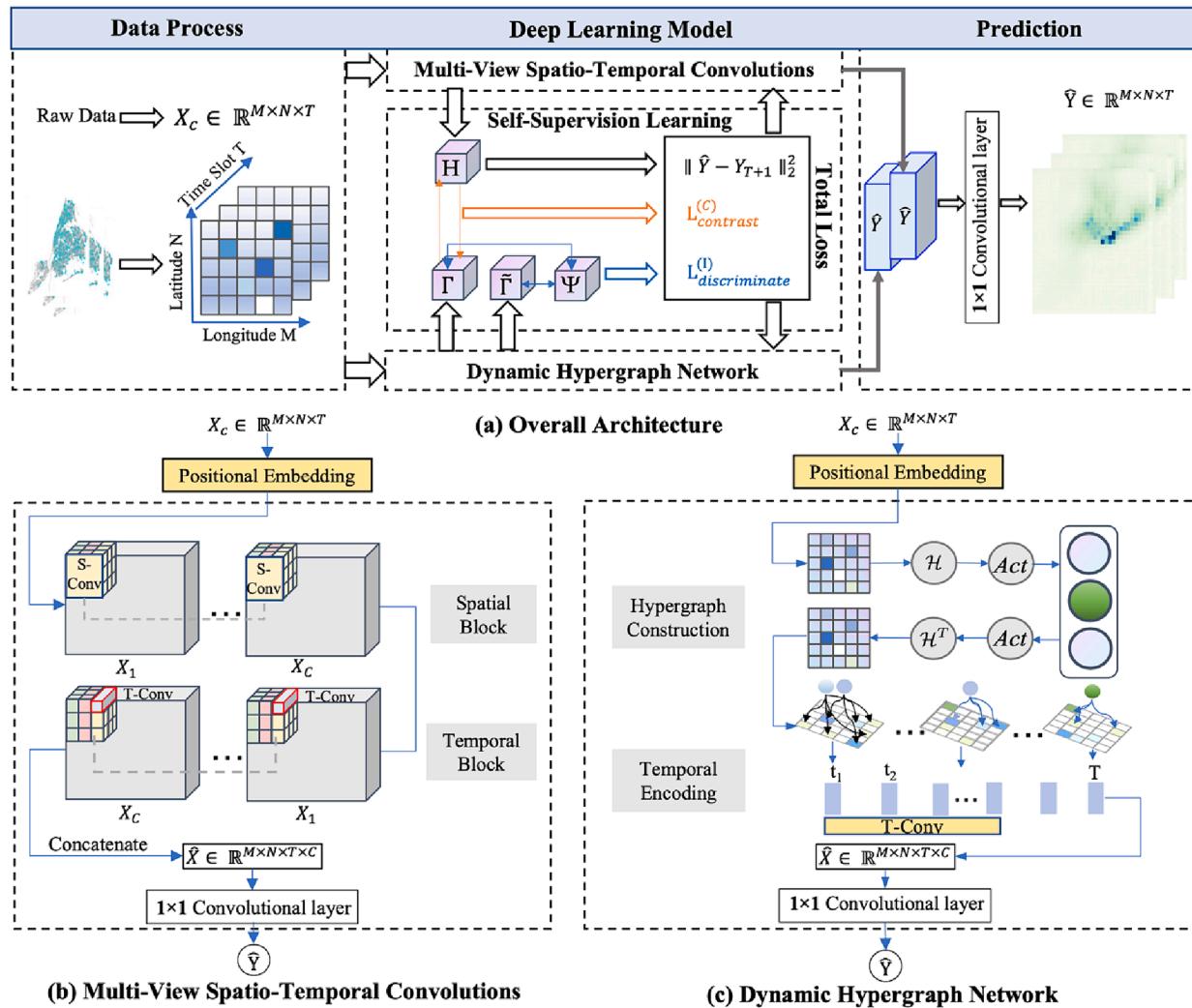


Fig. 7. Illustrative architecture of the proposed sparse spatio-temporal dynamic hypergraph learning (SST-DHL).

features after spatial convolution, respectively. $W^{(R)} \in \mathbb{R}^{L_{(M)} \times L_{(N)}}$ is the convolution kernel, where $R = M \times N$, and $b^{(R)} \in \mathbb{R}^d$ is the bias. The kernel size is denoted by $L_{(M)}$ and $L_{(N)}$ along the longitude and latitude dimensions, respectively. $*$, and $\sigma(\cdot)$ represent the process of convolution operation, dropout and ReLU, respectively. Finally, the spatial representation $H^{(R)} \in \mathbb{R}^{R \times T \times C \times d}$ is conducted with residual connection addition prior embedding $E_{r,c}$ to prevent gradient vanishing.

Temporal Traffic Accident Pattern Encoding. Further, we use a temporal convolutional network to aggregate traffic accident patterns across time to characterize the temporal dependence of traffic accident incidence across time. The formal description of the aggregating process is as follows:

$$H_{r,c}^{(T)} = \sigma(\delta(W^{(T)} * H_r^{(R)} + b^{(T)}) + H_t^{(R)}) \quad (6)$$

where $H_r^{(R)} \in \mathbb{R}^{T \times d}$, $H_t^{(R)} \in \mathbb{R}^d$ are the feature tensor learned from spatial convolution. (T) represents total timestamps. Finally, $H^{(T)} \in \mathbb{R}^{R \times T \times C \times d}$ are generated to capture the temporal-evolving patterns.

4.3. Global dynamic hypergraph network.

We used several local convolution blocks in the previous model to capture the local dependencies, yet these blocks ignored spatio-temporal correlation from distant ranges. It can be difficult to forecast sparse traffic accidents if one merely considers the local spatio-temporal

influence of the surrounding environment without considering the global spatio-temporal dependency. From a spatial perspective, regions with similar city functions (e.g., intersections, retail centers, school zones) exhibit high correlated even if they are far apart in geographic city space. From a temporal perspective, we have similar conclusions from Section 3.2 that the occurrence of traffic accidents is cyclical and dependent on longer time scales. To this end, we employ a hypergraph learning architecture inspired by (Li et al., 2022), which consists of a long temporal encoder and a regional hypergraph spatial model for capturing cross-regional global dependencies at global scales.

Dynamic Global Hypergraph Relation Encoding. In the conventional graph neural network approach, the data relationships between different nodes are pairwise. However, the data structure of traffic accidents in real situations may go beyond pairwise connection and even more complex. In the case of such multi-modal data, the situation of data association modeling may be more complicated. Unlike conventional graph, hypergraph can connect multiple nodes in a hyperedge, offering a framework to depict the complex higher-order dependencies inherent in traffic accident occurrences.

To adaptively capture the complex higher-order dependencies of traffic accident occurrence in hypergraphs, we construct trainable hypergraph structures that automatically and explicitly capture cross-regional traffic accident dependencies. First, the hypergraph is generally defined as $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}, W_t)$, where \mathcal{V} is the nodes containing all regions, \mathcal{E} denotes the set of all hyperedges, and W_t represents the

weights of hyperedges at time stamps t . The hyperedges serve as connectors among multiple regions, facilitating the representation of complex relationships. Thus, an incidence matrix $\mathcal{H}_t \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$ can be used to represent the hypergraph \mathcal{G}_t for further operations. In our research, the set of hyperedges equals the product of the number of regions and the accident severities, represented as $|\mathcal{E}| = RC$. By doing so, the higher-order connections between different regions and different severity categories can fully interact. Within our hypergraph learning framework, we elucidate global region relationships through hypergraph-guided information transfer patterns between distinct regions and hyperedges. The hypergraph messaging patterns is defined using the following form:

$$\Gamma_t^{(R)} = \sigma(\mathcal{H}_t^T * \sigma(\mathcal{H}_t * E_t)) \quad (7)$$

where, $\mathcal{H}_t \in \mathbb{R}^{|\mathcal{V}| \times RC}$ represents the weights that can be learned between nodes and hyperedges obtained by hypergraph construction. $E_t \in \mathbb{R}^{RC \times d}$ contains all accident severity degrees of the region combination vector. Note that \mathcal{H}_t corresponds to the t -th time stamp, indicating the capability of our hypergraph structure to capture the dynamic global dependence over time evolution.

