



Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction

Xiaowei Gao^{a,*}, Xinke Jiang^b, James Haworth^a, Dingyi Zhuang^c, Shenhao Wang^e, Huanfa Chen^d, Stephen Law^f

^a SpaceTimeLab, University College London (UCL), London, UK

^b School of Computer Science, Peking University (PKU), Beijing, China

^c Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

^d The Bartlett Centre for Advanced Spatial Analysis, University College London (UCL), London, UK

^e Department of Urban and Regional Planning, University of Florida, Gainesville, FL, USA

^f Department of Geography, University College London (UCL), London, UK

ARTICLE INFO

Keywords:

Spatiotemporal sparse data mining
Traffic crash prediction
Uncertainty quantification
Zero-inflated Tweedie distribution

ABSTRACT

Traffic crashes present substantial challenges to human safety and socio-economic development in urban areas. Developing a reliable and responsible traffic crash prediction model is crucial to address growing public safety concerns and improve the safety of urban mobility systems. Traditional methods face limitations at fine spatiotemporal scales due to the sporadic nature of high-risk crashes and the predominance of non-crash characteristics. Furthermore, while most current models show promising occurrence prediction, they overlook the uncertainties arising from the inherent nature of crashes, and then fail to adequately map the hierarchical ranking of crash risk values for more precise insights. To address these issues, we introduce the **Spatiotemporal Zero-Inflated Tweedie Graph Neural Networks (STZITD-GNN)**, the first uncertainty-aware probabilistic graph deep learning model in road-level daily-basis traffic crash prediction for multi-steps. Our model combines the interpretability of the statistical Tweedie family with the predictive power of graph neural networks, excelling in predicting a comprehensive range of crash risks. The decoder employs a compound Tweedie model, handling the non-Gaussian distribution inherent in crash data, with a zero-inflated component for accurately identifying non-crash cases and low-risk roads. The model accurately predicts and differentiates between high-risk, low-risk, and no-risk scenarios, providing a holistic view of road safety that accounts for the full spectrum of probability and severity of crashes. Empirical tests using real-world traffic data from London, UK, demonstrate that the STZITD-GNN surpasses other baseline models across multiple benchmarks, including a reduction in regression error of up to 34.60% in point estimation metrics and an improvement of above 47% in interval-based uncertainty metrics.

1. Introduction

Traffic crashes persist as a major barrier to sustainable urban development (Dai et al., 2018; Tu et al., 2020; Barahimi et al., 2021; Rodrigue, 2020). The recent World Health Organisation (WHO) report reveals that road traffic crashes claimed approximately 1.19 million lives worldwide in 2021, making them the leading cause of mortality and disability and imposing an economic burden of up to US\$1.8 trillion, approximately 10%–12% of the global Gross Domestic Product (GDP) (World Health Organization, 2023).

To support reductions in traffic crashes, there is a growing emphasis on the development of models aimed at predicting the location,

severity, and causes of crashes. The early approaches relied mainly on linear regression models (Mountain et al., 1996; Greibe, 2003), models that incorporate random parameters (El-Basyouny and Sayed, 2009; Dinu and Veeraragavan, 2011), and geographically weighted regression (GWR) models (Li et al., 2013; Xu and Huang, 2015) to predict regional traffic crashes. Recent research has turned to deep learning-based methods, notably graph neural networks (GNNs), which are able to handle large datasets and capture complex spatial and temporal representations of urban road networks. Examples include time-varying GNN (Zhou et al., 2020b), multiview fusion GNN (Trirat et al., 2023) and subgraph-based GNN (Zhao et al., 2023). Despite

* Corresponding author.

E-mail addresses: xiaowei.gao.20@ucl.ac.uk (X. Gao), XinkeJiang@stu.pku.edu.cn (X. Jiang), j.haworth@ucl.ac.uk (J. Haworth), dingyi@mit.edu (D. Zhuang), shenhaowang@ufl.edu (S. Wang), huanfa.chen@ucl.ac.uk (H. Chen), stephen.law@ucl.ac.uk (S. Law).

<https://doi.org/10.1016/j.aap.2024.107801>

Received 27 July 2024; Received in revised form 23 September 2024; Accepted 24 September 2024

Available online 2 October 2024

0001-4575/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

substantial advancements in model development, accurately predicting road crashes at a fine spatiotemporal granularity road presents several challenges:

- **Spatially Imbalanced Crash Occurrence:** The prediction of road crashes is challenging due to the uneven spatial distribution of crashes. Firstly, zero-inflation is inherent to crash data, where areas with no crashes are much more common than those with crashes. This spatially skewed distribution causes models to be biased towards predicting zero crashes, complicating the effective training of predictive algorithms (Wu et al., 2022). Secondly, where crashes do occur, they are predominantly minor, leading to an abundance of low-risk scores across the spatial grid. This creates a misleading representation that underestimates the potential severity of less frequent, more severe crashes (Shirazi and Lord, 2019; Saha et al., 2020). The combination of these factors — widespread zero-crash areas and the dominance of minor crashes in crash-prone zones — results in a challenging scenario for models that need to understand and predict the full spectrum of crash severities across different areas (Wang et al., 2023b).
- **Temporal Constraints with Uncertainty:** Building on the spatial challenges, the temporal dynamics of traffic crashes introduce additional complexity. Particularly, predictive models for traffic crashes face the challenge of accurately forecasting not just immediate but also future conditions, necessitating reliable multi-step predictions (Soltani and Qadikolaei, 2024). The nature of traffic crashes, characterised by their non-linearity and the variable time intervals between events, adds complexity to this task. As the prediction horizon extends, uncertainty increases, particularly because future traffic conditions and potential crash triggers can change substantially over time (Qian et al., 2022). Most current deep learning methods do not fully address these challenges. They tend to overlook the complexities of aleatoric uncertainty, which arises from the inherent randomness in the data. These models often rely on assumptions of uniform variance across all data points, producing point estimates from average outcomes (Wang et al., 2023b; Zhuang et al., 2022; Jiang et al., 2023). This approach fails to capture the temporal variability and the unpredictable nature of crashes, resulting in predictions that do not adequately reflect the true range of future possibilities.

To address the intricate challenges posed by the spatial and temporal distribution of crash data, we introduce the Spatio-temporal Zero-Inflated Tweedie Graph Neural Network (STZITD-GNN). This model combines spatio-temporal graph machine learning with statistical methods, enabling enhanced uncertainty quantification over multi-step prediction intervals. It is specifically designed to handle the probabilistic assessment of non-recurrent, road-level traffic crashes, providing a robust solution to predict spatially and temporally varying crash data.

The proposed framework integrates a cohesive encoder that combines a Gated Recurrent Unit (GRU) for capturing temporal dynamics and a Graph Attention Network (GAT) for spatial relationships, ensuring a synchronised capture of spatio-temporal dependencies across individual roads and time steps. The decoder leverages a Tweedie (TD)-based distribution, a flexible compound Poisson-gamma model that simultaneously models both exact zero occurrences and continuous positive values. The choice of the TD distribution is motivated by its established efficacy in fields challenged by nonnormal distributions, such as crash frequency estimation (Wang et al., 2019), insurance claims (Smyth and Jørgensen, 2002), actuarial studies (Shi, 2016), and meteorological precipitation (Dunn, 2004). This approach allows the model to output the TD distribution parameters, explicitly determining the upper and lower bounds for the prediction intervals of each sample, thus providing a probabilistic representation of each road's crash risk scores.

Given the inherent extreme imbalance that traditional TD models struggle to adequately capture (Khosravi et al., 2010; Abe and Yadohisa, 2017), we have enhanced our model by incorporating an additional parameter, π , to estimate the probability mass of zero values. This adjustment leads to the creation of the Zero-Inflated Tweedie (ZITD) distribution on the decoder, which offers a more comprehensive and responsible depiction of real crash scenarios. This refinement ensures that our model not only captures the complexities of traffic crash data, but also enhances the accuracy and reliability of its predictions. Specifically, our main contributions can be summarised as follows.

- To address the limitations of ST-GNN in modelling complex spatial and temporal relationships, our framework introduces an end-to-end spatio-temporal encoder. This encoder is specifically designed to simultaneously model pairwise relationships and integrate them into a cohesive embedding. This approach enhances the ability of our model to capture and synthesise the intricate dynamics between spatial and temporal data points.
- An innovative ZITD distribution is employed within our model to effectively address the uncertainties associated with zero-inflation and the long-tailed characteristics prevalent in road crash data. This approach enhances the model's capability to accurately represent the complex distribution patterns observed in traffic crash data.
- To the best of our knowledge, the STZITD-GNN is the first work to use probabilistic framework for granular, multistep road-level crash forecasting. The use of Likelihood-based Loss minimisation rather than the traditional regression ones enhances its uncertainty-aware predictive accuracy and reliability.
- Extensive testing with real-world data confirms the superior performance of our framework compared to existing state-of-the-art approaches. Notably, it achieves a reduction in regression error of up to 34.60% in point estimation metrics and shows a significant improvement, increasing at least by 55.07% in uncertainty-based metrics.

This manuscript is structured as follows. In Section 2, we review spatiotemporal modelling methodologies in predicting traffic risk. Section 3 provides a detailed description of the proposed model. Section 4 compares our empirical results with state-of-the-art benchmark models, further enhancing our assertions with illustrative visualisations. In Section 5, we conclude the article by highlighting the advantages of our model and suggesting avenues for future improvements.

2. Literature review

Predicting traffic crashes is essential for urban planning and mobility management. Initially, research in this area predominantly utilised traditional statistical methods and classical machine learning techniques, such as regression models (Chang, 2005; Lord et al., 2005; Caliendo et al., 2007; Bergel-Hayat et al., 2013), Bayesian networks (Martin et al., 2009; Mujalli and De Oña, 2011; Hossain and Muro-machi, 2012), tree-based models (Chong et al., 2004; Chang and Chen, 2005; Wang et al., 2010; Lin et al., 2015), and the k-nearest-neighbour method (Sayed and Abdelwahab, 1998; Tang and Gao, 2005; Lv et al., 2009). These early studies were generally confined to small geographical areas and demonstrated limited effectiveness in capturing the nonlinear and dynamic spatio-temporal patterns of the contributing factors (Chang and Chen, 2005; Anderson, 2009; Zhang et al., 2014). In addition, these methods often analyse crash data in isolation, neglecting critical interdependencies between different locations (Wang et al., 2013a, 2021). As a result, their broader applicability to city-wide analysis with large data size was constrained, highlighting the need for more advanced models capable of addressing these limitations.

Recent advances in deep learning techniques have shown promise in urban traffic crash prediction by incorporating spatio-temporal characteristics of the data. Convolutional neural networks (CNNs), recurrent

neural networks (RNNs), and graph-based networks, such as graph-convolution networks (GCNs), have demonstrated their effectiveness in capturing intricate spatial patterns, capturing temporal dependencies, and modelling complex interactions (Wang and Cao, 2021; Liu et al., 2020). As summarised in Table 1, Chen et al. (2016) were among the first to explore city level crash prediction employing a stack denoise autoencoder. They integrated human mobility GPS data and historical incident point data at the grid level to map the real-time crash situation in Tokyo. However, while their approach attempted to capture a larger spatial area, it did not consider urban geo-semantic information for precise and long-term crash prediction. Later, Chen et al. (2018) introduced a Stack Denoise Convolutional Autoencoder that incorporated spatial dependencies using stacked CNNs. Nonetheless, both of these studies overlooked the influence of temporal factors. Although RNNs are capable of handling temporal information, they are more suitable for short-term temporal learning (Ren et al., 2018). To better address the temporal aspect, Ren et al. (2018) used Long-Short-Term Memory (LSTM) to consider influential temporal factors between several locations. Their work represented the first attempt to incorporate temporal factors into deep learning models for the prediction of city-wide traffic crashes. Similarly, Moosavi et al. (2019) employed LSTM networks for country-wide crash prediction, demonstrating their efficacy in predicting a binary outcome. Building on the LSTM framework, Yuan et al. (2018) proposed a more advanced model known as the Convolutional Long Short-Term Memory (Hetero-ConvLSTM) neural network based on spatial heterogeneous data. This model aimed to jointly address spatial heterogeneity and capture temporal features in state-wide crash prediction. However, their approach relied on pre-defined moving windows based on the spatial features of Iowa, United States, limiting its generalisability to other cities. Bao et al. (2019) integrated CNNs, LSTM as well as ConvLSTM into a spatiotemporal convolutional long short-term memory network (STCL-Net). Their work presented a more practical and feasible approach for the prediction of city-wide traffic crashes. By combining the strengths of these models, STCL-Net effectively captured spatiotemporal dependencies in crash data and influential factors, leading to improved prediction performance.

To capture local and global dynamics with hierarchical spatial information, Zhu et al. (2019) proposed the Deep Spatio-temporal Attention Learning Framework, which incorporates a spatio-temporal attention mechanism to effectively capture the dynamic impact of traffic crashes at different spatial levels. One notable aspect of their work was the emphasis on using real traffic administrative areas instead of manually divided grids. However, this approach does not incorporate spatial relationships, which limits its capacity to model the spatial heterogeneities that characterise traffic crashes. In a similar vein, Huang et al. (2019b) introduced the Deep Dynamic Fusion Network (DFN) framework, which aggregates heterogeneous external factors across both spatial and temporal dimensions. Using a temporal aggregation layer, the DFN framework is designed to automatically capture external influences from latent temporal dimensions.

Although the mentioned studies effectively captured temporal and spatial patterns in traffic crash data, yielding promising prediction outcomes, their lack of consideration of latent spatial correlations within urban structures can disrupt regional patterns in practical urban science and lead to prediction inaccuracies. Therefore, Graph-based networks have become a key method for city-wide traffic crash prediction by leveraging spatial graph features from actual urban networks (Xue et al., 2022; Cheng et al., 2022; Li et al., 2022).

Zhou et al. (2020b) developed the RiskOracle framework by constructing an urban graph of regions and integrating the Differential Time-varying Graph Neural Network (DTGN) to identify dynamic correlations between historical traffic conditions and risks, enabling simultaneous predictions of traffic flow and crashes. To tackle zero-inflation in crash data, they introduced a data enhancement strategy based on a priori knowledge (PKDE), defining subregional crash labels for training. This approach adjusts zero crash values to a range between 0

and 1, maintaining the hierarchy of actual crash levels. Expanding on this foundation, Zhou et al. (2022b) refined the DTGN framework by embedding an LSTM-based hierarchical sequence learning architecture, accommodating multiscale historical spatio-temporal data for nuanced long-term and short-term predictions. Wang et al. (2021,b) both focus on the integration of geographical and semantic relationships through different graph structures. While Wang et al. (2021) used crash, road, and POI graphs to capture regional semantics and introduced modules for geographical and semantic spatio-temporal correlations, Wang et al. (2021b) combined attention-based LSTM, CNN, and GCN for multi-view predictions of traffic crashes at various scales. Although both approaches tackled the zero-inflation issue with promising results, they share limitations in addressing dynamic geographical semantics and alignment issues in graph structures. In contrast, Trirat et al. (2023) proposed a multi-view graph neural network incorporating dynamic and static similarity information, offering a more dynamic approach to understanding traffic crashes. Using the Huber loss function, a parameter-based regression loss, their model aimed at robust regression that effectively accommodates zero inflation, showcasing an advanced strategy for precise prediction of urban traffic crashes.

