Uncertainty-Aware Crime Prediction With Spatial Temporal Multivariate Graph Neural Networks

Zepu Wang[1]*, Xiaobo Ma[2], Huajie Yang[3], Weimin Lyu[4], Yang Liu[5], Peng Sun[6]*, Sharath Chandra Guntuku[1]
[1]University of Pennsylvania [2]University of Arizona [3]City University of Macau
[4]Stony Brook University [5]University of Toronto [6]Duke Kunshan University
zepu@seas.upenn.edu, peng.sun568@duke.edu

ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10889685

***Abstract*—Crime prediction (CP) plays a pivotal role in urban analytics, contributing significantly to personal safety and societal stability. Unlike conventional time series forecasting, CP faces unique difficulties due to the inherent sparsity of crime incidents, particularly within small spatial regions and limited time windows. This sparsity, coupled with the non-Gaussian distribution of crime data—characterized by an excess of zero events and over-dispersion—presents a critical challenge for the signal processing community. In this regard, we propose a novel framework, Spatial-Temporal Multivariate Zero-Inflated Negative Binomial Graph Neural Networks (STMGNN-ZINB), which integrates diffusion and convolutional graph networks to capture spatial, temporal, and multivariate dependencies. By leveraging a Zero-Inflated Negative Binomial distribution, the STMGNN-ZINB effectively models the over-dispersed and zero-heavy nature of crime data, significantly improving both prediction accuracy and confidence interval estimation. Experimental results on real-world datasets demonstrate that our STMGNN-ZINB outperforms state-of-the-art CP methods, offering a robust tool for crime early warning and explicable insights into urban crime dynamics.**

***Index Terms*—Time series, spatial-temporal signal, crime prediction, graph neural networks.**

I. Introduction

Accurate crime prediction is essential for optimizing police deployment strategies and infrastructure planning, which play a crucial role in enhancing urban safety and improving the quality of life in cities [1]–[3]. As urban populations grow and cities become more complex, the ability to predict crime with precision has become increasingly valuable [4]. Effective crime prediction enables law enforcement agencies to allocate resources efficiently, proactively respond to potential criminal activities, and reduce overall crime rates. However, the complexity and sparsity of crime data present significant challenges in developing accurate predictive models.

With the advancement of deep learning technologies, researchers have gained new tools to model the intricate patterns found in crime data [5]–[10]. Traditional statistical methods, while valuable, often fail to capture the rich spatial and temporal dynamics present in urban crime. As a result, more sophisticated neural network architectures, such as recurrent neural networks (RNNs) [2], convolutional neural networks (CNNs) [11], and graph neural networks (GNNs) [3], [12], have been explored to model the spatial-temporal correlations and the interrelationships among different crime types. Among these, spatial-temporal GNNs (STGNNs) have shown considerable promise in capturing complex spatial and temporal dependencies in urban data [13]–[16]. These models leverage the graph structure of urban environments, where nodes represent locations and edges represent spatial or temporal interactions, making them well-suited for crime prediction.

Despite these advancements, a major limitation persists in existing models: the implicit assumption that crime data follows a Gaussian distribution [2], [17]–[19]. This assumption oversimplifies the true nature of crime data, particularly the variance structure, and leads to models that struggle to handle the unique characteristics of urban crime. Crime data is typically sparse, with many regions experiencing long periods of no incidents, resulting in numerous zero values in the datasets [20], [21]. Moreover, crime occurrences are not only infrequent but also exhibit over-dispersion, where the variance exceeds the mean. These properties make traditional models inadequate for capturing the underlying distribution.

To address these challenges, the Zero-Inflated Negative Binomial (ZINB) distribution has been introduced as a more appropriate tool for modeling sparse and over-dispersed data. This distribution has been successfully applied in other urban prediction tasks, such as sparse travel demand modeling [17] and traffic risk forecasting [22]. The ZINB distribution captures both the zero-inflation and the over-dispersed nature of crime data, making it a natural choice for crime prediction. Additionally, uncertainty quantification, which is often overlooked in traditional models, becomes crucial when dealing with sparse data, as it allows for more informed decision-making in law enforcement and urban planning [23], [24].