Finally, the global dependencies $\Gamma_t^{(R)} \in R^{RC \times d}$ are generated by encoding the global dependencies between regions in different accident scenarios. Through accurate modeling of correlations among regions and likely hyperedge representations, regions exhibiting higher-order correlation of traffic accidents can be aligned with the input global urban occurrence patterns, accomplishing the goal of data augmentation. By doing so, the problem of the skewed distribution of traffic accident occurrences can be well mitigated.

Global Temporal Relation Encoding. Utilizing a temporal convolutional network with a fusion kernel size, we merge data in the temporal dimension to include the temporal patterns of traffic accidents into our embedding process. The process of global temporal relation encoding is defined using the following form:

$$\Gamma_r^{(T)} = \sigma(\eta(V * \Gamma_r^{(R)}) + c) \quad (8)$$

where $V \in \mathbb{R}^{L_{(R)} \times 1}$ is the trainable transformation parameter for the temporal convolution and c is the bias. For the r -th region, $\Gamma_r^{(T)} \in R^{T \times d}$ contains embedding vecots for all T time stamps.

4.4. Two-Stage Self-Supervision learning paradigm.

Previous studies have demonstrated the effectiveness of encoders in capturing both local and global spatio-temporal dependencies of traffic accidents. Our main objective is to effectively model these dependencies, particularly when confronted with sparse data, to ensure accurate forecasting. However, conventional supervised learning approaches encounter difficulties in predicting sparse data accurately. To address this issue, we propose a two-stage self-supervised learning framework within the STT-DHL paradigm. This framework consists of two key components: i) data enhancement via a hypergraph infomax network, as proposed by (Feng et al., 2019), which enriches the global semantic relationships of accident occurrences, and ii) a contrastive learning objective to facilitate interaction between local and global spatio-temporal dependencies.

Hypergraph Infomax Network. To augment the primary embedding space of traffic accident data, we developed a hypergraph infomax network, drawing inspiration from the efficacy of self-supervised learning in data enhancement (Chen et al., 2021) and the graph encoding function proposed by (Veličković et al., (2018)). This network facilitates learning the consistency of spatio-temporal representations across both node and graph levels. As our traffic accident hypergraph constitutes a single graph, we utilize a corruption function, denoted as \mathcal{C} , randomly creating a corrupted hypergraph [$\tilde{\mathcal{G}} = \mathcal{C}(\mathcal{G})$]. Specifically, we denote $\Gamma^{(R)}$ and $\tilde{\Gamma}^{(R)}$ as the local spatial embeddings derived

from the original hypergraph \mathcal{G} and the corrupted hypergraph $\tilde{\mathcal{G}}$, respectively. Subsequently, a readout function is applied to generate a global spatial representation $\Phi_{t,c} = \sum_{r=1}^R \frac{\Gamma_{r,t,c}}{R}$, which encapsulates the global information at the time step t . We further need a discriminator function \mathcal{D} as an indicator function to maximize the local mutual information:

$$D(\Gamma_{r,t,c}, \Phi_{t,c}) = \text{&sigm}\left(\Gamma_{r,t,c} * A^{(I)} * \Phi_{t,c}\right) \\ D\left(\tilde{\Gamma}_{r,t,c}, \tilde{\Phi}_{t,c}\right) = \text{&sigm}\left(\tilde{\Gamma}_{r,t,c} * A^{(I)} * \tilde{\Phi}_{t,c}\right) \quad (9)$$

where \mathcal{D} serves as an indicator function for mutual information, aiming to maximize the mutual information between the graph-enhanced representation $\Gamma_{r,t,c}$ and the global graph information $\Phi^{(R)}$. Subsequently, we proceed to train the hypergraph neural network to distinguish between the original and corrupted graphs based on the node embeddings according $\Phi_{t,c}$, with the following equation:

$$\mathcal{L}^{(I)} = \frac{1}{2N} \sum_{r=1}^R \left(\log \mathcal{D}(\Gamma_{r,t,c}, \Phi_{t,c}) + \log \left(1 - \mathcal{D}(\tilde{\Gamma}_{r,t,c}, \tilde{\Phi}_{t,c}) \right) \right) \quad (10)$$

By minimizing the generative loss $\mathcal{L}^{(I)}$, the global background information from the whole city is connected to a single region $\Gamma_{r,t,c}$ to achieve an enhanced self-supervised signal.

Local-Global Cross-View Contrastive Learning. To enhance the integration of local spatio-temporal networks with global hypergraph dependency encoders, we employ cross-perspective contrast learning, aiming to model both local and global traffic accident patterns effectively. The contrast learning model comprises two components: i) mitigating the sparsity issue inherent in traffic accident data by enabling collaborative monitoring between our local and global dependency encoders, representing two contrast viewpoints; ii) leveraging our spatio-temporal traffic accident pattern modeling to mitigate noisy information and generate improved representations of traffic accident data.

Specifically, we utilize embedding vectors from local and global relationship modeling as positive training pairs, iterating over all regions. Negative pairings are created by using embeddings from different regions within the local and global views. This approach enables us to model both local and global traffic accident patterns, facilitating cross-perspective contrast learning and the integration of spatio-temporal networks.

$$\mathcal{L}^{(C)} = \sum_{r=1}^R \log \frac{\exp(\cos(\bar{H}_r, \bar{\Gamma}_r))}{\sum_r \exp(\cos(\bar{H}_r, \bar{\Gamma}_r))} \quad (11)$$

where $\bar{H}_r, \bar{\Gamma}_r \in R^d$ represent the means of local and global embeddings produced across the temporal dimension. $\text{Cos}(\cdot)$ is the cosine similarity, quantifying the similarity between two embeddings. To enhance our SST-DHL model's ability to differentiate between locations based on their traffic accident occurrence patterns across time steps, we apply regularization to the aforementioned contrastive loss.

4.5. Model optimization

We elucidate how our SST-DHL model acquires knowledge by merging two-stage self-supervised learning tasks with a joint optimal goal. The upcoming time slot $T+1$ is predicted in our SST-DHL framework by mean prior T embeddings. The approach is officially described as:

$$\hat{Y}_r = \frac{1}{T} \sum_{t=1}^T \sum_{d=1}^d W_d \Gamma_{r,t,c}^{(T)}(d) \quad (12)$$

The set of weight parameters for model training is $W_d \in R^{d \times d}$. To construct the joint loss L , we combined the mean squared error (MSE) loss and previous loss $L^{(I)}$ and $L^{(C)}$ together. The joint loss function is

formulated as:

$$\mathcal{L} = \|\hat{Y} - Y_{T+1}\|_2^2 + \lambda_1 \mathcal{L}^{(I)} + \lambda_2 \mathcal{L}^{(O)} + \lambda_3 \|\Theta\|_2^2 \quad (13)$$

where $\hat{Y} \in \mathbb{R}^R$ denotes the predict number of traffic accidents for regions in R . The ground truth of a traffic accident's occurrence is represented by Y_{T+1} . The regularisation weights for the balancing loss are $\lambda_1, \lambda_2, \lambda_3$ in this case. Further regularisation is done using a weight-decay term. $\|\Theta\|_2^2$ represents the L_2 norm.