Recent attempts to apply graph-based deep learning for road-level crash prediction have shown promise, but also highlight areas for improvement. Yu et al. (2021) developed a Deep Spatiotemporal Graph Convolutional Network with a unique three-layer network, treating the road graph, spatio-temporal data, and embeddings separately. They tackled zero inflation through undersampling, balancing the dataset between risky and non-risky roads, and framing crash prediction as a binary classification. While this method effectively predicts crashes at the road level, the binary outcome overlooks the varying degrees of crash risk across roads. Similarly, Wu et al. (2022) designed a fully connected dynamic network to predict road-level crashes by focussing on dynamic spatio-temporal correlations. To mitigate zero inflation, they introduced a cost-sensitive learning module that enhances positive sample classification accuracy by considering the imbalance between zero and non-zero crash values. However, this work only considered limited road segments in New York City with binary crash occurrence outputs, highlighting the need for further development to capture the full spectrum of crash severity and provide a more detailed understanding of road-level crash scenarios.

In summary, despite significant progress in urban traffic crash prediction yielding promising results, existing studies have yet to fully address uncertainty quantification in their predictions and provide multi-step crash assessments at the detailed road level.

3. Methodology

In this section, we first introduce the concepts and define the framework for predicting road-level traffic crashes in Section 3.1. We then illustrate the statistical innovations from the TD distribution to ZITD in the context of uncertainty quantification in Section 3.2. Subsequently, we introduce our new STZITD-GNN model, a spatio-temporal probabilistic modelling framework that combines the strengths of GRU and GAT in Section 3.3. The section ends with an in-depth exploration of the loss function employed within our model in Section 3.4. The major notations used in this section are listed in Table 2.

3.1. Preliminaries and problem definition

This section presents the basic concepts and terminology relevant to the study, followed by a description of the problem.

Definition 1 (Road Traffic Crash Risk Values). A traffic crash refers to an crash that involves one or more road users that leads to physical injury, loss of life, or property damage. We denote the set of roads as V of size $N = |V|$, where v_i signifies the i th road. The traffic crash risk on the roads is represented as $Y \in \mathbb{R}^{N \times T}$ for a duration of total T

Table 1
Summary of deep learning methods in traffic crash prediction.

Reference	Methodology	Spatial Unit	Zero-Inflation Issue	Prediction Steps	Prediction Outcome
Chen et al. (2016)	Stack Denoise Autoencoder	Regions	N/A	Single	Crash Risk Score
Ren et al. (2018)	Long Short-term Memory network	Regions	N/A	Multiple	Crash Risk Score
Chen et al. (2018)	Stack Denoise Convolutional Autoencoder	Regions	N/A	Single	Crash Risk Score
Yuan et al. (2018)	Convolutional Long Short-Term Memory	Regions	N/A	Multiple	Crash Risk Score
Moosavi et al. (2019)	Long Short-term Memory Network	Regions	N/A	Single	Crash Occurrence
Zhu et al. (2019)	Deep Spatial-Temporal Attention Learning Framework	Regions	N/A	Multiple	Crash Risk Score
Huang et al. (2019b)	Dynamic Fusion Network	Regions	N/A	Single	Crash Occurrence
Bao et al. (2019)	Spatiotemporal Convolutional LSTM	Regions	N/A	Single	Crash Risk Score
Zhou et al. (2020b)	DTGN with PKDE	Regions	Yes	Single	Crash Risk Score
Wang et al. (2021)	GSNet with Weighted Loss	Regions	Yes	Single	Crash Risk Score
Wang et al. (2021b)	MVMT-ST Networks	Regions	Yes	Single	Crash Risk Score
Yu et al. (2021)	Deep ST-GCN with Undersampling	Roads	Yes	Single	Crash Occurrence
Wu et al. (2022)	MADGCN with Cost-sensitive Loss	Roads	Yes	Multiple	Crash Occurrence
Zhou et al. (2022b)	Enhanced DTGN	Regions	Yes	Multiple	Crash Risk Score
Trirat et al. (2023)	MV-GNNs with Huber Loss	Regions	Yes	Multiple	Crash Risk Score

Table 2
Main notations in this paper.

Notation	Description
Y / X	Road-level traffic crash risk score / Road-level spatiotemporal traffic feature
$V / E / A$	Road set / Edge set / Road adjacency matrix
$v_i / (v_i, v_j)$	The i th road / Edge between the i th road and j th.
N / T	Numbers of roads / Time length
C_{it}	Number of crashes on road v_i at time slot t
l_{it}^j	The j th cardinality values of crash severities in the time slot t for road v_i
$d / F / F'$	Feature dimension/Temporal feature dimension/Spatio-Temporal feature dimension
ST	Spatiotemporal deep learning Model
$f_{TD}(\cdot)$	Probabilistic density function of Tweedie distribution
$\rho / \phi / \mu / \pi$	Index parameter / Dispersion parameter/ Mean parameter/ Zero-inflated parameter
$\theta / \lambda / \alpha, \gamma$	Natural parameter / Poisson mean parameter / Gamma parameters
Z_T / Z	Temporal embedding / Spatiotemporal embedding
$\text{ReLU}(\cdot) / \sigma(\cdot) / \text{LeakyReLU}(\cdot) / [- \cdot]$	ReLU activation/ Sigmoid activation/LeakyReLU activation/ Concatenation
$W_r, W_a, W_c, W_g, b_r, b_a, b_c, \bar{a}$	Learnable weight matrices of GRU and GAT
$W_\pi, W_\mu, W_\phi, W_\rho, b_\pi, b_\mu, b_\phi, b_\rho$	Learnable weight matrices of the four Parameter Encoders
$h_i / r_i / u_i$	Hidden feature / Reset gate / Update gate
$\alpha_{i,j}$	Attention value between road v_i and v_j
M	Number of attention heads
ϵ	The minimum value
η	L2 normalisation weight-parameter

periods, and y_{it} denotes the crash risk in the t th time slot for the road v_i . Thus, $Y_t \in \mathbb{N}^N$ signifies the crash risk for all roads in the t th time interval, where y_{it} is its constituent element. The calculation of traffic crash risk score y_{it} for the road v_i at the time slot t is given by:

$$y_{it} = \sum_{j=1}^3 C_{i,k}^t \times l_j, \tag{1}$$

where $C_{i,k}^t$ denotes the total number of k crashes with the severity level l_j at time slot t for road v_i . We allocate the crash point to its closest road. As per prior research (Wang et al., 2021; Trirat et al., 2023), l is assigned the values 1, 2, and 3, representing minor injury, serious injury, and fatal crash severities, respectively. y_{it} equals 0 if there are no traffic crashes.

Definition 2 (Spatio-temporal Features). The embedding features are represented as $X \in \mathbb{R}^{N \times T \times d}$, where d is the dimension of the features. We denote x_{it} as the spatio-temporal feature of the road v_i at the time slot t th. Consequently, $X_t \in \mathbb{R}^{N \times d}$ signifies the features of all roads at time slot t . Our study includes spatial features such as road types, road lengths and widths, road conditions, POIs, and census characteristics

for each lower layer super output area (LSOA) of roads. Temporal features, which vary daily, encompass weather information such as sunrise and sunset times, humidity, visibility, rainfall, etc., along with the identification as a holiday or a working day.

Definition 3 (Road Connection Graph). The road network is represented as a graph $\mathcal{G} = (V, E, A)$, where each road segment is defined as a node in the set V . The set of edges E represents the connections between roads, while $A \in \mathbb{R}^{N \times N}$ denotes the adjacency matrix that formally describes these connections. In the adjacency matrix, an entry $A_{i,j} = 1$ indicates the existence of an edge between roads v_i and v_j , denoted as $(v_i, v_j) \in E$, while $A_{i,j} = 0$ implies no direct connection, represented as $(v_i, v_j) \notin E$.

This approach differs from the common method of using intersections as nodes and road segments as edges, instead treating each road segment as a node. This choice enables a more granular analysis of crash risks and allows the incorporation of road-specific characteristics directly into the node representations, thus enabling node prediction outcomes to inform road-level crash risk (Tang et al., 2022). Fig. 1 provides a visual representation of this road-based graph structure.

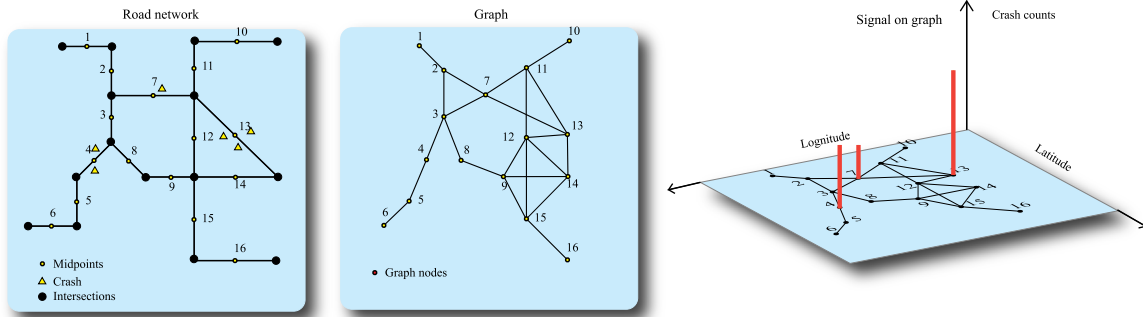


Fig. 1. Road-based Graph Structure: Visualisation of how the physical road network is transformed into a mathematical graph representation, with road segments as nodes and their connections as edges (Zhang and Cheng, 2020).

The illustration demonstrates how individual road segments are represented as nodes and how their interconnections form the edges of the graph. This visualisation helps to conceptualise the transformation of a physical road network into a mathematical graph structure, which is fundamental to our spatial-temporal analysis approach.

Tobler’s First Law of Geography posits that spatial correlation is strongly linked to geographic proximity (Tobler, 1970; Yu et al., 2020; Zhou et al., 2021; Wang et al., 2023a). In the context of traffic risk, this principle manifests itself as spatial spillover effects, where risk at one location can influence nearby areas. To capture this spatial attribute and also to deal with the signal sparsity, we employ a diffusion process based on spatiotemporal smoothing techniques (Perozzi et al., 2014; Li et al., 2017; Hamilton, 2020). Specifically, we adopt the approach proposed by Zhang and Cheng (2020), using the adjacency matrix A to model the spread of traffic risks to neighbouring roads and also incorporate temporal smoothing.

Problem Definition: Our model aims to predict the future crash risk score in the next p time windows (as daily) and the confidence interval of the predicted results per road. The task is to use historical records $X_{1:t}, Y_{1:t}$ and the graph structure A as input data for training, to predict the probabilistic density function $f(Y_{t+1:t+p})$ of the distribution of $Y_{t+1:t+p}$.

$$ST_{\theta}([X_{1:t}, Y_{1:t}], A) \rightarrow f(Y_{t+1:t+p}), \quad (2)$$

where θ denotes the learned parameters for the ST-GNN model (also referred to as ST). The multistep predictions are made for a 14-day output, as $p = 14$.

3.2. From TD distribution to ZITD distribution

This subsection provides an in-depth exploration of the progression from the TD to the ZITD distribution in the specific context of road crash prediction. Building on the work of Lord and Mannering (2010) and Smyth and Jørgensen (2002), we recognise the effectiveness of TD distribution to address overdispersed data, indicating its potential applicability in multi-step road-level traffic crash prediction.

As illustrated in Section 3.1, when we aggregate crash data (incorporating various severities) from different roads or identical roads over different time intervals, the resulting distribution of crash counts tends to exhibit overdispersion. In detail, the crash value applied to both crash counts and the associated severity — is the prevalence of ‘excess’ zeros, implying more zero entries than would be predicted under a corresponding Poisson process. Meanwhile, in nonzero crash values, a clearly left-skewed but long-tailed distribution presents a smaller proportion of higher severity crashes. In these scenarios, the variance of the crash counts outweighs the mean crash count, creating a deviation from the Gaussian distribution, as underlined by the work of Halder et al. (2019) and Wang et al. (2023b).

3.2.1. TD distribution

In accordance with Eq. (1), we designate y_k as the crash risk value of the road i on time slot t . As a special case of Exponential Dispersion Models (EDMs), the TD model encompasses a broad family of statistical distributions (Tweedie et al., 1984; Jørgensen, 1987; Jiang et al., 2023). A random variable that follows the TD distribution has a probability density function f_{TD} (Jørgensen, 1987), as follows :

$$f_{TD}(y_k | \theta, \phi) \equiv a(y_k, \phi) \exp \left[\frac{y_k(\theta) - \kappa(\theta)}{\phi} \right], \quad (3)$$

here, $\theta \in \mathbb{R}$ represents the natural parameter, while $\phi \in \mathbb{R}^+$ stands for the dispersion parameter. The normalising functions $a(\cdot)$ and $\kappa(\cdot)$ are parameter functions for ϕ and θ , respectively, detailed further in the subsequent discussion. For a TD distribution of EDMs, the mean and variance of variable y_k are given by:

$$E(y_k) = \mu = \kappa(\theta)', \quad Var(y_k) = \phi \kappa(\theta)'', \quad (4)$$

In these equations, $\kappa(\theta)'$ and $\kappa(\theta)''$ represent the first and second derivatives of $\kappa(\theta)$, respectively. Meanwhile, the function $\mu \geq 0$ works as the mean parameter. Since the mean $E(y_k)$ and variance $Var(y_k)$ are dependent, $\kappa(\theta)''$ can be directly expressed as $V(\mu)$, thereby providing a direct mean-variance relationship (Saha et al., 2020; Mallick et al., 2022). Additionally, the TD family encompasses a number of vital distributions for distinct index parameters ρ as demonstrated by $V(\mu) = \mu^\rho$. These include the Normal ($\rho = 0$), Poisson ($\rho = 1$), Gamma ($\rho = 2$), Inverse Gaussian ($\rho = 3$), and Compound Poisson-Gamma distribution ($1 < \rho < 2$) and could be referred to as Appendix A. The Compound Poisson-Gamma distribution is particularly applicable, given its focus on parameterising zero-inflated and long-tailed data, which aligns with the observed characteristics of real traffic crash data. The μ , ϕ and ρ are the three parameters that govern the probability and expected value of the risk of traffic crashes, which can be found from Appendix B.

3.2.2. ZITD distribution

The TD model, introduced in Section 3.2.1, is designed to account for the occurrence of zeros by allowing a positive probability of zero outcomes, as demonstrated by formula $\mathbb{P}(y_k = 0) = \mathbb{P}(C_k = 0) = \exp(-\lambda)$. However, this model struggles to accurately represent crash data on roads where zero values overwhelmingly dominate, leading to an extremely unbalanced distribution that can misrepresent the actual crash rate (Mallick et al., 2022; Zhou et al., 2022a).