In this paper, inspired by other uncertainty quantification applications and analysis procedures in urban systems [17], [23], [25], we implement a effective framework called Spatial-Temporal Multivariate Zero-Inflated Negative Binomial Graph Neural Networks (STMGNN-ZINB), which combines the strengths of spatial-temporal graph neural networks and the ZINB distribution to address the unique challenges of crime prediction. Our framework not only provides accurate numeric predictions but also quantifies the uncertainty associated with these predictions, which is particularly important in sparse data scenarios. The main contributions of this work are summarized as follows:

*Corresponding authors.

- We leverage the Zero-Inflated Negative Binomial (ZINB) distribution to model crime data, effectively capturing both zero inflation and sparsity.
- We integrate the ZINB distribution with spatial-temporal multivariate GNNs, enabling precise modeling of the sparse, discrete uncertainty present in crime data.
- We conduct extensive experiments on two real-world crime datasets, demonstrating that our proposed STMGNN-ZINB framework outperforms existing models based on alternative distributional assumptions.

II. Preliminaries

A. Problem Definition

Suppose the historical crime data is embedded in a static graph G nodes with C multivariate crime features over T time intervals. For N nodes in graph G, in this paper, we construct the adjacency matrix A of G by scaling the normalized Laplacian matrix [26]. The historical time series features can then be represented as $X \in \mathbb{R}^{N\times T\times C}$.

The objective of this task is to learn a mapping function f that uses the historical crime data X and the graph structure G as inputs to forecast future crime data for Q time intervals. Our goal is not only to predict the expected values of future crime but also to estimate the confidence intervals for these predictions. Consequently, we denote our output as $\hat{X} \in \mathbb{R}^{N\times Q\times C\times Z}$, where Z represents the parameters of the assumed distribution of crime data.

B. Zero-Inflated Negative Binomial Distribution

In our model, we assume that the distribution of crimes follows the Zero-Inflated Negative Binomial (ZINB) distribution [17], [27]. The probability mass function (PMF) of a random variable following the ZINB distribution is given by:

$$P(Y=y) = \begin{cases} \pi + (1-\pi)(1-p)^r & \text{if } y=0, \\ (1-\pi)\binom{y+r-1}{y}p^y(1-p)^r & \text{if } y=1,2,3,\dots, \end{cases} \tag{1}$$

where π represents the probability of an extra zero, indicating zero inflation, which in this context, signifies the likelihood of no occurrences of a specific type of crime. The parameters r and p are the shape parameters of the traditional negative binomial distribution, where r affects the dispersion and p is the probability of success in each experiment.

C. Diffusion Graph Convolution Networks (DGCNs)

To capture spatial correlations within a predefined graph structure, we utilize Diffusion Graph Convolutional Networks (DGCNs). These models employ the concept of diffusion processes on graphs to effectively capture the spread of information across the graph's topology.

At the heart of DGCNs is the diffusion convolution operation, which can be considered a generalization of the traditional convolutional operations adapted for graph data. The underlying principle is to simulate a diffusion process on the graph, allowing information to propagate from a node to its neighbors through multiple steps or layers. This dynamic can be mathematically articulated using the graph Laplacian and its exponentiations to represent various degrees of diffusion.

The diffusion convolution operation in a DGCN is mathematically defined as:

$$H^{(l+1)} = \sigma\left(D^{-1}AH^{(l)}W^{(l)} + B^{(l)}H^{(l)}\right),$$

where $H^{(l)}$ denotes the node features (or hidden states) at layer l; A is the adjacency matrix of the graph, where A_{ij} represents the edge weight between nodes i and j, with $A_{ij}=0$ if no edge exists; D is the diagonal degree matrix, with each diagonal element D_{ii} being the sum of the weights of all edges connected to node i; $W^{(l)}$ is the weight matrix for layer l, which is learned during the training process; $B^{(l)}$ is the bias term for layer l; $\sigma(\cdot)$ is a non-linear activation function, such as the Rectified Linear Unit (ReLU).

This formulation enables the network to effectively balance the influence of immediate neighbors (through $D^{-1}AH^{(l)}$) and the retention of the current node's features (via $B^{(l)}H^{(l)}$), thus integrating both local and global information in the learning process.

D. Multivariate-Temporal Convolutional Networks (MTCNs)

Traditional Temporal Convolutional Neural Networks (TCNs) are a specialized variant of convolutional neural networks designed to process sequence data effectively. The core principle of TCNs is to apply a shared gated 1D convolution across a specified width w_l in the l^{th} layer, allowing the integration of information from w_l adjacent time points. In this study, we adapt traditional TCNs to simultaneously capture multivariate-temporal correlations. Specifically, we aim to incorporate information from all crime types at preceding time points for predicting a specific crime type at the current time point. This integration is facilitated by reshaping the multivariate and temporal dimensions together.