5. Experiments

In this section, we conduct extensive experiments on real traffic accident datasets, focusing on the following inquiries:

- **RQ1:** How well does SST-DHL perform in predicting traffic accidents compared to different baselines? (Performance Comparison)
- **RQ2:** What are the advantages of employing the hypergraph neural network with self-supervised learning components, and how do they enhance prediction accuracy? (Ablation Experiments)
- **RQ3:** How can we explain the improvement in model performance? How does the global spatiotemporal dependency captured by the dynamic hypergraph enhance the explainability of the model? (Model Explainability)

Following a description of our experimental setup, we present the evaluation findings related to the aforementioned research issues.

5.1. Experimental setting

Dataset Description. Two city accident datasets from New York City (USA) and London (UK) were adopted for the experiments. Both datasets were gathered from the Government Data Open Websites form the New York City and UK, where the main information focus is accident occurrence time, geographic location (latitude and longitude), and the accident type. Sample of two years from Jan 2018 to Dec 2019 were adopted to consider long-term patterns of accident occurrence. After data cleaning and eliminating invalid data without geographical location or time occurrence information, there were eventually 409,920 accidents in New York City and 56,284 accidents in London. Note that one might be confused by the fact that the number of accidents in New York City is 8 times higher than the number of accidents in London. This is mainly because New York City counts no person injury (property damage only) as accidents in its statistics, while the UK only counts accidents in which people were injured. Accordingly, we maintain the classification criteria employed by the respective local government agencies to identify accidents. Specifically, the 409,920 accidents that occurred in New York City were classified into four categories: 328,194 (80.24 %) in type I with no person injury, 65,368 (15.95 %) in type II with one person injury, 12,374 (3.02 %) in type III with two person injury, and 3,983 (0.97 %) in type IV with three or more person injury. Similarly, the 56,284 accidents that occurred in London were classified

into three categories: 47,415 (84.24 %) in type I named slight injury, 8,584 (15.25 %) in type II named serious injury, and 285 (0.51 %) in type III named fatal injury. The objective of this experiment is to assess the generalizability of the proposed model varying levels of data sparsity from heterogeneous data sources. Similar with (Zhou et al., 2022c), we assigned varying weights to different accident types for prediction based on the formula $y_{r,t} = \sum_{c=1}^C x_{r,t}(c) \times c$, where $x_{r,t}(c)$ denotes the number of accidents for accident type c .

Further, we apply a $1.5\text{km} \times 1.5\text{km}$ raster region to both cities and generate 1225 and 1224 distinct regions, respectively. We set the objective resolution of the forecast time period to 24 h, 4 h, and 1 h, respectively. The train, validate, and test dataset ratio is set to 6: 2: 2 based on the chronological sequence of traffic accidents. The datasets are described in detail in Table 1:

Evaluation Metrics. We employ root mean square error (RMSE), mean absolute error (MAE), and mean percentage error (MAPE) to evaluate the effectiveness of various methodologies. Following is how these metrics are expressed:

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{M} \sum_{i=1}^M (Y - Y_{T+1})^2} \\ MAE &= \sum_{i=1}^M |Y - Y_{T+1}| \\ MAPE &= \frac{1}{M} \sum_{i=1}^M \left| \frac{Y - Y_{T+1}}{Y_{T+1}} \right| \end{aligned} \quad (10)$$

where Y_{T+1} is the ground truth of traffic accident and \hat{Y} is the prediction frequency. The MAE is utilized to compute the overall error of predictions, as it is not skewed by outliers. RMSE is more sensitive to extreme values and can be employed to assess the stability of prediction results. MAPE is vulnerable to small values of the truth, which can show a level of bias. To ensure a fair comparison and minimize assessment bias, the prediction accuracy of all tested models is averaged across all test timestamps intervals. Lower values of MAE/RMSE/MAPE indicate enhanced performance in traffic accidents prediction tasks.

Baselines for Comparison. Against thoroughly testing our method, we compare SST-DHL to 8 baselines based on diverse spatiotemporal prediction solutions:

- **HA Historical averages.** We employ an average of historical traffic accident data to accomplish this purpose.
- **SVM (Shevade et al., 2000).** The distance between the samples and the hyperplane is minimized during SVM training by using a kernel function.
- **LSTM (Ren et al., 2018).** The temporal patterns of traffic accidents are captured by long short-term memory networks with LSTM hidden units that are tightly connected.
- **ConvLSTM (Shi et al., 2015)** It employs the convolutional long short-term memory (ConvLSTM) neural network model to conduct an

Table 1
Dataset description used for experiments.

Dataset	M ¹	N ¹	R ¹	Time span	Instance	Time interval	Timestamps	Sparisity (%) ²
New York City	35	35	1225	Jan 2018 to Dec 2019	409,920	24 h 4 h 1 h	730 4380 17,520	18.55 % 5.56 % 1.62 %
London	36	34	1224	Jan 2018 to Dec 2019	56,284	24 h 4 h 1 h	730 4380 17,520	5.71 % 0.98 % 0.24 %

¹ M and N represent the quantity of regions adjusting to the spatial row-and-column dimensions in a city's region map. R = M × N represents the number of regions in city.

² The sparsity rate of accident tensor X ∈ ℝ^{R×T×C} is calculated by dividing the number of non-zero elements (accidents happened) by the total number of elements.

- extensive investigation on the spatio-temporal prediction problem. Yuan et al. (2018) also used this framework to predict traffic accidents.
- **ST-Transformer** (Xu et al., 2021). A unique spatio-temporal transformer networks (STTNs) was used to obtain both long-range temporal dependencies and dynamically directed spatial dependencies.
 - **DCRNN** (Li et al., 2018). The method combines a sequence-sequence learning framework with a diffusion convolution operation to forecast temporal and spatial dynamics by simulating them using a diffusion mechanism.
 - **ASTGCN** (Guo et al., 2019). The approach utilizes spatio-temporal graph convolutional networks, incorporating attention mechanisms to identify spatial patterns and extract temporal properties.
 - **AdapGL** (Zhang et al., 2022). To understand the complicated dependencies, an adaptable graph learning method based on graph convolutional networks is used.

Hyperparameter Settings. Adam optimizer is used to optimize our SST-DHL with a learning rate of 0.001. We look for hidden states in the dimensions for the range $\{2^2, 2^3, 2^4, 2^5\}$. We set up SST-DHL for cross-region embedding propagation with 128 hyperedges for our dynamic hypergraph learning component. In order to model long-term temporal context, we stack four convolutional layers. There are four options for batch size: $\{4, 8, 16, 32\}$. Regularization terms $\lambda_1, \lambda_2, \lambda_3$ have a weight ranging from 0 to 1.

5.2. Performance comparison (RQ1)

To evaluate the effectiveness and robustness of the proposed models, we constructed experiments on two heterogeneous traffic accident datasets in New York City and London with different time intervals. Considering the actual prediction significance of traffic accidents, we divide the time intervals into 24 h, 4 h, and 1 h. The sparse ratio of traffic accidents is indicated in parentheses, representing the ratio of non-zero occurrences (traffic accident happened) to the total size of the accident tensor in the spatio-temporal accident tensor.

Table 2
Overall performance of traffic accident prediction in NYC in terms of MAE&RMSE&MAPE.