To address this issue, we introduce a refinement, known as the Zero-Inflated Tweedie (ZITD) model with Four-parametric compound distributions. The ZITD model includes a distinctive feature, a sparsity parameter (π) that specifically addresses the skewness of the data characterised by an excess of zeros. The parameter π signifies the likelihood of zero inflation, effectively distinguishing between zero and nonzero occurrences. In contrast, the complement probability $(1 - \pi)$

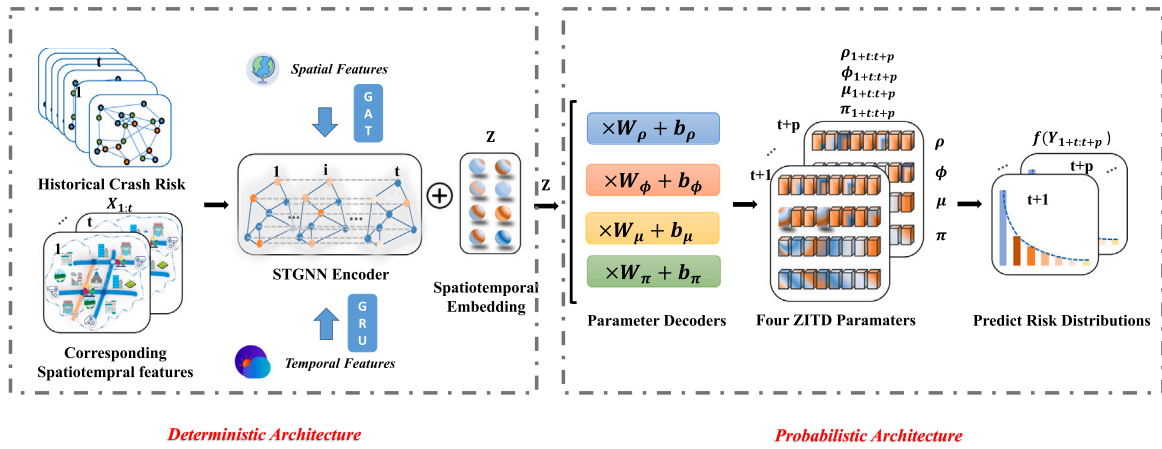


Fig. 2. The overall framework of STZITD-GNNs. STZITD-GNNs utilise the ST-GNN Encoder ST (composed of GRU and GAT encoders) to encode the history time window $1 : t$ crash risk $Y_{1:t}$, and road features $X_{1:t}$ with the use of the road connection graph in the spatial-temporal embedding of the road Z . After encoding, the four parameter decoders map Z into the ZITD parameter space and obtain $\pi_{t+1:t+p}, \mu_{t+1:t+p}, \phi_{t+1:t+p}, \rho_{t+1:t+p}$ for the predicted time window $t+1 : t+p$, which determine the predicted road traffic crash risk distribution $f(Y_{t+1:t+p})$. Note: This figure is a simplified representation. For a complete list of features, including weather and road characteristics, please refer to [Table 3](#).

aligns with the traditional TD, facilitating a balanced representation of the crash risk values across the spectrum.

$$y_k = \begin{cases} 0 & \text{with probability } \pi \\ Y & \text{with probability } 1 - \pi, Y \sim f_{TD}(Y|\mu, \phi, \rho) \end{cases} \quad (5)$$

Finally, denoting the ZITD model as $y_k \sim f_{ZITD}(\pi, \mu, \phi, \rho)$, we formulate the probability density function of this model as follows:

$$f_{ZITD}(y_k|\pi, \mu, \phi, \rho) = \begin{cases} \pi + (1 - \pi)f_{TD}(y_k = 0|\mu, \phi, \rho) & \text{if } y_k = 0 \\ (1 - \pi)f_{TD}(y_k|\mu, \phi, \rho) & \text{if } y_k > 0 \end{cases} \quad (6)$$

In accordance with this specification, we tackle the problem of extreme zero inflation by applying a linear weighting to the zero value, such that $\mathbb{P}(y_k = 0) = \pi + (1 - \pi)\exp(-\frac{1}{\phi} \frac{\mu^2 - \rho}{2 - \rho})$. This probability is significantly higher than the probability of observing a zero count from the TD distribution, intuitively indicating a substantial zero inflation. This modification ensures that the mean value of the ZITD distribution is $\mathbb{E}(y_k) = (1 - \pi)\mu$.

In the ZITD model, each parameter plays a distinct role in representing road-level crash risks. The sparsity parameter (π) is adjusted for instances without crashes, significantly influencing the dispersion parameter (ϕ) to provide a weighted interpretation suitable for left-skewed crash data. Higher ϕ values reflect the improved capability of the model to represent scenarios with lower risk accurately. Additionally, the index parameter (ρ) is essential to capture the long-tail distribution of crash risk values. Values near 2 for ρ indicate the robust performance of the model in these scenarios.

3.3. The framework of spatiotemporal zero-inflated tweedie graph neural networks (STZITD-GNNs)

In relation to the four parameters π, μ, ϕ, ρ of the ZITD probability distributions detailed above, we have designed a probabilistic spatio-temporal graph learning module, named STZITD-GNNs. This module extracts the spatiotemporal correlations $Z = ST_{\theta}(X_{1:t}, Y_{1:t}, A)$ under the assumption of ZITD for each road crash and predicts future results for the subsequent p time windows through $f_{ZITD}(X_{t+1:t+p}|Z)$, as described in Eq. (7).

$$f_{ZITD}(Y_{t+1:t+p}|\pi_{t+1:t+p}, \mu_{t+1:t+p}, \phi_{t+1:t+p}, \rho_{t+1:t+p}) = f_{ZITD}(Y_{t+1:t+p}|ST_{\theta}(X_{1:t}, Y_{1:t}, A)) = f_{ZITD}(X_{t+1:t+p}|Z). \quad (7)$$

As illustrated in [Fig. 2](#), instead of performing two independent estimations, our methodology starts with the deployment of a GRU

acting as a temporal encoder. Subsequently, a GAT is implemented for the purpose of spatial encoding. The combined output from the GRU and GAT is utilised to parameterise the ZITD distribution. The proposed approach is shown in Eq. (8). The architecture's encoding component applies a spatio-temporal embedding scheme complemented with an extra sparsity parameter and distribution parameters. The decoding component is responsible for probabilistic estimations of future multi-step crash risk values, which can be construed as the set of parameters of the outcome distribution. In contrast to the principles of variational autoencoders, which restrict latent embedding with a specific Gaussian distribution for the representation of latent variables ([Kipf and Welling, 2016b](#); [Hamilton, 2020](#)), we echo the statistical domain by allowing integration of deep learning methodologies into parameter learning.

$$\pi_{t+1:t+p}, \mu_{t+1:t+p}, \phi_{t+1:t+p}, \rho_{t+1:t+p} = ST_{\theta}(X_{1:t}, Y_{1:t}, A). \quad (8)$$

3.3.1. Parameter decoders

The detailed functions of the temporal encoder, GRU and the spatial encoder, GAT are listed in the [Appendices C and D](#). In this section, we highlight the parameter decoders in detail. To transmute the acquired spatio-temporal embeddings into ZITD parameter values, we have devised four parameter decoders to operate on Z . Consequently, four parameters, namely $\pi_{t+1:t+p}, \mu_{t+1:t+p}, \phi_{t+1:t+p}$, and $\rho_{t+1:t+p}$ can be computed from Z , as detailed below:

$$\begin{aligned} \pi_{t+1:t+p} &= \sigma(W_{\pi} \cdot Z + b_{\pi}) \\ \mu_{t+1:t+p} &= \text{ReLU}(W_{\mu} \cdot Z + b_{\mu}) \\ \phi_{t+1:t+p} &= \text{ReLU}(W_{\phi} \cdot Z + b_{\phi}) + \epsilon \\ \rho_{t+1:t+p} &= \sigma(W_{\rho} \cdot Z + b_{\rho}) + 1 + \epsilon \end{aligned} \quad (9)$$

Here, π lies within the range of $[0, 1]$, μ falls within the interval $[0, +\infty)$, ϕ exists within $(0, +\infty)$, and ρ spans the range of $(1, 2)$. The learnable weight matrices are represented as $W_{\pi}, W_{\mu}, W_{\phi}, W_{\rho}, b_{\pi}, b_{\mu}, b_{\phi}, b_{\rho} \in \mathbb{R}^{F' \times p}$. $\lim \epsilon \rightarrow 0$ represents the smallest value, and $\text{ReLU}(\cdot)$ signifies the ReLU activation function.

3.4. Learning framework of STZITD-GNNs

3.4.1. ZITD loss function

As previously mentioned, our encoding mechanism employs a spatio-temporal embedding architecture complemented by a sparsity parameter. The decoding component involves a probabilistic estimation of future road traffic crashes. To accurately predict these crashes, the overall learning objective can be seen as maximising the log-likelihood function: $\max \log f_{ZITD}(y_k|\pi, \mu, \phi, \rho)$. We use the negative likelihood as

our loss function to better fit the distribution into the data. Here, y denotes the ground-truth values corresponding to one of the predicted crashes with the parameters π, μ, ϕ, ρ . The log-likelihood of ZITD is made up of the parts $y = 0$ and $y > 0$:

For $y > 0$, $NLL_{y>0} = -\log f_{ZITD}(y > 0|\pi, \mu, \phi, \rho)$:

$$\begin{aligned} & \log f_{ZITD}(y > 0|\pi, \mu, \phi, \rho) \\ &= \log(1 - \pi) + \log f_{TD}(y > 0|\mu, \phi, \rho) \\ &= \log(1 - \pi) + \frac{1}{\phi} \left(y \frac{\mu^{1-\rho}}{1-\rho} - \frac{\mu^{2-\rho}}{2-\rho} \right) + \log a(y > 0, \phi, \rho) \\ &= \log(1 - \pi) + \frac{1}{\phi} \left(y \frac{\mu^{1-\rho}}{1-\rho} - \frac{\mu^{2-\rho}}{2-\rho} \right) - \log y \\ & \quad + \log \sum_{j=1}^{\infty} \frac{y^{-j\alpha} (\rho - 1)^{\alpha j}}{\phi^{j(1-\alpha)} (2-\rho)^j j! \Gamma(-j\alpha)} \\ & \geq \log(1 - \pi) + \frac{1}{\phi} \left(y \frac{\mu^{1-\rho}}{1-\rho} - \frac{\mu^{2-\rho}}{2-\rho} \right) - \log(j_{max} \sqrt{-\alpha} y) + j_{max}(\alpha - 1). \end{aligned} \quad (10)$$

In this equation, $j_{max} = \frac{y^{2-\rho}}{(2-\rho)\phi}$ and $\alpha = \frac{2-\rho}{1-\rho} < 0$. We optimise the lower bound; hence, $NLL_{y>0}$ can be optimised in this manner.

For $y = 0$, $NLL_{y=0} = -\log f_{ZITD}(y = 0|\pi, \mu, \phi, \rho)$:

$$\begin{aligned} \log f_{ZITD}(y = 0|\pi, \mu, \phi, \rho) &= \log(\pi) + \log(1 - \pi) f_{TD}(y = 0|\mu, \phi, \rho) \\ &= \log(\pi) + \log(1 - \pi) + \frac{1}{\phi} \left(-\frac{\mu^{2-\rho}}{2-\rho} \right) \end{aligned} \quad (11)$$

Here, π, μ, ϕ, ρ are chosen and calculated in accordance with the index of $y = 0$ or $y > 0$. The ultimate negative log-likelihood loss function is given by:

$$NLL_{STZITD} = NLL_{y=0} + NLL_{y>0} + \eta \Theta^2. \quad (12)$$

Model parameter Θ can be optimised by minimising the negative log-likelihood loss: $\hat{\Theta} = \text{argmin} NLL_{STZITD}$. η stands for the weight parameter for L_2 Normalisation.

It is worth noting that our model can be extended to other distributions by modifying the probability layer. The main reason is that we learn the parameters that determine the distributions. For example, if we choose the Gaussian distribution, we can parameterise the probability layer using the spatio-temporal embedding of the mean and variance, thereby quantifying the data uncertainty that follows the Gaussian distribution. Using the flexibility of the probability layer, we design a range of benchmark models for comparison to the STZITD-GNNs. Our code is available on Github.¹

3.4.2. Time analysis & training algorithms

We summarise the STZITD-GNN training procedure in Algorithm 1. Based on that, the time complexity can be analysed as follows (for the sake of simplicity, we denote the hidden union of embedding $F' \approx F$ and F as F):

- For temporal encoders, the time complexity can be formulated as $O(tN(dF + F^2 + F))$.
- For spatial encoders, we assume that the number of road edges is $|E|$. Hence, when the graph attention encoder is implemented, the complexity of computing attention becomes $O(M|E|F)$. Furthermore, when performing W_a multiplication, the cost is $O(MNF^2)$. This leads to a total time complexity of $O(M|E|F + MNF^2)$.
- When generating the four parameters, the four one-layer MLPs require a time complexity of $O(F^2p)$.

Algorithm 1 The training of STZITD-GNNs.

Require: Spatiotemporal road feature $X \in \mathbb{R}^{N \times T \times d}$, road network graph $G = (V, E, A)$, roads traffic crash $Y \in \mathbb{R}^{N \times T}$, model hyperparameters M, F', F , and total time slot T , time window t, p .

- 1: Calculate road traffic crash Y via Eq. (1).
- 2: Initialise parameters Θ via Xavier Initialiser.
- 3: **while** STZITD-GNNs is not converged **do** ▷ Train
- 4: Calculate temporal embedding Z_T via Eq. (C.2) by road feature $X_{1:t}$; ▷ Temporal Encode
- 5: Calculate spatiotemporal embedding Z via Eq. (D.3) by graph G and Z_T ; ▷ Spatial Encode
- 6: Obtain four parameters $\pi_{t+1:t+p}, \mu_{t+1:t+p}, \phi_{t+1:t+p}, \rho_{t+1:t+p}$ via output layer refer to Eq. (9);
- 7: Calculate distribution $f_{ZITD}(Y_{t+1:t+p}|\pi_{t+1:t+p}, \mu_{t+1:t+p}, \phi_{t+1:t+p}, \rho_{t+1:t+p})$ from four parameters via Eq. (6);
- 8: Minimising negative log-likelihood loss NLL_{STZITD} against $Y_{t+1:t+p}$ via Eq. (12) using Adam Optimiser;
- 9: **end while**
- 10: End optimising model parameters Θ ;

4. Experiments

In this section, we describe the detailed elements of the STZITD-GNNs model by analysing its performance in crash prediction in three boroughs of London, UK, namely Lambeth, Tower Hamlets and Westminster, each characterised by distinct sociodemographic traits, road networks, and traffic crash profiles. The crash data is accessed from the UK STATS19 dataset, maintained by the Department for Transport. It is made up of three interlinked tables that provide information about crashes, their severity, and related vehicle data.² We have configured the model to forecast 14 steps ahead, achieving stable temporal horizon results. Our evaluation focusses on two key aspects: first, demonstrating the model's ability to effectively quantify the uncertainty of road-level traffic crash data, and second, showing its precision in predicting.

Section 4.1 introduces the datasets used in this study and justifies the choice of data-driven strategies. This is followed by an outline of the experimental setup in Section 4.2. Subsequently, we describe the evaluation metrics in Section 4.3 and the baseline models used for comparison in Section 4.4. The subsequent analysis includes a comparison of the performance of our model against other methods in Section 4.5. It concludes with a sensitivity analysis in Section 4.6

4.1. Traffic crash data description

The study focuses on the three most vulnerable boroughs of Greater London, UK, known for their high crash frequency: Westminster, Lambeth, and Tower Hamlets, as illustrated in Fig. 3(a). The analysis covers the period from January 1, 2019, to December 31, 2019. During this time, these boroughs experienced an average of 3 to 5 daily crashes, as detailed in Table 3. The data reveal a complex and challenging pattern for road level analysis. A significant proportion of roads within these boroughs report no crashes during the study period, resulting in a pronounced zero-inflation problem. Specifically, after aggregating and processing the risk data, the road-level daily basis spatio-temporal graph risk matrix shows zero inflation rates of 95.72% in Westminster, 96.71% in Lambeth, and 96.28% in Tower Hamlets.