Model	Dataset	New York City									
		Time Interval	24 Hours (18.55 %)			4 Hours (5.56 %)			1 Hour (1.62 %)		
			Metrics	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE
HA		1.7203	2.4342	83.82	1.8903	2.3024	92.45	1.4728	1.6614	99.57	
SVM (2000)		1.6302	2.3704	76.34	0.0784	0.3452	99.97	1.1532	1.2342	100.0	
LSTM (2018)		1.3661	1.7758	62.69	0.0784	0.3452	99.97	1.1532	1.2342	100.0	
ConvLSTM (2018)		1.4478	2.0429	57.76	0.0789	0.3521	99.52	1.1378	1.2224	99.93	
ST-Transformer (2021)		1.4586	2.1294	54.71	0.0782	0.3501	98.25	1.1389	1.2238	100.0	
DCRNN (2018)		1.3907	1.9224	57.89	0.0732	0.3530	99.93	1.1387	1.2232	100.0	
ASTGCN (2021)		1.5377	1.8752	95.52	0.0791	0.3505	98.93	1.1352	1.2199	99.68	
AdapGL (2022)		1.2709	1.7445	55.63	0.0972	0.3381	92.91	1.1381	1.2225	99.95	
SST-DHL		1.0975	1.5206	49.03	1.3610	1.8118	71.45	0.9923	1.0286	92.74	
Improvements		+13.64 %	+12.83 %	+11.86 %	—	—	+23.09 %	+12.81 %	+15.86 %	+7.21 %	
Model	Dataset	London									
		Time Interval	24 Hours (5.71 %)			4 Hours (0.98 %)			1 Hour (0.24 %)		
			Metrics	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE
HA		1.5434	1.8322	99.98	1.2432	1.3543	100.0	1.3724	1.2857	100.0	
SVM (2000)		1.3201	1.5028	99.97	1.2017	1.4703	100.0	1.1972	1.3024	100.0	
LSTM (2018)		1.2788	1.4121	99.95	1.2012	1.3010	99.99	1.1924	1.2889	100.0	
ConvLSTM (2018)		1.2774	1.3934	99.80	1.1719	1.2411	99.99	1.1538	1.2089	100.0	
ST-Transformer (2021)		1.2775	1.3935	99.81	1.1718	1.2412	99.99	1.1540	1.2090	100.0	
DCRNN (2018)		1.2798	1.3956	99.99	1.1719	1.2411	99.99	1.1539	1.2089	100.0	
ASTGCN (2021)		1.2735	1.3891	99.54	1.1719	1.2410	99.99	1.1538	1.2089	100.0	
AdapGL (2022)		1.2929	1.4068	101.1	1.1603	1.2301	98.93	1.1541	1.2092	100.0	
SST-DHL		1.0681	1.2041	82.44	1.1646	1.2375	98.45	1.1443	1.2011	99.69	
Improvements		+16.12 %	+13.32 %	+17.18 %	—	—	+0.4 %	+1%	+0.6 %	+0.4 %	

Performance comparisons between our SST-DHL method and baselines for predicting traffic accidents are provided in Table 2. Further, to gain deeper insights into the model prediction results, Fig. 8 illustrates the comparison of different model prediction results and the real accident distribution for different time intervals in New York City. Note that to facilitate a more intuitive comparison of prediction effects across various time intervals, we standardized the predicted and actual values. This allows us to assess the model's performance without distorting the distribution and relative magnitudes of the results. Ultimately, in Fig. 8, colors denote the normalized intensity of traffic accidents within each region at the corresponding time step, where darker shades signify a greater occurrence of accidents. Our main conclusions are as follows:

i) Table 2 demonstrates that our model outperforms baselines for different time intervals for both datasets, with an average performance improvement of 7.21 %–23.09 %, particularly evident when the data sparsity exceeds 1 %. In particular, baselines seem to perform better than our model in the New York City dataset (4 h interval), which has lower MAE and RMSE values. This is because the baselines tend to predict all region's accident rates as 0, thus causing overfitting. While our model can stably learn the spatio-temporal patterns of traffic accidents, thus improving the metric of MAPE by 23.09 %. The proposed model has not consistently improved when predicting London accidents at 4 Hours and 1 Hour interval due to the extreme sparsity (below 1 %) in such cases. More tests have been conducted to demonstrate that our model significantly outperforms the conventional baselines when the range of data sparsity is 1 %–20 %.

ii) It is evident from Fig. 8 that our model generates superior spatio-temporal representation patterns across different levels of sparsity, which is more prevalent in actual applications. SST-DHL can distinguish the global regional relative risk, leading to improved predictive performance in critical geographic locations. Comparing our SST-DHL to the state-of-the-art baseline model, areas with relatively low traffic accident rates continue to benefit from significant enhancements.

In every scenario, SST-DHL outperforms alternative spatio-temporal prediction methods, demonstrating the superiority of our model for spatio-temporal dependent models of traffic accident prediction. Despite

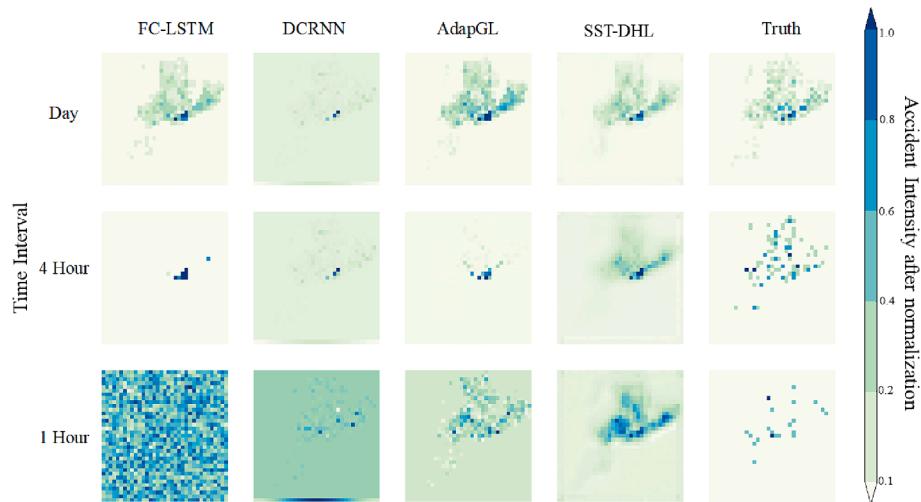


Fig. 8. Visualization of various models forecast results over regions in the entire city space of New York City.

the fact that GNN-based algorithms (AdapGL, ASTGCN, and DCRNN) capture spatial linkages through the propagation of higher-order information on area graphs, they solve traffic accident prediction difficulties inside supervised learning frameworks. However, poor supervised labeled data hinders their capacity to create high-quality spatio-temporal representation. So, the integration of hypergraph learning and self-supervised tasks is necessary to improve the encoding performance of complex spatio-temporal traffic accident patterns because conventional graph neural networks struggle to handle predictions from sparsely distributed traffic accident data.

Specifically, we attribute these significant improvements to: i) the capacity to capture the overall traffic accident pattern and maintain the global spatio-temporal signal across the city region, due to our hypergraph-dependent encoder. ii) The developed two-stage self-supervised learning paradigm incorporates additional self-supervised signals that provide comprehensive spatio-temporal representations in situations of sparse traffic accident data.

5.3. Model ablation and effectiveness analyses (RQ2)

Additionally, we perform a comprehensive investigation into the contributions of individual components to the predictive performance of SST-DHL. Specifically, model ablation experiments were conducted to examine the advantages of two crucial components of SST-DHL, named as multi-view spatio-temporal encoder and two-stage self-supervised learning progress. **Table 3** gives the results in detail.