Furthermore, the distribution of crash risk values, as depicted in Fig. 3(b), exhibits a non-Gaussian pattern as a long-tailed effect is evident, where some instances of high-risk crash occurrences contrast with most low-risk ones, creating an extended right tail in the distribution. This combination of zero-inflation and long-tailed distribution presents a significant challenge for traditional modelling approaches. The prevalence of zero-risk roads alongside the sporadic occurrence

¹ <https://github.com/STTDAnonymous/STTD>

² <https://www.gov.uk/guidance/road-crash-and-safety-statistics-guidance>

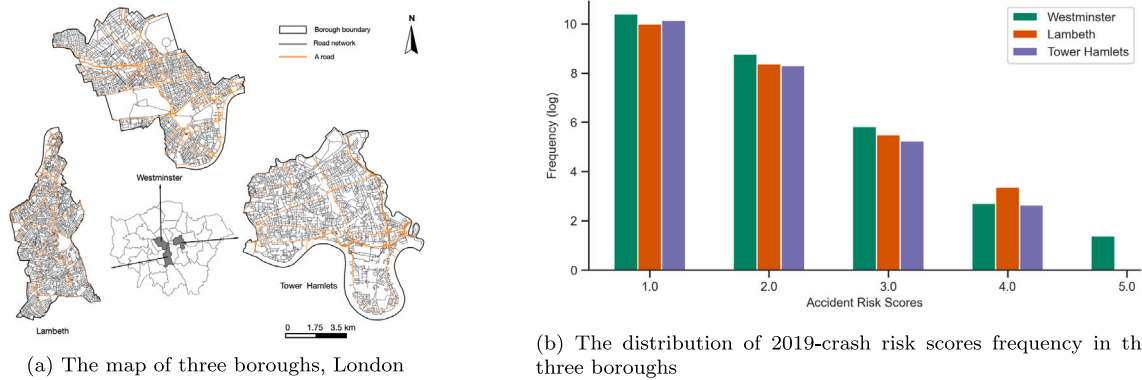


Fig. 3. Traffic crashes are depicted for the case studies.

Table 3
Data characteristics by region.

Regions	Data	Source	Classes	Observation Count
Westminster	Crash Counts	Department for Transport, UK	1	1,745
	Point of Interest	Ordnance Survey	20	25,237
	Roads	Ordnance Survey	8	4,822
	Date	/	4	365
	Meteorological Characteristics	The Met Office	8	365
	Socio-demographic Characteristics	Census 2011	8	960
Lambeth	Crash Counts	Department for Transport, UK	1	1,335
	Point of Interest	Ordnance Survey	20	10,102
	Roads	Ordnance Survey	8	5,659
	Date	/	4	365
	Meteorological Characteristics	The Met Office	8	365
	Socio-demographic Characteristics	Census 2011	8	1,456
Tower Hamlets	Crash Counts	Department for Transport, UK	1	1,236
	Point of Interest	Ordnance Survey	20	10,989
	Roads	Ordnance Survey	8	4,688
	Date	/	4	365
	Meteorological Characteristics	The Met Office	8	365
	Socio-demographic Characteristics	Census 2011	8	48,220

of high-risk events creates a highly skewed dataset that defies normal distribution assumptions, underscoring the suitability of our proposed STZITD-GNN model for this dataset.

4.2. Experiment setup

The train, validation and test data are all from 2019, as the ratio is 8:2:2. The parameters of STZITD-GNN are optimised with the Adam optimiser (Kingma and Ba, 2014), where the learning rate is set to 0.01, and weight decay is 0.01. In STZITD-GNN, $N_{epoch} = 20$, the hidden unit is set to 42, GAT head is set to 3. The baseline parameters are optimised using Adam Kingma and Ba (2014) with regularisation of L_2 and a dropout rate of 0.2. The GNNs in STZITD-GNNs and baselines are all two-layered. We also set the hidden units at 42. Similarly to previous work, we also employ the early stopping strategy with patience equal to 10 to avoid overfitting (Zhou et al., 2020a). STZITD-GNNs is implemented in Pytorch 1.9.0 with Python 3.8. All experiments were carried out on 1 NVIDIA GeForce RTX 3090, 24 GB.

4.3. Evaluation metrics

We use eight metrics to evaluate the performance of our STZITD-GNNs from four perspectives, including the predictive regression effect of road crash risk scores, the predictive effect of uncertainty-aware distributional characteristics, the accuracy of zero crash road identification, and the accurate identification of crashes occurring on predicted high-risk roads.

4.3.1. Crash accuracy metrics

For the aspect of the point estimate, we compare the prediction of the numerical crash risk scores. For probabilistic distribution outputs like our model, we use the mean value of the distribution as the numerical output. For specific metrics, we use Mean Absolute Error (MAE), Mean Absolute Percentage error (MAPE), and Root Mean Squared Error (RMSE) to evaluate the predicted crash risk scores of each road. Specific definitions are summarised:

$$\begin{aligned}
 MAE &= \frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N |y_{ij} - \hat{y}_{ij}|, \\
 MAPE &= \frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N \left| \frac{y_{ij} - \hat{y}_{ij}}{y_{ij}} \right|, \\
 RMSE &= \sqrt{\frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N (y_{ij} - \hat{y}_{ij})^2},
 \end{aligned} \tag{13}$$

where \hat{y}_{ij} and y_{ij} are the predicted (i.e. distribution mean) and ground-truth values of i th road at time slot j th respectively. $p = 14$ days, and the time interval j th is the predicted window index. Lower values of these metrics mean better performance with fewer errors.

4.3.2. Uncertainty quantification metric

To quantify the uncertainty of the results, we use two typical metrics: the mean prediction interval width (MPIW) and the prediction interval coverage probability (PICP) are based on a confidence interval 5%–95% with a confidence level set at 5% ($\alpha = 5\%$). MPIW gauges the

average width of the prediction interval, reflecting the distribution's specific characteristics. Ideally, a smaller MPIW indicates reduced variance and increased certainty, highlighting a more precise prediction. The PICP, on the other hand, is implemented to evaluate whether the ground truth value, denoted y_{ij} , lies within the predicted interval. A higher PICP indicates better performance. The goal is to achieve a prediction interval that is narrow and encompasses a significant number of data points.

$$\begin{aligned} \text{MPIW} &= \frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N (U_{ij} - L_{ij}), \\ \text{PICP} &= \frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N \mathbf{I}(U_{ij} < \hat{y}_{ij} < L_{ij}), \end{aligned} \quad (14)$$

L_{ij} and U_{ij} correspond to the lower and upper bound of the confidence interval for observation road i at time slot j th. The $\mathbf{I}(U_{ij} < \hat{y}_{ij} < L_{ij})$ equals 1 if the condition is true, else 0.

4.3.3. Zero crash identification metrics

Given the problem of zero-value data sparsity, highlighted in Fig. 3(b), it is essential to assess the ability of a model to accurately represent roads without crashes. We used the true zero rate (ZR) as a metric to quantify whether a model reflects the zero sparsity seen in the actual data. The formula is as follows:

$$\text{ZR} = \frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N \mathbf{I}(y_{ij} = 0 \cap \hat{y}_{ij} = 0), \quad (15)$$

where the indicator function, $\mathbf{I}(y_{ij} = 0 \cap \hat{y}_{ij} = 0)$, equals 1 if the condition holds true, and 0 otherwise.

4.3.4. Accuracy Hit Rate Metric on Crash Occurrence

To evaluate the effectiveness of our model in identifying roads with a higher likelihood of crash occurrence, we use the accuracy hit rate in a ($\text{AccHR}@a$) metric. Specifically, we focus on the top 20% ($a = 20\%$) of roads considered high-risk to verify if these areas actually correspond to actual crash occurrence. This alignment of predicted high-risk areas with real crash sites, as reported in Zhang and Cheng (2020), Zhang et al. (2019a), allows $\text{Acc}@a$ to effectively measure the precision of our model in forecasting roads at increased risk of crashes. The formula for calculating $\text{Acc}@a$ is as follows:

$$\text{AccHR}@a = \frac{1}{p} \sum_{j=1}^p \frac{\sum_{i=1}^N \mathbf{I}(y_{ij} > 0 \cap \hat{y}_{ij} \geq \text{Top}(\hat{y}:j,a))}{\sum_{i=1}^N \mathbf{I}(y_{ij} > 0)}, \quad (16)$$

The indicator function, $\mathbf{I}(y_{ij} > 0 \cap \hat{y}_{ij} \geq \text{Top}(\hat{y}:j,a))$, is rated as 1 when a crash (denoted as $y_{ij} > 0$ for a crash occurrence on the i th road in the j th time interval) aligns with the top a percent of predicted risks (where \hat{y}_{ij} meets or exceeds the threshold $\text{Top}(\hat{y}:j,a)$), and 0 otherwise.

4.4. Baseline models

In evaluating our proposed model, we carefully select baseline models that allow evaluation of the features of the proposed model: uncertainty quantification in multistep road-level crash risk prediction. Our selection criteria favour traditional point estimation methods and the most recent interval prediction models. This approach is in line with recent studies Zhuang et al. (2022). Our baselines include the mean-predict method (HA) and advanced spatial-temporal graph deep learning techniques such as STGCN and STGAT. Additionally, we consider parameterisation approaches such as STG-GNN, STNB-GNN, STTD-GNN, and STZINB-GNN. These probabilistic models have two or three parameter distributions, simpler than the four-parameter models we propose.

Our focus on GNN-based models is motivated by the inherent graph structure of road networks and the spatial dependencies in

crash risk (Zhou et al., 2020b). GNNs are uniquely suited to capture the complex relationships between road segments, incorporate heterogeneous road characteristics, and learn network-wide patterns while preserving local details (Wu et al., 2023). The evolution from traditional statistical methods to GNNs represents a significant advancement in modelling capabilities for road-level crash prediction. It is important to note that the ST-GNN architecture remains consistent across the probabilistic models (STG-GNN, STNB-GNN, STTD-GNN, STZINB-GNN, and STZITD-GNN), with the key distinction being the underlying statistical distribution used for parameter estimation in each model. This consistency allows for a direct comparison of different probabilistic approaches while controlling for the neural network architecture. Additionally, it also mitigates the boundary edge affected bias in the outcomes, as we compared all the methods using the same spatial network representation (Zhang and Cheng, 2020).

The baseline models' details are:

- **HA** (Liu and Guan, 2004): The historical average serves as the minimum baseline. It is calculated by averaging road traffic crashes in the same time intervals from historical data to predict the future road traffic crash.
- **STGCN** (Yu et al., 2018): Spatio-temporal graph convolution network, which utilises graph convolution and 1D convolution to capture spatial dependencies and temporal correlations, respectively.
- **STGAT** (Huang et al., 2019a): The Graph Attention Network (GAT), originally proposed by Veličković et al. (2017), enhances the representation of a node by focussing on its neighbouring nodes, assigning varying attention weights to each based on their importance. Building on the work of Huang et al. (2019a) and Veličković et al. (2017), we replace the Graph Convolutional Network (GCN) in the STGCN with GAT and then use LSTM to capture the temporal dependencies, resulting in the modified model, which we refer to as STGAT.
- **STG-GNN** (Wang et al., 2023b): Same as STZITD-GNN, we replace the statistical assumption of road traffic crashes with Gaussian distribution.
- **STNB-GNN** (Zhuang et al., 2022): Same as STZITD-GNN, we replace the statistical assumption of road traffic crashes with Negative Binomial Distribution.
- **STTD-GNN**: Same as STZITD-GNN, we replace the statistical assumption of road traffic risks with TD Distribution.
- **STZINB-GNN** (Zhuang et al., 2022): Unlike STZITD-GNN, we replace the statistical assumption of road traffic crashes with a zero-inflated negative binomial distribution.

4.5. Performance comparison

We evaluated the performance of our proposed model, STZITD-GNN, against several baseline models from four different aspects: accuracy of point estimation, uncertainty quantification, non-risk road identification and accurate prediction of crash occurrence, as shown in Tables 4. The results are indicated by upward arrows, representing improvements over the second-best models. These enhancements confirm the significant impact of probabilistic assumptions on the accuracy and stability of our results, consistent with recent findings by Wang et al. (2023b). The consistent performance of the STZITD-GNN model underscores the critical importance of selecting appropriate distributional assumptions that align with the characteristics of traffic crash data.

For point estimation accuracy, the superior performance of the STZITD-GNN model across all regression error metrics highlights the effectiveness of the ZITD distribution. This distribution captures the long-tail dynamics of road traffic crash risk data with greater accuracy, making the mean value derived from the four parameters of the distribution decoder more representative of the actual risk scores. STZITD-GNN significantly improves the accuracy of the estimation of the crash

Table 4

Comparison of multi-step prediction performance (14 days) in three boroughs. The bold font indicates the best performing model, while underline indicates the second best. To maintain clarity, detailed names of all GNNs in the final five statistically integrated deep learning models are omitted. Instead, abbreviations such as STG, STNB, STTD, STZINB, and STZITD are used to emphasise the most significant modifications within the GNN models.