Multi-View Spatio-Temporal Convolution. We created three variations to evaluate the efficacy of our spatio-temporal convolutional network for representing multi-view interdependence. As shown in

Table 3, the local relational representation in spatial and temporal is disabled by the notations “w/o S-Conv,” and “w/o T-Conv0,” respectively. The complete multi-view local encoder is eliminated when using “w/o Local”. By contrasting it with the aforementioned model alternatives, the findings in **Table 3** show that SST-DHL successfully extracts knowledge from many viewpoints of traffic accident data that correspond with geographical, temporal, and semantic information. Each specific view’s semantic encoding is complimentary to the others, which improves the performance of traffic accident prediction as a whole.

Two-Stage Self-Supervised Learning. In addition, we conducted ablation experiments to determine the effectiveness of a two-stage self-supervised learning architecture for enhancing the spatio-temporal representation of traffic accidents based on regional self-identification. In this part, we develop five variations of the comparison procedure as shown in **Table 3**:

i). “w/o Hyper.” Instead of examining regional hypergraph relations, we depend exclusively on local spatial encoders for prediction. ii). “w/o GlobalTem.” We do not examine global temporal encoders but instead rely on local temporal encoders for prediction. iii). “w/o Infomax.” The hypergraph infomax network is deleted in order to perform global context injection of self-supervised auxiliary signals. iv). “w/o ConL.” In our system, deactivate bracketed cross-view contrast learning for interaction between local and global spatially dependent encoders. v). “w/o Global.” There is no two-stage self-supervised learning in this variant, we only use multi-view spatio-temporal convolutional models for prediction.

As shown in **Table 3**, the adoption of our self-supervised learning significantly enhances SST-DHL’s performance in the majority of the tested scenarios. This once more emphasizes the benefits of using self-supervised signals to successfully handle the sparsity issue and skewed distribution of traffic accident data, leading to enhanced spatiotemporal representations. Our hypergraph information network reinforced the global regional dependency by introducing a self-supervised learning task into the embedding space of traffic accidents. The improved performance is attributed to the fact that contrastive learning allows probing important information directly from the data, promoting the generation of more robust feature representations. Thus, the primary representation problem and the improved representation job cooperate to provide superior global regional embeddings.

5.4. Deep analysis of model explainability (RQ3)

In this section, we demonstrate the ability of our SST-DHL model to explain the high-order complex correlations of traffic accident occurrence. Specifically, we address the explainability of the model through

Table 3
Ablation experiments on the SST-DHL paradigm.

Model	New York City			London		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
w/o S-Conv	1.1823	1.3204	57.82	1.2422	1.3453	84.32
w/o Hyper	1.2021	1.4234	58.64	1.2301	1.3254	85.53
w/o Local	1.2413	1.5233	60.72	1.1503	1.5362	88.21
w/o Hyper	1.2984	1.5815	59.26	1.2303	1.3827	90.08
w/o GlobalTem	1.3001	1.7302	67.32	1.1923	1.3302	88.21
w/o Infomax	1.1981	1.6502	65.12	1.1728	1.2314	87.42
w/o ConL	1.1823	1.5901	62.21	1.2232	1.3342	91.23
w/o Global	1.3233	1.6101	57.82	1.2421	1.4032	92.34
SST-DHL	1.0975	1.5206	49.03	1.0681	1.2041	82.44

global dynamic hypergraph, local dynamic hypergraph, and hyperedges, respectively.

Global Dynamic Hypergraph Network. A learnable dynamic hypergraph is trained by our SST-DHL to capture the higher-order correlation of the spatio-temporal distribution of traffic accidents. The dynamic hypergraph is defined as $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$. In the initial state, we define each hyperedge to be able to connect all regions, and then learn the weights of different regions within each hyperedge adaptively. Finally, the incidence matrix $H_t \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{V}|}$ are well learned through our SST-DHL model. Fig. 9 shows the dynamic hypergraph, where Fig. 9(a-c) is the dynamic hypergraph \mathcal{G}_t at different time and Fig. 9(d) is the global hypergraph $\mathcal{G} = \sum_{t=1}^T \mathcal{G}_t$. Each hypergraph consists of $35 \times 35 = 1225$ regions and 128 hyperedges, where the node size is the frequency of accident occurrence, and each hyperedge (red line) connects the top five highest correlation regions learned by SST-DHL. Within 128 hyperedges, we can dynamically capture the complex high-order correlations of traffic accidents occurring near and far among all city regions.

Local Dynamic Hypergraph Network. It should be noted that the location relationships of regions in Fig. 10 are not the actual Euclidean distances, but the regions with high relevance after hyperedge capture are automatically clustered. As shown in Fig. 10(a), we construct a local dynamic hypergraph by randomly selecting 20 hyperedges in the global dynamic hypergraph for the exposition. Each of the hyperedges contains 5 highest-relevance regions, including 95 regions in total. It means that the same regions may also belong to different hyperedges at the same time. Further, we plotted the heat map in Fig. 10(b) to preserve the latitude and longitude location relationships of the regions. We can observe a clear representation of how the dynamic hypergraph captures higher-order correlations across long distances. Specifically, it captures regions with similar accident occurrence patterns, even if they are distant from each other. For example, regions colored by green capture regions with a higher frequency of accident occurrence, and regions colored by gray contains the lower frequency regions.

Deep Understanding of Hyperedge. The core of the hypergraph is to describe the higher-order correlations among regions through hyperedges. As shown in Fig. 11, we discuss the similar region features extracted from the hyperedge by sampling the hyperedge in terms of accident severity, temporal dimension, and spatial dimension, respectively.

As shown in Fig. 11(a), we first sample six hyperedges ($\epsilon_{10}, \epsilon_{25}, \epsilon_{40}, \epsilon_{61}, \epsilon_{120}, \epsilon_{120}$) and generate a matrix of 4×5 for each hyperedge. Each row in the matrix represents the top 5 globally most relevant regions extracted by this hyperedge for a specific traffic accident category. To maintain consistency in the analysis, min-max normalization is applied to all data. The shades of color in Fig. 11 reveal that the regions with high correlation at different accident categories have similar traffic accident occurrence patterns. This result validates the ability of our hypergraph neural network to capture the global dependence of different geographical regions under different accident categories.

Further, without loss of generality, we select four hyperedges ($\epsilon_{40}, \epsilon_{61}, \epsilon_{120}$) and generate similar feature matrices for each hyperedge over time. In Fig. 11(b), a total of $4 \times 10 = 40$ matrices are displayed,

where the meaning of each matrix is the same as in Fig. 11(a). We can find that each hyperedge captures similar traffic accident occurrence patterns in highly correlated regions within the same day. The dynamic hyperedges also capture the global dependencies with temporal evaluation, which can be illustrated by the fact that the color of the correlation matrix changes as time shifts.

In addition, we visualized the correlation matrix among all regions within a particular hyperedge as shown in Fig. 11(c). As depicted by hyperedges $\epsilon_{16}, \epsilon_{30}$, and ϵ_{46} , various hyperedges represent distinct higher-order characteristics of traffic accident incidence. For example, the global regions' higher-order features captured by the super edges ϵ_{16} and ϵ_{30} are similar to the true distribution of traffic accidents. While the higher-order features captured by the hyperedges ϵ_{46} may be higher-order dependencies not observed from the original traffic accidents.

6. Conclusions

In this research, we propose a Sparse Spatio-Temporal Dynamic Hypergraph network (SST-DHL) to predict accident frequency in city level. Initially, we transformed the traffic accident dataset into a tensor format suitable for spatio-temporal prediction. Subsequent analysis of traffic accident occurrence patterns confirmed the presence of sparsity and complex spatio-temporal correlation features in traffic accidents. To this end, the proposed SST-DHL contains three key elements that work together to increase the accuracy of traffic accident prediction. First, we employed a multi-view spatio-temporal encoder to capture the local spatio-temporal characteristics of traffic accident occurrence from the adjacent regions and continuous timestamps. Second, a hybrid dynamic hypergraph network is conducted to examine the global spatio-temporal dependence of traffic accident occurrence. Third, a two-stage self-supervised learning paradigm is introduced on the hypergraph structure, which enhances the representation of sparse traffic accident data and can effectively detect the complex higher-order correlations on traffic accidents. The efficacy of our SST-DHL approach is demonstrated through multi-timescale experiments conducted on two real-life urban accident datasets.

The findings indicate that our proposed SST-DHL outperforms other baselines in several aspects. Firstly, it exhibits an average performance improvement ranging from 7.21 % to 23.09 % compared to baselines such as LSTM, ConvLSTM, ST-Transformer, ASTGCN, AdapGL, etc. Second, the proposed SST-DHL achieves superior performance on the New York City accident dataset and also stable performance on the new heterogeneous dataset with different accident determination criteria from London, which confirms the robustness and generalizability of our methods for different sparse dataset. Additionally, an in-depth analysis of model interpretability from a hypergraph learning perspective reveals that our model not only enhances prediction accuracy but also produces interpretable learned weights that represent the underlying reasons for improved accuracy. In summary, this study successfully predicts accident frequency using a spatio-temporal deep neural network, highlighting the efficacy of introducing a Dynamic Hypergraph Network to enhance higher-order correlations among traffic accidents. Furthermore, self-supervised learning proves effective in enhancing pattern

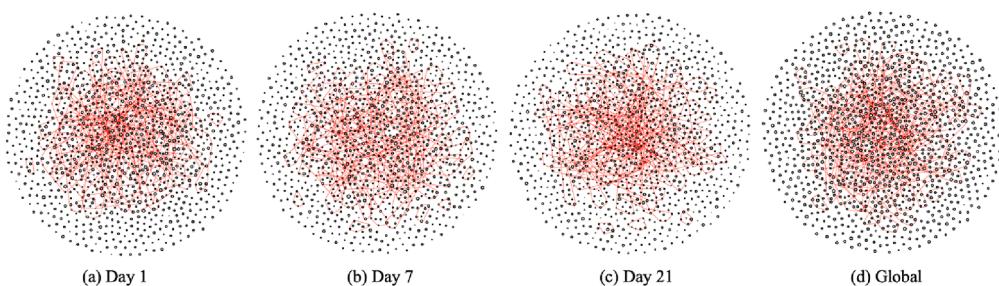


Fig. 9. Visualization of global dynamic hypergraph network for all regions across the temporal dimension.

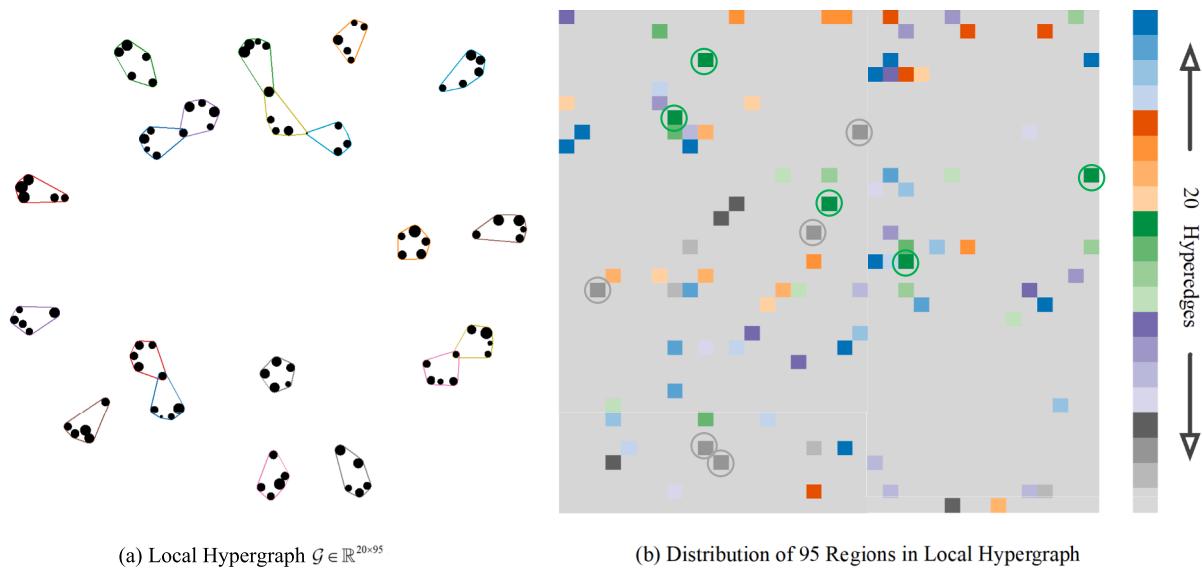


Fig. 10. Visualization of local hypergraph network.

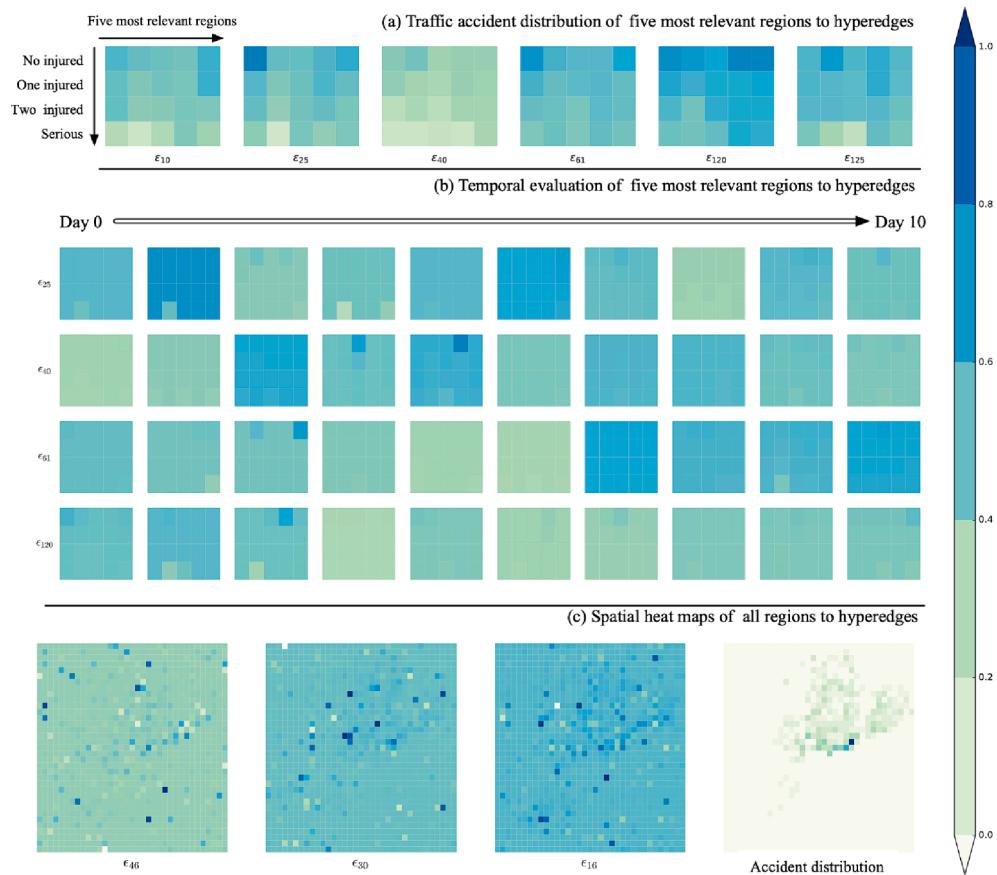


Fig. 11. Visualization relevance weights of hyperedges from hypergraph learning. (a) Type of accident distribution of five most relevant regions to hyperedges, (b) Temporal evaluation of five most relevant regions to hyperedges, (c) Spatial heat maps of all regions to hyperedges.

representation on large sparse accident datasets. These findings have implications for the development of advanced traffic management systems aimed at reducing accidents.