Metric	Model	HA	STGCN	STGAT	STG	STNB	STTD	STZINB	STZITD
Dataset		Lambeth							
Point Estimation	MAE	0.1348	0.0613	<u>0.0302</u>	0.1182	0.0802	0.0315	0.0544	0.0238 _{21.19%}
	MAPE	0.4142	0.4023	<u>0.3203</u>	0.4046	0.0363	0.0426	<u>0.0265</u>	0.0135 _{49.06%}
	RMSE	0.2107	0.1436	0.1122	0.1827	0.1390	<u>0.1098</u>	0.1191	0.0947 _{113.75%}
Uncertainty Quantification	MPIW	/	/	/	0.4871	0.1946	0.1349	<u>0.0454</u>	0.0204 _{55.07%}
	PICP	/	/	/	0.7328	0.8903	<u>0.9873</u>	0.9796	0.9899 _{100.26%}
Non crash Quantification	ZR	0.5203	0.5305	0.5986	0.1992	0.5621	<u>0.6838</u>	0.6408	0.7870 _{115.09%}
Hit Rate	<i>AccHR@20</i>	0.4520	0.6113	0.6422	0.2666	0.4471	<u>0.7123</u>	0.6184	0.7659 _{107.52%}
Dataset		Tower Hamlets							
Point Estimation	MAE	0.0939	0.2385	<u>0.0305</u>	0.1741	0.2475	0.0433	0.2135	0.0271 _{111.15%}
	MAPE	0.2896	0.2017	<u>0.1205</u>	0.2662	0.0399	<u>0.0246</u>	0.0594	0.0221 _{110.16%}
	RMSE	0.1987	0.1325	0.1355	0.3896	0.1548	<u>0.1186</u>	0.1368	0.0918 _{122.60%}
Uncertainty Quantification	MPIW	/	/	/	0.5321	0.2263	0.0810	<u>0.0296</u>	0.0156 _{147.30%}
	PICP	/	/	/	0.6439	0.9442	<u>0.9845</u>	0.9845	0.9871 _{100.26%}
Non crash Quantification	ZR	0.4998	0.4896	0.5234	0.1351	0.5158	0.6420	<u>0.7526</u>	0.8876 _{117.94%}
Hit Rate	<i>AccHR@20</i>	0.4752	0.5869	<u>0.6950</u>	0.2647	0.5022	0.6368	0.5827	0.7224 _{103.94%}
Dataset		Westminster							
Point Estimation	MAE	0.1467	0.0784	0.0847	0.1055	0.3266	<u>0.0579</u>	0.3240	0.0357 _{138.34%}
	MAPE	0.1884	0.8976	0.6012	2.9295	0.0330	0.0973	0.0415	0.0259 _{121.52%}
	RMSE	0.5999	0.2384	0.2433	0.6681	0.2909	<u>0.1552</u>	0.2751	0.1015 _{134.60%}
Uncertainty Quantification	MPIW	/	/	/	0.8542	0.6272	0.2021	<u>0.0493</u>	0.0259 _{147.46%}
	PICP	/	/	/	0.6036	0.9363	<u>0.9801</u>	0.9796	0.9893 _{100.94%}
Non crash Quantification	ZR	0.5981	0.1431	0.2922	0.2071	0.5393	0.5079	<u>0.6288</u>	0.7328 _{116.54%}
Hit Rate	<i>AccHR@20</i>	0.4217	0.5020	0.4808	0.2933	0.4503	<u>0.6075</u>	0.5139	0.6898 _{113.55%}

risk score, demonstrating a 49.06% increase in MAPE compared to the second-best model in the Lambeth case study, which has the most significant zero-inflation. Traditional deterministic models like HA, STGCN, and STGAT show varied performance, with HA consistently underperforming. GAT-based models, utilising predefined adjacency matrices that embed domain spatial knowledge, generally outperform GCN models by incorporating spatial proximity and contextual similarity, thereby better representing spatial dependencies. Among probabilistic models, those that assume Gaussian distributions perform poorly. They are significantly outperformed by deterministic models and show far worse performance 50% compared to the STZITD-GNN model, which uses skewed distribution assumptions. This underscores the importance of selecting appropriate distributional assumptions for probabilistic models.

Second, the variation in uncertainty quantification across different models highlights the significant impact of distributional assumptions. The four-parameter distribution model (STZITD-GNN) shows substantial improvements, most notably in effectively capturing a narrow distribution while encompassing a wider range of actual crash risk values. In contrast, the worst-performing model, STG-GNNs, illustrates how improper distributional assumptions can degrade the quality of the estimation, introducing greater uncertainty into the results. The MPIW of this model is approximately 1.5 times that of the least uncertain three-parameter model, due to its constraints that enforce equal mean and variance in the distribution. Although ideally, outcome distributions would have a narrow interval, traditionally a wider MPIW suggests a higher PICP, indicating a trade-off in measuring uncertainty (Wang et al., 2023b; Zhu et al., 2022; Zhuang et al., 2022). The STZINB-GNN and STTD-GNN models show that in scenarios characterised by extreme zero-inflation and heavier tails, a higher MPIW is associated with better PICP performance. This correlation suggests that broader prediction intervals, which encompass more observations, are beneficial in more extreme cases as they cover a wider range of observed values, despite reducing the ability to achieve narrower intervals. However, the STZITD-GNN model shows a promising improvement in MPIW, ranging

from 47.30% to 55.07% in all case studies, highlighting its efficacy. In terms of PICP, it also shows a marginal increase of approximately 0.2% over the next best model, STTD-GNNs. This modest improvement still validates the impact of leveraging long-tailed and zero-inflated TD distributions, which contribute to a larger PICP and reduced variance in the predicted outcomes.

In scenarios characterised by a high prevalence of non-crash roads, the STZITD-GNN model not only aims to precisely quantify crash risk values but also effectively differentiates non-crash roads. The integration of the sparsity parameter (π) significantly enhances the ability of the traditional TD assumption to address the prevalent zeros in the dataset, thereby reducing uncertainties associated with zero outcomes. The TD and ZINB distributions provide distinct parametric approaches to tackle the issue of zero-inflation. The TD distribution naturally integrates zero-inflation through its end-to-end structure, while the ZI parameter in the ZINB model requires an additional interpretative step like ZITD model does. For example, in the Tower Hamlets and Westminster cases, the ZI parameter significantly increases performance in zero-rate predictions compared to the NB distribution, showing improvements of 16.54% and 17.94%, respectively. In extreme zero-inflation scenarios, such as in Lambeth, both the STTD and STZINB models perform poorly, with a 15.09% discrepancy in ZR results compared to the STZITD-GNN model. Although the ZI paradigm is integrated into both the TD and the NB models, the ZITD model has a continuous parameter space, which can be advantageous for fitting data with a mix of zero inflation and continuous positive values. In contrast, traditional ZINB models operate under a discrete probabilistic assumption and struggle with complex data structures, particularly when the positive values are continuous rather than merely counts.

In analyses that correlate high-risk roads with actual crashes, the STZITD-GNN model consistently demonstrates robust performance. In particular, it achieved a significant 76.59% precision in identifying actual crashes in the Lambeth case. Across all three distinct cases, the model maintains a robust and accurate identification rate, with a variance of only 7.61%. In contrast, models employing unsuitable

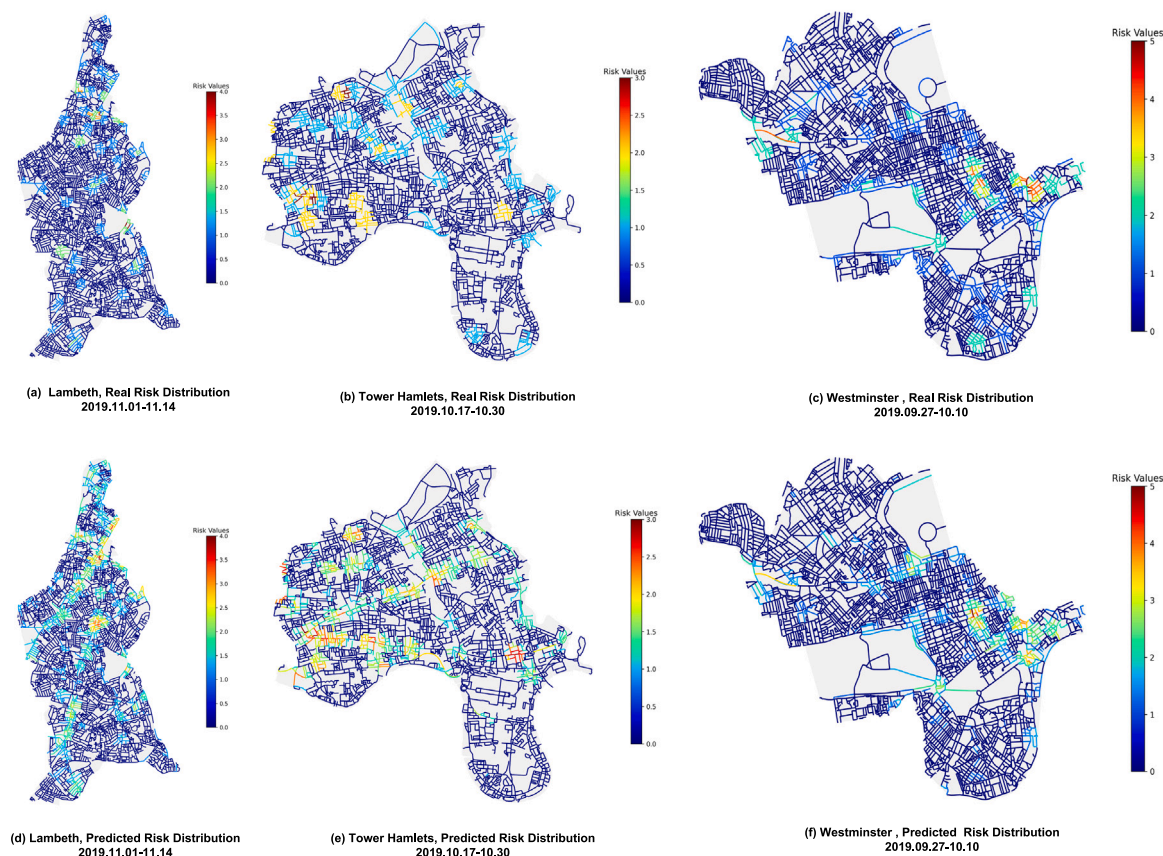


Fig. 4. Real v.s. Predicted Crash Risks on Lambeth, Tower Hamlets, and Westminster.

distribution assumptions, such as STG-GNN, tend to significantly underpredict crash occurrences, often yielding the poorest results. This discrepancy underscores the importance of selecting the appropriate distribution model again.

4.5.1. Visualisation of predicted road-level crash

To illustrate the results, we have generated Choropleth maps of road-level crashes across three distinct boroughs over a 14-day period. Considering the sparsity of crash data, which complicates the visualisation, we selected the three most risk-intensive periods of the data for display, as shown in Fig. 4. Each borough reveals distinct crash patterns that correlate closely with the model's predictions, offering a granular view of its predictive accuracy.

These visualisations highlight the general sparsity of traffic crash data, with most roads being non-risk (blue). Only a few roads exhibit higher crash risk levels (red), while others are associated with low-risk crashes. The STZITD-GNN model excels in identifying a range of risk levels, from zero-risk to high-risk. It performs particularly well in Westminster, accurately identifying a broader range of risk-prone roads, aided by the area's lower zero inflation rate and denser graph structure. Furthermore, the model consistently and accurately maps the spatial connectivity of high-risk crash roads across all three boroughs, with risk areas typically clustering along major thoroughfares such as A11 Whitechapel Road and A13 in Tower Hamlets and A3 in Lambeth.

In the three distinct boroughs, the STZITD-GNN model's ability to handle diverse spatial patterns and variations in daily crashes demonstrates its adaptability and consistent performance. However, the model occasionally misclassifies zero-risk roads as low-risk, particularly in Lambeth and Tower Hamlets. Although these errors generally fall within lower-risk categories and do not significantly impact the identification of higher-risk roads with actual crash occurrences, refining this aspect of the model could further improve its overall accuracy.

4.6. Sensitivity analysis

4.6.1. Temporal stability of metrics

Evaluating temporal stability is crucial for evaluating a model's performance over multiple prediction time steps, particularly in data-driven methods, where superior results in specific instances might mask potential underfitting issues (Ding et al., 2022; Mannering, 2018). Our primary goal is to achieve consistently low uncertainty, as indicated by the MPIW metric in three boroughs, depicted in Fig. 5. It illustrates the temporal variation of MPIW for different models over the 14-day prediction horizon. The x -axis represents each day of the prediction period, allowing us to observe how the uncertainty (as measured by MPIW) changes as we forecast further into the future. We strive to maintain a narrow yet comprehensive MPIW throughout the prediction steps to ensure reliable prediction intervals for traffic crashes.

Analysis shows that the MPIW of the top performing models, including STZITD-GNNs, is significantly lower than that of the baseline models, emphasising their superior performance throughout the study period. Therefore, we focus on the temporal dynamics of these leading models, presenting baseline performances as averages for contextual clarity.

As indicated in Fig. 5, the MPIW is typically lowest on the first day of testing and generally increases over time. However, this increase is smoother in our STZITD-GNNs model compared to the more erratic fluctuations observed in baseline models. This smoother increase reflects our model's ability to handle the inherent uncertainties and the variability of risk levels, which might be less accurately captured by models with misspecified distributions. Moreover, boroughs like Westminster and Lambeth, known for higher instances of zero-crash risk values, show more fluctuating MPIWs, suggesting greater uncertainties in prediction intervals due to the sparsity of traffic crashes. Despite these challenges, our model consistently delivers stable and accurate predictions across different boroughs.

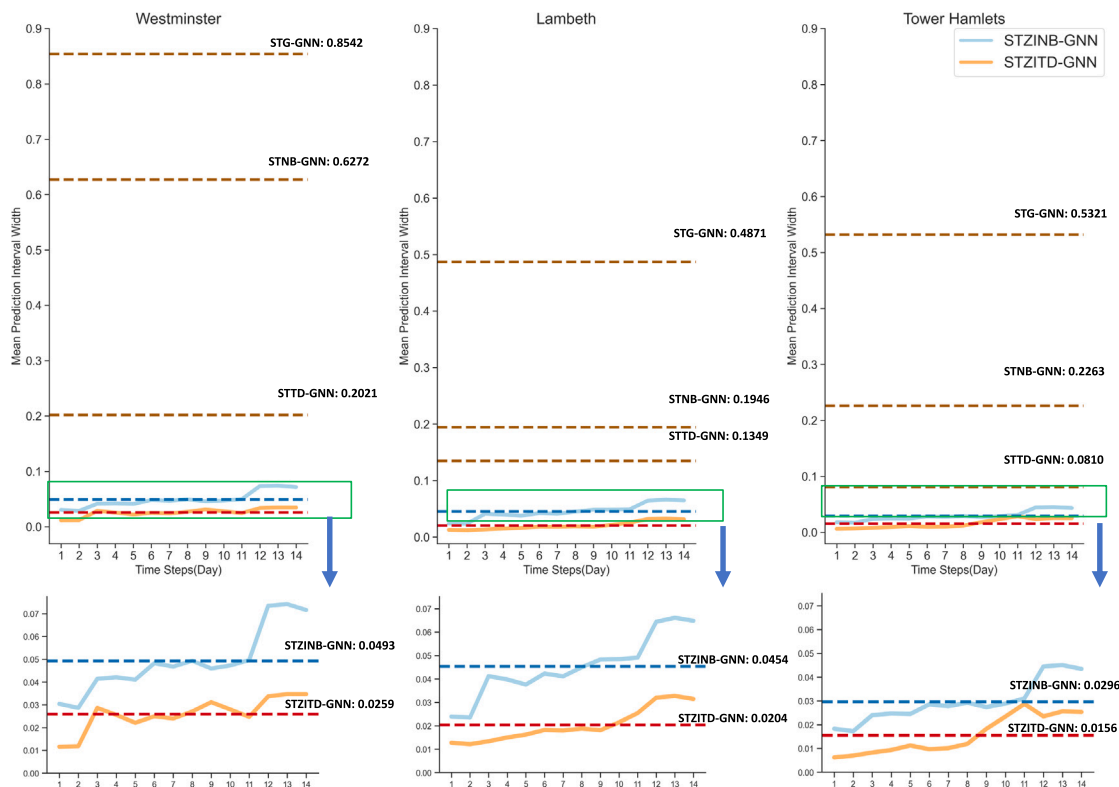


Fig. 5. Temporal variation of MPIW for baseline and STZITD-GNN models.

In the evaluation of the temporal stability for the estimation of points, depicted in Fig. 6, the STZITD-GNN customised approach emphasises the advantages of applying the most appropriate probabilistic assumptions, enhancing spatiotemporal representations, and managing the various aspects of traffic crashes. It is apparent that the STG-GNN model consistently shows the weakest performance across all temporal prediction periods, with deterministic models also performing poorly. Furthermore, models based on the Negative Binomial (NB) distribution exhibit unsatisfactory performance characterised by higher fluctuation and higher error rates. This inadequacy stems from the focus of NB models on binary crash-occurrence outcomes, which often overlook the complexities of long-tail crash risk values. In contrast, our model exhibits significantly less variance across all three evaluated metrics at each time step, demonstrating stability even towards the end of the prediction horizon. This robust performance is attributed to the model’s ability to dynamically fine-tune prediction intervals using four critical parameters, effectively addressing the challenges posed by the sparsity and skewed distribution of traffic crash data.

In terms of the zero rate, deterministic deep learning models tend to conflate no-risk with lower-risk values. This occurs because deterministic models do not adequately differentiate between the absence of risk and the presence of minimal risk. As a result, these models often misclassify low-risk scenarios as zero-risk, particularly in situations with extreme zero inflation, leading to overfitting. For the hit rate, the issue is inversely presented as low-risk predictions often being incorrectly classified as no-risk. This misclassification is due to the model’s inability to capture subtle variations in risk levels.

The final component of our evaluation focusses on quantifying both non-crash roads and high-risk roads where actual crashes occurred, as shown in Fig. 7. The STZITD-GNN model excels consistently in both metrics, maintaining stable performance with less than 18% variance throughout the 14-step period. Unlike deterministic models, which exhibit less variability but struggle significantly with accurately

identifying true noncrash roads, often misclassifying all roads as having no crash risk, the STZITD-GNN model effectively addresses these challenges.

In scenarios of extreme zero inflation, deterministic models frequently misclassify roads, conflating no-risk with lower risk values, thereby leading to overfitting. Conversely, probabilistic models using Gaussian distributions struggle to manage the high prevalence of zero values and lower risk values, resulting in poor performance and high sensitivity in the ZR metrics over time. This problem occurs because each time step may present unique scenarios of zero-valued roads, which Gaussian models often misinterpret due to their uniform mean and variance assumptions.

In the identification of actual crashes on high-risk crash roads, as illustrated in Fig. 7(b), the ZITD distribution performs exceptionally well. It employs a weighted dispersion parameter (ϕ) that effectively distinguishes lower crash risks from non-crash roads, accurately reflecting the distribution of low, medium, and high-risk roads where actual crashes occurred, shown as long-tailed data distributions. This adaptability to the complex temporal dynamics of road-level crash data at each time step not only mitigates overfitting for non-crash scenarios, but also stabilises the model performance, enabling it to reliably capture real crash occurrences in critical geographic locations.

4.6.2. Analysis of long-tail crash data characteristics

To explore the impact of inherently long-tailed data on our model, we created a three-dimensional surface plot over four information sources, as illustrated in Fig. 8. Due to visualisation constraints, it is not feasible to represent all four parameters in a single plot. We therefore focus on the TD parameters ϕ , ρ , and μ , with the understanding that the sparsity parameter π significantly influences the original TD parameters. This plot illustrates the predicted crash values, with the actual crash values y represented by the colour gradient.

A distinct trend is evident in the learnt values of ρ , consistently approaching 2, suggesting a preference for more skewed distributions

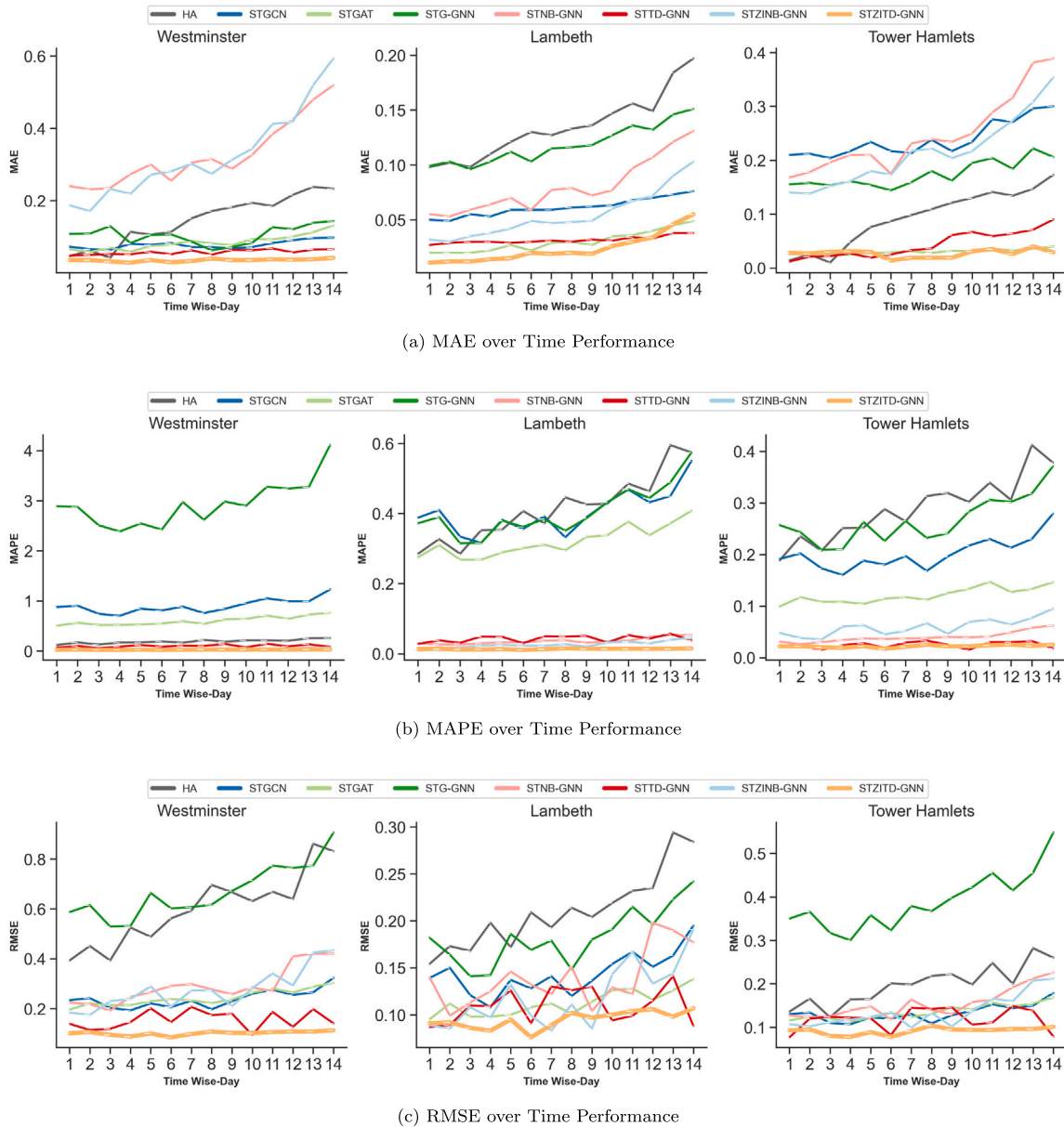


Fig. 6. Time-wise Comparison for the Point Estimation Metrics.

as ρ increases from 1 to 2 (Tweedie et al., 1984; Kurz, 2017). This trend corresponds with the long-tailed nature of our data and validates the probabilistic assumptions. The increase in ρ serves as an adaptive response to the component of the loss function $y \frac{\mu^{1-\rho}}{1-\rho}$ ($y > 0$), effectively penalising the underestimation of true, although infrequent, higher values of crash risk within the long tail. This adjustment directs the model to focus more on such kinds of traffic crash rates, continually benefiting from significant enhancements in model sensitivity and accuracy.

Additionally, significant variations in the values ϕ are observed, particularly concerning data with zero value, which serve as the dispersion parameter. For example, in Westminster, ϕ is significantly lower than in other boroughs, indicating fewer zero-inflation issues. Traditionally, ϕ in the TD distribution is crucial for managing the frequency of zero occurrences to effectively address zero inflation. However, due to the highly imbalanced nature of our data, a single ϕ parameter often fails to capture the prevalence of zeros, as demonstrated by the overlap of plots in areas of lower and zero crash values. In our ZITD model, an enhanced ϕ , augmented by the sparsity parameter π , manages

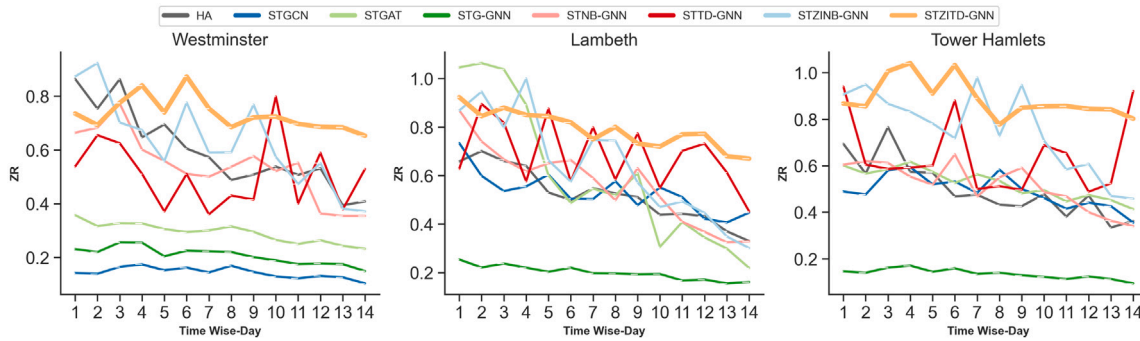
these features, offering a precise depiction of the inherent sparsity and discrete uncertainty in road-level crash predictions.

Moreover, the model parameter μ , which denotes the mean of the predicted crash values, shows a strong correlation with the actual crash values y , as evidenced by the colour alignment in the graphs. This correlation not only substantiates the effectiveness of the model in estimating mean points, but also affirms its capability to accurately mirror and adapt to the complexities of urban crash data.

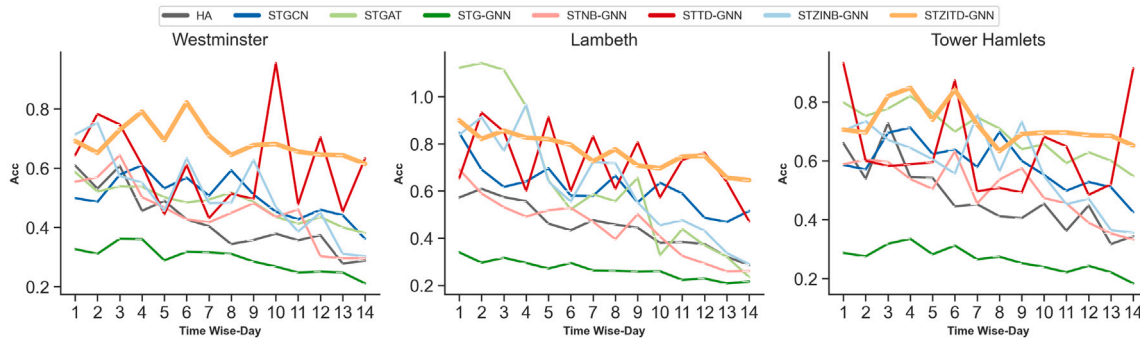
5. Discussion, limitation and conclusion

5.1. Discussion

This paper introduces STZITD-GNN, a probabilistic deep learning framework that, to the best of our knowledge, is the first model of uncertainty-awareness specifically tailored for the prediction of road-level crashes on a daily basis. Developed from data-driven insights, the STZITD-GNN utilises predefined distribution characteristics of traffic

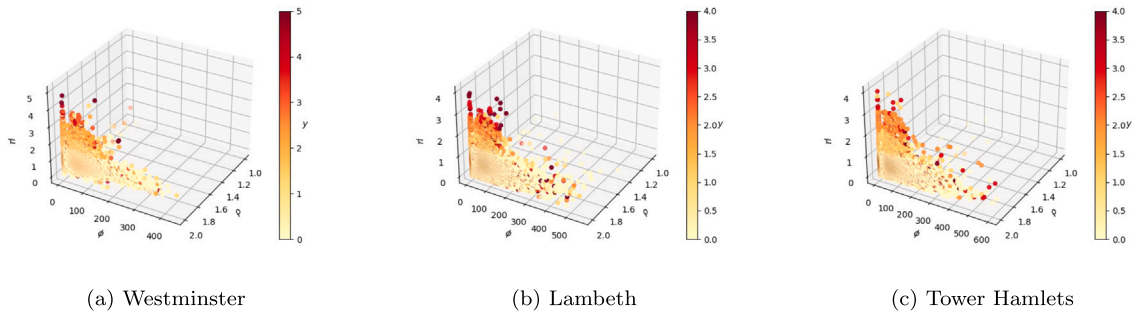


(a) Zero Rate over Time Performance



(b) Accurate Hit Rate on Crash Occurrence at Top 20% Risky Roads over Time Performance

Fig. 7. Time-wise Comparison for the Non-crash Roads Quantification and Accurate Crash Hit Metrics.



(a) Westminster

(b) Lambeth

(c) Tower Hamlets

Fig. 8. Three Tweedie parameter visualisation on Lambeth, Tower Hamlet, and Westminster with multi-step predict. The colour bar shows the predicted crash from the highest to zero.

crash data to enhance the representation of sparse traffic crash scenarios. It effectively detects finer-level correlations in traffic crashes, thereby delivering precise predictions with narrow prediction intervals and reduced uncertainties. Through systematic multi-timescale experiments conducted on three boroughs of London, our STZITD-GNN model demonstrated superior performance across all metrics, with improvements of up to 49.06% in MAPE for point estimation, 55.07% in MPIW for uncertainty quantification for all levels of road risk, and 76.59% precision in identifying actual crashes in high-risk areas. The key contributions of this study are summarised as follows.

- Innovative Distribution Extension:** We have adapted the conventional TD to a ZITD distribution, introducing a sparsity parameter, π , integrated with the dispersion parameter ϕ . This modification tailors the model to our data-driven assumptions, representing the dispersed and extreme zero-inflated nature of crash data at the road level. Therefore, our methodology is specifically designed to avoid the pitfall of focusing solely on high-risk roads. Instead, it offers a holistic view of road safety, considering the full range of risk profiles.

- Integration with ST-GNN:** The incorporation of the ZITD encoder into the ST-GNN framework makes the STZITD-GNN model an end-to-end framework to handle the complex spatiotemporal dynamics of crashes. It surpasses traditional point estimation methods by enabling the mapping of output variance and prediction intervals. Additionally, the model facilitates an interpretable, uncertainty-aware process that elucidates the underlying factors contributing to improved prediction accuracy.
- Empirical Validation Across Boroughs:** The model has shown improved predictive accuracy and minimal uncertainties compared with benchmark models in three London boroughs, each featuring unique spatial crash patterns in multi-step scenarios. This performance confirms the robustness and generalisability of the model across diverse urban settings, even with sparse datasets. Additionally, the model's stable performance across heterogeneous datasets, which vary in crash determination criteria, further underscores its reliability and effectiveness in urban traffic risk assessment.

In addition, it should be recognised that the challenges of overdispersion and zero inflation extend beyond traffic crashes, affecting a wide array of urban issues such as crime rates (Zhang and Cheng, 2020). This universality suggests that our proposed model could be applicable to a broader range of contexts beyond its current usage.

5.2. Limitations

The computational requirements of our STZITD-GNN model require further consideration, especially when considering its application to larger urban areas or more extensive road networks. As the complexity of the graph structure increases with the number of road segments, the time complexity of our model grows correspondingly (Tang et al., 2022). Future research should explore advanced techniques in AI efficiency and graph compression to address these scalability challenges.