While the proposed SST-DHL model demonstrates the effectiveness in the traffic accident prediction task, we must acknowledge several limitations for future research and development: i). While SST-DHL

performs well within a data sparsity range of 1 % to 20 %, it encounters difficulties below 1 %. Addressing this challenge may involve exploring alternative modeling approaches or incorporating additional data sources for improved prediction accuracy. ii). To enhance real-time prediction capabilities, exploring finer-scale time divisions is crucial. Refining the temporal granularity allows the model to capture short-

term variations and provide more accurate prediction of traffic incidents as they unfold. iii). While our study focuses on identifying patterns within traffic accident data, integrating multiple sources of traffic-related datasets, such as traffic flow, vehicle trajectory, weather conditions, etc., in future research can enrich the model's understanding of complex interactions and improve prediction accuracy. Addressing these challenges in future studies can further enhance the framework's robustness and applicability in real-world scenarios, ultimately contributing to the development of more effective traffic management systems and accident prevention strategies.

CRediT authorship contribution statement

Pengfei Cui: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Xiaobao Yang:** Writing – review & editing, Validation, Supervision, Funding acquisition, Formal analysis. **Mohamed Abdel-Aty:** Writing – review & editing, Validation, Supervision, Investigation, Formal analysis. **Jinlei Zhang:** Writing – review & editing, Investigation, Formal analysis. **Xuedong Yan:** Validation, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2023YFC3009600), the Key Program of the National Natural Science Foundation of China (No. 62333016) and the National Natural Science Foundation of China and the Civil Aviation Administration of China under Grant (No. U2333206). The authors also appreciate the contribution of the UCF SST lab at which the first author was a visiting scholar.

References

- Alvarez-Rodriguez, U., Battiston, F., de Arruda, G.F., Moreno, Y., Perc, M., Latora, V., 2021. Evolutionary dynamics of higher-order interactions in social networks. *Nat. Hum. Behav.* 5 (5), 586–595.
- Avuglah, R.K., Adu-Poku, K.A., Harris, E., 2014. Application of ARIMA models to road traffic accident cases in Ghana. *Int. J. Stat. Appl.* 4 (5), 233–239.
- Bergel-Hayat, R., Debbah, M., Antoniou, C., Yannis, G., 2013. Explaining the road accident risk: Weather effects. *Accid. Anal. Prev.* 60, 456–465.
- Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. *Accid. Anal. Prev.* 39 (4), 657–670.
- Chang, L.-Y., 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. *Saf. Sci.* 43 (8), 541–557.
- Chen, H., Wang, Y., Guo, T., Xu, Chang, Deng, Y., Liu, Z., Ma, S., Xu, Chunjing, Xu, Chao, Gao, W., 2021. Pre-trained image processing transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12299–12310.
- Contisciani, M., Battiston, F., De Bacco, C., 2022. Inference of hyperedges and overlapping communities in hypergraphs. *Nat Commun* 13 (1), 7229.
- Cui, P., Yang, X., Ma, L., Mu, C., 2023. Modeling non-parametric effects of two-vehicle speed on crash risk at intersections: Leveraging two-dimensional additive logistic regression beyond univariable approach. *J. Transp. Safety Security* 1–22.
- De Ona, J., López, G., Mujalli, R., Calvo, F.J., 2013. Analysis of traffic accidents on rural highways using latent class clustering and bayesian networks. *Accid. Anal. Prev.* 51, 1–10.
- Dong, C., Chang, N., 2023. Overview of the identification of traffic accident-prone locations driven by big data. *Digital Transp. Safety* 2 (1), 67–76.
- Dong, C., Clarke, D.B., Yan, X., Khattak, A., Huang, B., 2014. Multivariate random-parameter zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. *Accid. Anal. Prev.* 70, 320–329.
- Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects. *Accid. Anal. Prev.* 82, 192–198.
- Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y., 2019. Hypergraph neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 3558–3565.
- Gao, Y., Zhang, Z., Lin, H., Zhao, X., Du, S., Zou, C., 2022. Hypergraph learning: Methods and practices. *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (5), 2548–2566.
- Guo, S., Lin, Y., Feng, N., Song, C., Wan, H., 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *AAAI* 33 (1), 922–929.
- Hassani, K., Khasahmadi, A.H., 2020. Contrastive multi-view representation learning on graphs. *Int. Conf. Mach. Learn.* PMLR 4116–4126.
- Hong, D., Dey, R., Lin, X., Cleary, B., Dobriban, E., 2022. Group testing via hypergraph factorization applied to COVID-19. *Nat. Commun.* 13 (1), 1837.
- Huang, C., Zhang, C., Zhao, J., Wu, X., Yin, D., Chawla, N., 2019. MiST: A Multiview and Multimodal Spatial-Temporal Learning Framework for Citywide Abnormal Event Forecasting, in: The World Wide Web Conference on - WWW '19. Presented at the The World Wide Web Conference, ACM Press, pp. 717–728.
- Islam, Z., Abdel-Aty, M., Cai, Q., Yuan, J., 2021. Crash data augmentation using variational autoencoder. *Accid. Anal. Prev.* 151, 105950.
- Ji, J., Yu, F., Lei, M., 2022. Self-supervised spatiotemporal graph neural networks with self-distillation for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 24, 1580–1593.
- Jin, G., Li, F., Zhang, J., Wang, M., Huang, J., 2023. Automated dilated spatio-temporal synchronous graph modeling for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 1–11 (in press).
- Kim, J.-K., Ulfarsson, G.F., Kim, S., Shankar, V.N., 2013. Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accid. Anal. Prev.* 50, 1073–1081.
- Lemonakis, P., Eliou, N., Karakasidis, T., 2021. Investigation of speed and trajectory of motorcycle riders at curved road sections of two-lane rural roads under diverse lighting conditions. *J. Saf. Res.* 78, 138–145.
- Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. *Accid. Anal. Prev.* 135, 105371.
- Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. *arXiv preprint arXiv:1707.01926*.
- Li, Z., Huang, C., Xia, L., Xu, Y., Pei, J., 2022. Spatial-Temporal Hypergraph Self-Supervised Learning for Crime Prediction, in: 2022 IEEE 38th International Conference on Data Engineering (ICDE). Presented at the 2022 IEEE 38th International Conference on Data Engineering (ICDE), IEEE, Kuala Lumpur, Malaysia, pp. 2984–2996.
- Lin, Z., Li, M., Zheng, Z., Cheng, Y., Yuan, C., 2020. Self-attention convlstm for spatiotemporal prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 11531–11538.
- Lin, L., Wang, Q., Sadek, A.W., 2015. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. *Transp. Res. Part C: Emerg. Technol.* 55, 444–459.
- Liu, Y., Ao, X., Qin, Z., Chi, J., Feng, J., Yang, H., He, Q., 2021b. Pick and choose: A GNN-based imbalanced learning approach for fraud detection. *Proc. Web Conf.* 2021, 3168–3177.
- Liu, X., Liang, Y., Huang, C., Zheng, Y., Hooi, B., Zimmermann, R., 2022. When do contrastive learning signals help spatio-temporal graph forecasting?, in: Proceedings of the 30th International Conference on Advances in Geographic Information Systems. pp. 1–12.
- Liu, X., Zhang, F., Hou, Z., Wang, Z., Mian, L., Zhang, J., Tang, J., 2021a. Self-supervised learning: Generative or contrastive. *IEEE Trans. Knowl. Data Eng.* 35 (1), 857–876.
- Lv, Y., Tang, S., Zhao, H., 2009. Real-time highway traffic accident prediction based on the k-nearest neighbor method. In: In: 2009 International Conference on Measuring Technology and Mechatronics Automation. IEEE, pp. 547–550.
- Ma, C., Dai, G., Zhou, J., 2021. Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method. *IEEE Trans. Intell. Transp. Syst.* 23 (6), 5615–5624.
- Ma, L., Xiong, H., Wang, Z., Xie, K., 2019. Impact of weather conditions on middle school students' commute mode choices: Empirical findings from Beijing, China. *Transp. Res. Part D: Transp. Environ.* 68, 39–51.
- Ma, C., Zhao, Y., Dai, G., Xu, X., Wong, S.-C., 2022. A novel STFSA-CNN-GRU hybrid model for short-term traffic speed prediction. *IEEE Trans. Intell. Transp. Syst.* 24 (4), 3728–3737.
- Martius, O., Schwierz, C., Davies, H.C., 2006. A refined Hovmöller diagram. *Tellus A: Dyn. Meteorol. Oceanogr.* 58 (2), 221–226.
- Moosavi, S., Samavatian, M.H., Parthasarathy, S., Teodorescu, R., Ramnath, R., 2019. Accident risk prediction based on heterogeneous sparse data: new dataset and insights. In: In: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 33–42.
- Ren, H., Song, Y., Wang, J., Hu, Y., Lei, J., 2018. A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction, 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 3346–3351.
- Sameen, M.I., Pradhan, B., 2017. Severity prediction of traffic accidents with recurrent neural networks. *Appl. Sci.* 7 (6), 476.
- Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., Murthy, K.R.K., 2000. Improvements to the SMO algorithm for SVM regression. *IEEE Trans. Neural Netw.* 11 (5), 1188–1193.
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 28.
- Tao, S., Corcoran, J., Hickman, M., Stinson, R., 2016. The influence of weather on local geographical patterns of bus usage. *J. Transp. Geogr.* 54, 66–80.
- Velicković, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D., 2018. Deep Graph Infomax. *arXiv preprint arXiv:1809.10341*.