Furthermore, our parameter-based GNN model aims to interpret different types of networks, recognising the complexities of urban road structures and their potential impact on crashes. As demonstrated by Amini et al. (2016), road configurations exhibit distinct structural properties and traffic patterns. For instance, grid networks, characterised by high cyclomatic numbers and degrees of connectivity, may require a more nuanced understanding of crash risks at intersections due to their regular structure. In contrast, radial networks, with lower connectivity measures, emphasise central nodes and primary arterials, potentially influencing crash patterns differently, especially under high traffic demand conditions. These varying patterns highlight the need for future research to explore how best to incorporate diverse structural characteristics into the GNN framework, enhancing our understanding of their influence on crash occurrences across different urban contexts.

While our model does not explicitly include traffic flow data as a measure of exposure, previous studies have shown that historical crash characteristics and characteristics of the road network can serve as practical embeddings in crash prediction models, supporting the self-supervised learning paradigm (Zhang and Cheng, 2020; Jin et al., 2023, 2024; Cui et al., 2024). The strong performance of our model suggests that the combination of spatial, temporal, and road characteristics effectively captures the underlying patterns of crash risk. However, future work that incorporates direct measures of traffic exposure could potentially further refine prediction accuracy and enhance the generalisability of the model across diverse urban settings.

Addressing these limitations in future research will contribute to developing more comprehensive, efficient, and broadly applicable models for traffic crash prediction, capable of adapting to the diverse and complex nature of urban environments around the world.

5.3. Conclusion

In conclusion, the development of the STZITD-GNN model marks a significant advancement in the prediction of urban road-level traffic crashes. The high accuracy and granularity of our model's predictions have significant implications for urban planning and traffic safety policies. By identifying high-risk road segments with greater precision, city planners and traffic engineers can prioritise safety interventions more effectively. For instance, resources for road improvements or increased traffic enforcement could be allocated to areas predicted to have the highest crash risks. Furthermore, the ability of the model to quantify the uncertainty in its predictions can help decision makers to assess the confidence level of risk assessments in different areas, potentially informing more nuanced and targeted safety strategies. However, this is just the beginning of an ongoing effort to refine and improve this model. By integrating additional data sources and continually improving our methodologies, our goal is to develop a tool that not only mitigates traffic risks but also significantly contributes to global urban safety strategies.

CRedit authorship contribution statement

Xiaowei Gao: Writing – review & editing, Writing – original draft, Visualization, Resources, Methodology, Data curation, Conceptualization. **Xinke Jiang:** Methodology, Conceptualization. **James Haworth:** Writing – review & editing, Writing – original draft, Supervision. **Dingyi Zhuang:** Methodology, Formal analysis, Conceptualization. **Shenhao Wang:** Writing – review & editing, Supervision. **Huanfa Chen:** Writing – review & editing, Supervision. **Stephen Law:** Writing – review & editing.

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.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) utilised GPT-4 for preliminary grammar checks and proofreading. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Appendix A. Tweedie table

Tweedie EDMs	ρ	$\kappa(\theta)$	ϕ	θ	$a(y, \phi)$
Normal	0	$\frac{\theta^2}{2}$	σ^2	μ	$\exp(-y^2)/2\sigma^2 - \log 2\pi/2$
Poisson	1	$\exp(\theta)$	1	$\log \mu$	$1/y!$
Poisson-Gamma	(1,2)	$\frac{(\theta-\rho\theta)^{\frac{2-\rho}{2}}}{2-\rho}$	ϕ	$\frac{\mu^{1-\rho}}{1-\rho}$	Eq. (B.1)
Gamma	2	$-\log(-\theta)$	ϕ	$-1/\mu$	$\frac{\phi^{-1/\phi} y^{1/\phi-1}}{\Gamma(1/\phi)}$

where μ, σ is the mean and the standard values of Gaussian Distribution.

Appendix B. Tweedie parameters

The frequency of traffic crashes C_k adheres to a Poisson distribution represented by $Pois(\lambda)$ with a mean of λ , expressed as $C_k \sim Pois(\lambda)$. This is natural, as frequency often appears as the main prediction target in the previous literature (Zhou et al., 2020b; Trirat et al., 2023; Bonera et al., 2024). Currently, the severity of the crash $l_k^{(j)}$ is modelled as independent and identically distributed *i.i.d* Gamma random variables, $C_k \perp l_k^{(j)}$ and $\forall j$, denoted as $Gamma(\alpha, \gamma)$. These variables have a mean of $\alpha\gamma$ and a variance of $\alpha\gamma^2$, such as $l_k^{(j)} \stackrel{iid}{\sim} Gamma(\alpha, \gamma)$. C_k and $l_k^{(j)}$ together form the model of the occurrence of a single crash. Consequently, the traffic crash risk y_k is formulated as a Poisson sum of these *i.i.d*. Gamma random variables. Considering Eq. (1) and considering $1 < \rho < 2$, we can finally reformulate the random variable y_k as follows:

$$y_k = \begin{cases} 0 & \text{if } C_k = 0, \\ \sum_{j=1}^{C_k} l_k^{(j)} = l_k^{(1)} + l_k^{(2)} + \dots + l_k^{(C_k)} & \text{if } C_k > 0. \end{cases} \quad (\text{B.1})$$

In the scenario where $C_k = 0$, we correspondingly have $y_k = 0$. The probability density at zero for traffic crash risk, therefore, becomes $\mathbb{P}(y_k = 0) = \mathbb{P}(C_k = 0) = \exp(-\lambda)$. In contrast, when $C_k > 0$, y_k is deduced from the sum of C_k independent Gamma random variables. For clarity, following the work of Zhou et al. (2022a), we re-parametrise

the TD distribution, setting $\theta = \mu^{1-\rho}/(1-\rho)$ and $\kappa(\theta) = \mu^{2-\rho}/(2-\rho)$. This results in the revised TD distribution as:

$$f_{TD}(y_k|\theta, \phi) = f_{TD}(y_k|\mu, \phi, \rho) = a(y_k, \phi, \rho) \exp\left[\frac{1}{\phi}\left(y_k \frac{\mu^{1-\rho}}{1-\rho} - \frac{\mu^{2-\rho}}{2-\rho}\right)\right], \quad (\text{B.2})$$

where

$$a(y_k, \phi, \rho) = \begin{cases} 1 & \text{if } y_k = 0, \\ \frac{1}{y_k} \sum_{j=1}^{\infty} \frac{y_k^{-j\alpha} (\rho-1)^{\alpha j}}{\phi^{j(1-\alpha)(2-\rho)} j! \Gamma(-j\alpha)} & \text{if } y_k > 0. \end{cases} \quad (\text{B.3})$$

In this case, μ, ϕ and ρ are the three parameters that govern the probability and expected value of the risk of traffic crashes. Parameters in the Gamma and Poisson distributions (λ, α, γ) can be obtained as (Kuruz, 2017; Halder et al., 2019; Mallick et al., 2022), :

$$\lambda = \frac{1}{\phi} \frac{\mu^{2-\rho}}{2-\rho}, \quad \alpha = \frac{2-\rho}{\rho-1}, \quad \gamma = \phi(\rho-1)\mu^{\rho-1}. \quad (\text{B.4})$$

Appendix C. Temporal encoder – GRU

Traffic crashes demonstrate a significant temporal correlation, underscoring the idea that the dynamic temporal features of the roads encapsulate a wealth of information. This observation has been acknowledged and exploited in prior methodologies (Wang et al., 2021; Zhao et al., 2019). As shown in Fig. 2, to seize these temporal proximity features, we implement a GRU (Chung et al., 2014). The GRU, in comparison to traditional LSTM networks, offers the advantage of learning relatively long-term dependencies without facing issues of vanishing and exploding gradients. In the GRU framework, there isn't a segregation of internal and external states, as seen in LSTM networks. Instead, it addresses the gradient vanishing and exploding problem by directly integrating a linear dependency between the current network state h_t and the preceding state h_{t-1} . Consequently, while a GRU preserves the capabilities of an LSTM network, its architecture remains more streamlined (Gu et al., 2019; Mahmoud et al., 2021; Ma et al., 2022).

Specifically, this component accommodates the feature sequence over past time windows $1:t$, thereby learning temporal dependencies. We represent the temporal embedding as \mathcal{Z}_T , taking into account both historical spatiotemporal road features and traffic crashes as input:

$$\mathcal{Z}_T = \text{GRU}(X_{1:t}, Y_{1:t}), \quad (\text{C.1})$$

Once the input features have been processed via the GRU, we obtain historical temporal embeddings \mathcal{Z}_T for all roads. These are subsequently directed to graph neural encoders. The GRU, in particular, consists of two gates: the update gate and the reset gate. The encoder processes a single input X_t, Y_t at time slot t and prior hidden features h_{t-1} in sequence:

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, X_t, Y_t] + b_r), \\ u_t &= \sigma(W_u \cdot [h_{t-1}, X_t, Y_t] + b_u), \\ \tilde{h}_t &= \tanh(W_c \cdot [r_t \cdot h_{t-1}, X_t, Y_t] + b_c), \\ h_t &= (1 - u_t) \cdot h_{t-1} + u_t \cdot \tilde{h}_t \end{aligned} \quad (\text{C.2})$$

In the equation above, h_t and h_{t-1} represent hidden features at the current and preceding time steps, respectively, while \tilde{h}_t denotes the candidate state that temporarily stores information from reset and update gates (r_t, u_t). The variables W and b (i.e. $W_r, W_u, W_c, b_r, b_u, b_c$) represent learnable weight and bias matrices, respectively. Importantly, r_t acts as the reset gate, determining the extent of integration of hidden features from previous time steps into features at the current time step. Meanwhile, u_t serves as the update gate, controlling how much information from previous time steps will be carried over to the next time step. The $\tanh(\cdot)$ is the hyperbolic tangent function. We perform the GRU gate operations step by step and select the final layer's hidden features h_t as the temporal embedding $\mathcal{Z}_T \in \mathbb{R}^{N \times F}$, where F denotes the dimension of road temporal embedding.

Appendix D. Spatial encoder – GAT

Within our research, road data is fundamentally perceived as having an inherent graph structure. This perspective facilitates the interpretation of node features as graph signals. To accommodate such graph-structured information, we employ the GAT (Veličković et al., 2017), an effective methodology renowned for enabling potential model deepening while concurrently incorporating dynamic spatial correlation into our network, achieved through real-time adjustment of attention weights in response to data changes (Wang et al., 2022; Zhang et al., 2019b).

Distinct from GCN of a spectral-domain convolution (Kipf and Welling, 2016a), GAT operates as a spatial-domain convolutional network. Instead, GAT emphasises the magnitude of the graph signal value at distinct points and their inter-distance to execute the convolution process. This trait empowers GAT to alleviate the over-smoothing issue associated with GCN deepening, thus increasing prediction precision regarding peaks and valleys (Zhao et al., 2019).

Initially, we compute the attention coefficient α_{ij} between roads v_i and v_j as per the following equation:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\bar{a}^T [W_a \mathcal{Z}_{T_i} \| W_a \mathcal{Z}_{T_j}]))}{\sum_{A_{i,o} > 0} \exp(\text{LeakyReLU}(\bar{a}^T [W_a \mathcal{Z}_{T_i} \| W_a \mathcal{Z}_{T_o}]))}, \quad (\text{D.1})$$

In this equation, $A_{i,o} > 0$ represents the neighbourhood or connected roads of the road v_i . $W_a \in \mathbb{R}^{F \times F'}$ and $\bar{a} \in \mathbb{R}^{2F'}$ are learnable weights, where F' denotes the hidden dimension. We apply the concatenation function $[\cdot \| \cdot]$ to concatenate the embedding of v_i and its neighbours, utilising the learnable parameter W_a and mapping function \bar{a} . Subsequently, we employ an activation function $\text{LeakyReLU}(\cdot)$, and normalise the attention coefficient α_{ij} using the $\exp(\cdot)$ function. The embedding of road v_i is consequently updated with a linear combination of the neighbourhood of road v_i and itself, wherein these computed attention coefficients are used.

In order to preemptively address potential inaccuracies in attention assignments, we have incorporated multiple attention assignments, otherwise known as multiple attention heads, into our approach. This tactic greatly amplifies the expressive capacity of the model. The multi-head attention mechanism facilitates concurrent consideration of information from a wide range of representation subspaces at various positions. This enables each head to learn and focus on different aspects of the input data, thus improving the model's ability to comprehend and accurately represent the data (Reza et al., 2022; Tang and Zeng, 2022; Ding et al., 2023). Herein, M symbolises the number of attention mechanisms, with the output embeddings of these mechanisms being concatenated in the following manner:

$$\mathcal{Z}_i = \parallel_{m=1}^M \sigma \left(\sum_{A_{i,j} > 0} \alpha_{ij}^m W_a^m \mathcal{Z}_{T_j} \right), \quad (\text{D.2})$$

In this equation, superscript m denotes a different head. The embedding of the road v_i is optimised by aggregating the embedding of its neighbours with the attention coefficients α_{ij}^m , and we apply learnable weights W_a^m for different heads. After this, we apply a nonlinear sigmoid function and concatenate different heads of embedding. The multi-head attention mechanisms hereby ensure stability in the attention mechanisms learning process and encapsulate diverse perspectives of road-level graph features. In particular, in Eq. (D.2), the dimension of the output feature is $M \times F'$, rather than F' . To ensure consistency in the output feature's dimension, we compute the average of each road's embedding in the last layer as follows:

$$\mathcal{Z}_i = \sigma \left(\frac{1}{M} \sum_{m=1}^M \sum_{A_{i,j} > 0} \alpha_{ij}^m W_a^m \mathcal{Z}_{T_j} \right). \quad (\text{D.3})$$

Finally, we input temporal embedding \mathcal{Z}_T into GAT to capture spatial dependency and obtain spatiotemporal $\mathcal{Z}_i \in \mathbb{R}^{F'}$ of road v_i . After that, we use \mathcal{Z} to infer the four-parameter STZITD-GNNs model.

References

- Abe, H., Yadohisa, H., 2017. A non-negative matrix factorization model based on the zero-inflated tweedie distribution. *Comput. Statist.* 32 (2), 475–499.
- Amini, B., Peiravian, F., Mojarradi, M., Derrible, S., 2016. Comparative analysis of traffic performance of urban transportation systems. *Transp. Res. Rec.* 2594 (1), 159–168.
- Anderson, T.K., 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accid. Anal. Prev.* 41 (3), 359–364.
- Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accid. Anal. Prev.* 122, 239–254.
- Barahimi, A.H., Eydi, A., Aghaie, A., 2021. Multi-modal urban transit network design considering reliability: multi-objective bi-level optimization. *Reliab. Eng. Syst. Saf.* 216, 107922.
- Bergel-Hayat, R., Debarh, M., Antoniou, C., Yannis, G., 2013. Explaining the road accident risk: Weather effects. *Accid. Anal. Prev.* 60, 456–465.
- Bonera, M., Barabino, B., Yannis, G., Maternini, G., 2024. Network-wide road crash risk screening: a new framework. *Accid. Anal. Prev.* 199, 107502.
- 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.
- Chang, L.-Y., Chen, W.-C., 2005. Data mining of tree-based models to analyze freeway accident frequency. *J. Saf. Res.* 36 (4), 365–375.
- Chen, C., Fan, X., Zheng, C., Xiao, L., Cheng, M., Wang, C., 2018. Sdcae: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data. In: 2018 Sixth International Conference on Advanced Cloud and Big Data. CBD, IEEE, pp. 328–333.
- Chen, Q., Song, X., Yamada, H., Shibasaki, R., 2016. Learning deep representation from big and heterogeneous data for traffic accident inference. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30, No. 1.
- Cheng, T., Zhang, Y., Haworth, J., 2022. Network SpaceTime AI: Concepts, methods and applications. *J. Geod. Geoinf. Sci.* 5 (3), 78.
- Chong, M.M., Abraham, A., Paprzycki, M., 2004. Traffic accident analysis using decision trees and neural networks. *arXiv preprint cs/0405050*.
- Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- Cui, P., Yang, X., Abdel-Aty, M., Zhang, J., Yan, X., 2024. Advancing urban traffic accident forecasting through sparse spatio-temporal dynamic learning. *Accid. Anal. Prev.* 200, 107564.
- Dai, L., Derudder, B., Liu, X., 2018. The evolving structure of the Southeast Asian air transport network through the lens of complex networks, 1979–2012. *J. Transp. Geogr.* 68, 67–77.
- Ding, H., Lu, Y., Sze, N., Chen, T., Guo, Y., Lin, Q., 2022. A deep generative approach for crash frequency model with heterogeneous imbalanced data. *Anal. Methods Accid. Res.* 34, 100212.
- Ding, C., Sun, S., Zhao, J., 2023. MST-GAT: A multimodal spatial-temporal graph attention network for time series anomaly detection. *Inf. Fusion* 89, 527–536.
- Dinu, R., Veeraragavan, A., 2011. Random parameter models for accident prediction on two-lane undivided highways in India. *J. Saf. Res.* 42 (1), 39–42.
- Dunn, P.K., 2004. Occurrence and quantity of precipitation can be modelled simultaneously. *Int. J. Climatol.: J. R. Meteorol. Soc.* 24 (10), 1231–1239.
- El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. *Accid. Anal. Prev.* 41 (5), 1118–1123.
- Greibe, P., 2003. Accident prediction models for urban roads. *Accid. Anal. Prev.* 35 (2), 273–285.
- Gu, Y., Lu, W., Qin, L., Li, M., Shao, Z., 2019. Short-term prediction of lane-level traffic speeds: A fusion deep learning model. *Transp. Res. C* 106, 1–16.
- Halder, A., Mohammed, S., Chen, K., Dey, D., 2019. Spatial risk estimation in tweedie compound Poisson double generalized linear models. *arXiv preprint arXiv:1912.12356*.
- Hamilton, W.L., 2020. Graph representation learning. *Synth. Lect. Artif. Intell. Mach. Learn.* 14 (3), 1–159.
- Hossain, M., Muromachi, Y., 2012. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accid. Anal. Prev.* 45, 373–381.
- Huang, Y., Bi, H., Li, Z., Mao, T., Wang, Z., 2019a. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6272–6281.
- Huang, C., Zhang, C., Dai, P., Bo, L., 2019b. Deep dynamic fusion network for traffic accident forecasting. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. pp. 2673–2681.
- Jiang, X., Zhuang, D., Zhang, X., Chen, H., Luo, J., Gao, X., 2023. Uncertainty quantification via spatial-temporal tweedie model for zero-inflated and long-tail travel demand prediction. *arXiv preprint arXiv:2306.09882*.
- Jin, J., Huang, H., Yuan, C., Li, Y., Zou, G., Xue, H., 2023. Real-time crash risk prediction in freeway tunnels considering features interaction and unobserved heterogeneity: a two-stage deep learning modeling framework. *Anal. Methods Accid. Res.* 40, 100306.
- Jin, J., Liu, P., Huang, H., Dong, Y., 2024. Analyzing urban traffic crash patterns through spatio-temporal data: A city-level study using a sparse non-negative matrix factorization model with spatial constraints approach. *Appl. Geogr.* 172, 103402.
- Jørgensen, B., 1987. Exponential dispersion models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 49 (2), 127–145.
- Khosravi, A., Nahavandi, S., Creighton, D., Atiya, A.F., 2010. Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Trans. Neural Netw.* 22 (3), 337–346.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kipf, T.N., Welling, M., 2016a. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Kipf, T.N., Welling, M., 2016b. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*.
- Kurz, C.F., 2017. Tweedie distributions for fitting semicontinuous health care utilization cost data. *BMC Med. Res. Methodol.* 17, 1–8.
- Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted Poisson regression for county-level crash modeling in California. *Saf. Sci.* 58, 89–97.
- Li, Y., Yu, R., Shahabi, C., Liu, Y., 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.
- Li, R., Zhong, T., Jiang, X., Trajcevski, G., Wu, J., Zhou, F., 2022. Mining spatio-temporal relations via self-paced graph contrastive learning. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '22, pp. 936–944.
- 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. C* 55, 444–459.
- Liu, J., Guan, W., 2004. A summary of traffic flow forecasting methods. *J. Highw. Transp. Res. Dev.* 21 (3), 82–85.
- Liu, J., Li, T., Xie, P., Du, S., Teng, F., Yang, X., 2020. Urban big data fusion based on deep learning: An overview. *Inf. Fusion* 53, 123–133.
- Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. A* 44 (5), 291–305.
- Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accid. Anal. Prev.* 37 (1), 35–46.
- Lv, Y., Tang, S., Zhao, H., 2009. Real-time highway traffic accident prediction based on the k-nearest neighbor method. In: 2009 International Conference on Measuring Technology and Mechatronics Automation, Vol. 3. IEEE, pp. 547–550.
- Ma, C., Zhao, Y., Dai, G., Xu, X., Wong, S.-C., 2022. A novel STFS-CNN-GRU hybrid model for short-term traffic speed prediction. *IEEE Trans. Intell. Transp. Syst.*
- Mahmoud, N., Abdel-Aty, M., Cai, Q., Yuan, J., 2021. Predicting cycle-level traffic movements at signalized intersections using machine learning models. *Transp. Res. C* 124, 102930.
- Mallick, H., Chatterjee, S., Chowdhury, S., Chatterjee, S., Rahnavard, A., Hicks, S.C., 2022. Differential expression of single-cell RNA-seq data using tweedie models. *Stat. Med.* 41 (18), 3492–3510.
- Mannering, F., 2018. Temporal instability and the analysis of highway accident data. *Anal. Methods Accid. Res.* 17, 1–13.
- Martin, J.E., Rivas, T., Matías, J., Taboada, J., Argüelles, A., 2009. A Bayesian network analysis of workplace accidents caused by falls from a height. *Saf. Sci.* 47 (2), 206–214.
- 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: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 33–42.
- Mountain, L., Fawaz, B., Jarrett, D., 1996. Accident prediction models for roads with minor junctions. *Accid. Anal. Prev.* 28 (6), 695–707.
- Mujalli, R.O., De Oña, J., 2011. A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks. *J. Saf. Res.* 42 (5), 317–326.
- Perozzi, B., Al-Rfou, R., Skiena, S., 2014. Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 701–710.
- Qian, W., Zhang, D., Zhao, Y., Zheng, K., Yu, J.J., 2022. Uncertainty quantification for traffic forecasting: A unified approach. *arXiv preprint arXiv:2208.05875*.
- Ren, H., Song, Y., Wang, J., Hu, Y., Lei, J., 2018. A deep learning approach to the citywide traffic accident risk prediction. In: 2018 21st International Conference on Intelligent Transportation Systems. ITSC, IEEE, pp. 3346–3351.
- Reza, S., Ferreira, M.C., Machado, J., Tavares, J.M.R., 2022. A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Syst. Appl.* 202, 117275.
- Rodrigue, J.-P., 2020. The Geography of Transport Systems. Routledge.
- Saha, D., Alluri, P., Dumbaugh, E., Gan, A., 2020. Application of the Poisson-tweedie distribution in analyzing crash frequency data. *Accid. Anal. Prev.* 137, 105456.
- Sayed, T., Abdelwahab, W., 1998. Comparison of fuzzy and neural classifiers for road accidents analysis. *J. Comput. Civ. Eng.* 12 (1), 42–47.
- Shi, P., 2016. Insurance ratemaking using a copula-based multivariate tweedie model. *Scand. Actuar. J.* 2016 (3), 198–215.

- Shirazi, M., Lord, D., 2019. Characteristics-based heuristics to select a logical distribution between the Poisson-gamma and the Poisson-lognormal for crash data modelling. *Transp. A: Transp. Sci.* 15 (2), 1791–1803.
- Smyth, G.K., Jørgensen, B., 2002. Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *ASTIN Bull.: J. IAA* 32 (1), 143–157.
- Soltani, A., Qadikolaie, M.R., 2024. Space-time analysis of accident frequency and the role of built environment in mitigation. *Transp. Policy*.
- Tang, S., Gao, H., 2005. Traffic-incident detection-algorithm based on nonparametric regression. *IEEE Trans. Intell. Transp. Syst.* 6 (1), 38–42.
- Tang, Y., Qu, A., Chow, A.H., Lam, W.H., Wong, S.C., Ma, W., 2022. Domain adversarial spatial-temporal network: A transferable framework for short-term traffic forecasting across cities. In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. pp. 1905–1915.
- Tang, J., Zeng, J., 2022. Spatiotemporal gated graph attention network for urban traffic flow prediction based on license plate recognition data. *Comput.-Aided Civ. Infrastruct. Eng.* 37 (1), 3–23.
- Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46 (sup1), 234–240.
- Trirat, P., Yoon, S., Lee, J.-G., 2023. Multi-view graph convolutional networks for traffic accident risk prediction. *IEEE Trans. Intell. Transp. Syst.*
- Tu, W., Zhu, T., Xia, J., Zhou, Y., Lai, Y., Jiang, J., Li, Q., 2020. Portraying the spatial dynamics of urban vibrancy using multisource urban big data. *Comput. Environ. Urban Syst.* 80, 101428.
- Tweedie, M.C., et al., 1984. An index which distinguishes between some important exponential families. In: *Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee International Conference*, Vol. 579. pp. 579–604.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Wang, S., Cao, J., 2021. AI and deep learning for urban computing. *Urban Inform.* 815–844.
- Wang, S., Chen, A., Wang, P., Zhuge, C., 2023a. Predicting electric vehicle charging demand using a heterogeneous spatio-temporal graph convolutional network. *Transp. Res. C* 153, 104205.
- Wang, W., Jiang, X., Xia, S., Cao, Q., 2010. Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation. *Saf. Sci.* 48 (10), 1248–1262.
- Wang, Y., Jing, C., Xu, S., Guo, T., 2022. Attention based spatiotemporal graph attention networks for traffic flow forecasting. *Inform. Sci.* 607, 869–883.
- Wang, B., Lin, Y., Guo, S., Wan, H., 2021. GSNet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 5. pp. 4402–4409.
- Wang, C., Quddus, M.A., Ison, S.G., 2013a. The effect of traffic and road characteristics on road safety: A review and future research direction. *Saf. Sci.* 57, 264–275.
- Wang, Q., Wang, S., Zhuang, D., Koutsopoulos, H., Zhao, J., 2023b. Uncertainty quantification of spatiotemporal travel demand with probabilistic graph neural networks. *arXiv preprint arXiv:2303.04040*.
- Wang, S., Zhang, J., Li, J., Miao, H., Cao, J., 2021b. Traffic accident risk prediction via multi-view multi-task spatio-temporal networks. *IEEE Trans. Knowl. Data Eng.* 1.
- Wang, K., Zhao, S., Jackson, E., 2019. Functional forms of the negative binomial models in safety performance functions for rural two-lane intersections. *Accid. Anal. Prev.* 124, 193–201.
- World Health Organization, 2023. *Global Status Report on Road Safety 2023*. World Health Organization.
- Wu, M., Jia, H., Luo, D., Luo, H., Zhao, F., Li, G., 2022. A multi-attention dynamic graph convolution network with cost-sensitive learning approach to road-level and minute-level traffic accident prediction. *IET Intell. Transp. Syst.* *itr2.12254*.
- Wu, Y., Ye, Y., Zeb, A., Yu, J.J., Wang, Z., 2023. Adaptive modeling of uncertainties for traffic forecasting. *arXiv preprint arXiv:2303.09273*.
- Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accid. Anal. Prev.* 75, 16–25.
- Xue, J., Jiang, N., Liang, S., Pang, Q., Yabe, T., Ukusuri, S.V., Ma, J., 2022. Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nat. Mach. Intell.* 4 (3), 246–257.
- 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., Lee, Y., Sohn, K., 2020. Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). *Transp. Res. C* 114, 189–204.
- Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence. IJCAI*.
- 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*. pp. 984–992.
- Zhang, Y., Cheng, T., 2020. Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. *Comput. Environ. Urban Syst.* 79, 101403.
- Zhang, Y., Cheng, T., Ren, Y., 2019a. A graph deep learning method for short-term traffic forecasting on large road networks. *Comput.-Aided Civ. Infrastruct. Eng.* 34 (10), 877–896.
- Zhang, C., James, J., Liu, Y., 2019b. Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. *IEEE Access* 7, 166246–166256.
- Zhang, C., Yan, X., Ma, L., An, M., 2014. Crash prediction and risk evaluation based on traffic analysis zones. *Math. Probl. Eng.* 2014.
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., Li, H., 2019. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 21 (9), 3848–3858.
- Zhao, S., Zhao, D., Liu, R., Xia, Z., Cheng, B., Chen, J., 2023. GMAT-DU: Traffic anomaly prediction with fine spatiotemporal granularity in sparse data. *IEEE Trans. Intell. Transp. Syst.*
- Zhou, X., Lu, P., Zheng, Z., Tolliver, D., Keramati, A., 2020a. Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliab. Eng. Syst. Saf.* 200, 106931.
- Zhou, H., Qian, W., Yang, Y., 2022a. Tweedie gradient boosting for extremely unbalanced zero-inflated data. *Comm. Statist. Simulation Comput.* 51 (9), 5507–5529.
- Zhou, Z., Wang, Y., Xie, X., Chen, L., Liu, H., 2020b. RiskOracle: a minute-level citywide traffic accident forecasting framework. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 01. pp. 1258–1265.
- Zhou, Z., Wang, Y., Xie, X., Chen, L., Zhu, C., 2022b. Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective. *IEEE Trans. Knowl. Data Eng.* 34 (8), 3786–3799.
- Zhou, Z., Wang, Y., Xie, X., Qiao, L., Li, Y., 2021. STUaNet: Understanding uncertainty in spatiotemporal collective human mobility. In: *Proceedings of the Web Conference 2021*. pp. 1868–1879.
- Zhu, L., Li, T., Du, S., 2019. Ta-stan: A deep spatial-temporal attention learning framework for regional traffic accident risk prediction. In: *2019 International Joint Conference on Neural Networks. IJCNN, IEEE*, pp. 1–8.
- Zhu, Y., Ye, Y., Liu, Y., James, J., 2022. Cross-area travel time uncertainty estimation from trajectory data: a federated learning approach. *IEEE Trans. Intell. Transp. Syst.* 23 (12), 24966–24978.
- Zhuang, D., Wang, S., Koutsopoulos, H., Zhao, J., 2022. Uncertainty quantification of sparse travel demand prediction with spatial-temporal graph neural networks. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. pp. 4639–4647.