- Wang, S., Cao, J., Yu, P.S., 2022b. Deep learning for spatio-temporal data mining: A survey. *IEEE Trans. Knowl. Data Eng.* 34 (8), 3681–3700.
- Wang, L., Egorova, E.K., Mokryakov, A.V., 2018b. Development of hypergraph theory. *J. Comput. Syst. Sci. Int.* 57, 109–114.
- Wang, B., Luo, X., Zhang, F., Yuan, B., Bertozzi, A.L., Brantingham, P.J., 2018. Graph-based deep modeling and real time forecasting of sparse spatio-temporal data. arXiv preprint arXiv:1804.00684.
- Wang, Q., Gan, S., Chen, W., Li, Q., Nie, B., 2021b. A data-driven, kinematic feature-based, near real-time algorithm for injury severity prediction of vehicle occupants. *Accid. Anal. Prev.* 156, 106149.
- Wang, J., Zhang, Y., Wei, Y., Hu, Y., Piao, X., Yin, B., 2021a. Metro passenger flow prediction via dynamic hypergraph convolution networks. *IEEE Trans. Intell. Transp. Syst.* 22 (12), 7891–7903.
- Wang, J., Zhang, Y., Wang, L., Hu, Y., Piao, X., Yin, B., 2022a. Multitask hypergraph convolutional networks: A heterogeneous traffic prediction framework. *IEEE Trans. Intell. Transp. Syst.* 23 (10), 18557–18567.
- Wang, Y., Zhu, D., 2022. SHGCN: a hypergraph-based deep learning model for spatiotemporal traffic flow prediction, in: In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, pp. 30–39.
- World Health Organization, 2019. World health statistics 2019: Monitoring health for the SDGs, sustainable development goals. World Health Organization, Geneva.
- Xu, M., Dai, W., Liu, C., Gao, X., Lin, W., Qi, G.-J., Xiong, H., 2021. Spatial-Temporal Transformer Networks for Traffic Flow Forecasting. arXiv preprint arXiv: 2001.02908.
- Xu, R., Luo, F., 2021. Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest. *Saf. Sci.* 135, 105125.
- Xue, J., Jiang, N., Liang, S., Pang, Q., Yabe, T., Ukkusuri, S.V., Ma, J., 2022. Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nat. Mach. Intell.* 4 (3), 246–257.
- Young, J.-G., Petri, G., Peixoto, T.P., 2021. Hypergraph reconstruction from network data. *Commun. Phys.* 4 (1), 1–11.
- Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. *Accid. Anal. Prev.* 51, 252–259.
- Yu, L., Du, B., Hu, X., Sun, L., Han, L., Lv, W., 2021. Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neurocomputing* 423, 135–147.
- Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep Learning framework for traffic forecasting. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 3634–3640.
- Yuan, Z., Zhou, X., Yang, T., 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, KDD '18, pp. 984–992.
- Zhang, C., James, J.Q., Liu, Y., 2019. Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. *IEEE Access* 7, 166246–166256.
- Zhang, X., Xu, Y., Shao, Y., 2022b. Forecasting traffic flow with spatial-temporal convolutional graph attention networks. *Neural Comput. Appl.* 34 (18), 15457–15479.
- Zhang, W., Zhu, F., Lv, Y., Tan, C., Liu, W., Zhang, X., Wang, F.-Y., 2022a. AdapGL: An adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks. *Transp. Res. Part C: Emerg. Technol.* 139, 103659.
- Zheng, C., Fan, X., Wang, C., Qi, J., 2020. GMAN: A graph multi-attention network for traffic prediction. *Proc. AAAI Conf. Artif. Intell.* 34 (1), 1234–1241.
- Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., Wang, Z., 2019. Traffic accident's severity prediction: A deep-learning approach-based CNN network. *IEEE Access* 7, 39897–39910.
- Zhou, Y., Zheng, H., Huang, X., Hao, S., Li, D., Zhao, J., 2022. Graph neural networks: Taxonomy, advances, and trends. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13 1, 1–54.
- Zhou, Y., Li, J., 2019. Research of network traffic anomaly detection model based on multilevel autoregression. In: In: 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), pp. 380–384.
- Zhou, Z., Wang, Y., Xie, X., Chen, L., Liu, H., 2020. RiskOracle: A minute-level citywide traffic accident forecasting framework. *AAAI* 34 (1), 1258–1265.
- Zhou, Z., Wang, Y., Xie, X., Chen, L., Zhu, C., 2022c. Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective. *IEEE Trans. Knowl. Data Eng.* 34 (8), 3786–3799.
- Zhou, Y., Zheng, H., Huang, X., Hao, S., Li, D., Zhao, J., 2022b. Graph neural networks: Taxonomy, advances, and trends. *ACM Trans. Intell. Syst. Technol. (TIST)* 13 (1), 1–54.