Each TCN layer H_l receives input from the preceding layer H_{l-1} and updates as follows:

$$H^{(l+1)} = f(\Gamma_l * H^{(l)} + b), \tag{2}$$

where Γ_l represents the convolution filter for the layer, $*$ denotes the shared convolution operation, and b represents the bias. If the previous hidden layer follows $H_{l-1} \in \mathbb{R}^{B\times|V|\times(w_{l-1}C)}$, then the convolution filter is $\Gamma_l \in \mathbb{R}^{(w_lC)\times(w_{l-1}C)}$, ensuring that $H_l \in \mathbb{R}^{B\times|V|\times(w_lC)}$. Notably, if there is no padding in each TCN layer, it follows that $w_l < w_{l-1}$.

The primary motivation for capturing multivariate-temporal correlations is to leverage the streamlined architecture of traditional TCNs for rapid training. Moreover, given that the temporal and multivariate dimensions typically exhibit shorter lengths compared to the spatial dimension, they can be treated as a single combined dimension. This approach helps mitigate the risk of extracting spurious relationships from the training data, which can lead to overfitting—a common issue when neural networks attempt to model relationships between two dimensions with extensive lengths [28].

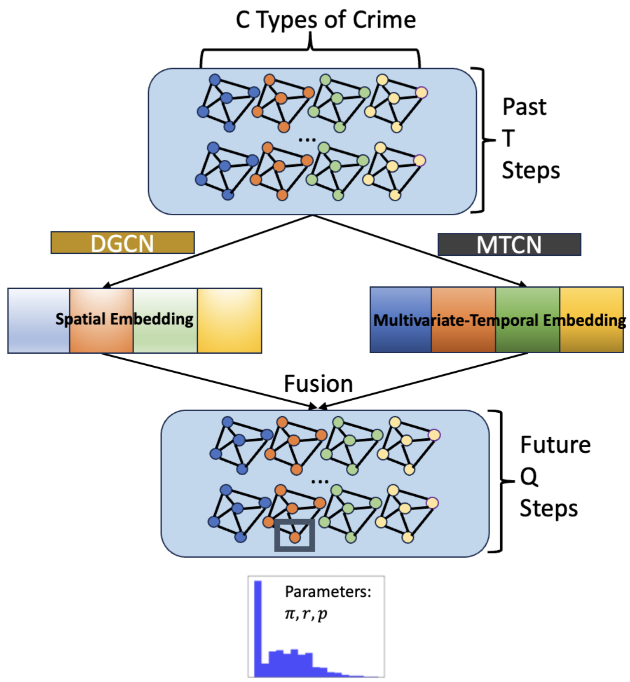


Fig. 1: Overal Framework of the proposed STMGNN-ZINB.

E. General Structure

As illustrated in Figure 1, we utilize DGCNs to capture spatial dependencies, resulting in spatial embeddings π_1, p_1, and r_1. Concurrently, MTCNs are employed to capture multivariate temporal correlations, yielding multivariate temporal embeddings π_2, p_2, and r_2. These embeddings are then fused into combined parameters π, p, and r using the Hadamard product. This method of integration is inspired by recent advances in uncertainty quantification techniques [17].

To address the issue where zeros result in infinite values for the KL-divergence-based variational lower bound [17], we opt to directly utilize the negative likelihood as our loss function. This approach allows for a more accurate fitting of the distribution to the data, circumventing the limitations posed by zeros in the calculation of the KL-divergence.

$$\text{NLL} = -\sum_{i=1}^{N} \begin{cases} \log[\pi_i + (1-\pi_i)\cdot(1-p_i)^{r_i}] & \text{if } y_i = 0, \\ \log[(1-\pi_i)\cdot\frac{\Gamma(r_i+y_i)}{\Gamma(r_i)\Gamma(y_i+1)} p_i^{y_i}(1-p_i)^{r_i}] & \text{if } y_i > 0. \end{cases} \quad (3)$$

III. Experiments

A. Experiment Setup

1) Data: Following the methodology in [29], our experiments utilize two distinct crime datasets from New York City (NYC) and Chicago (CHI). These datasets cover various types of crime incidents, including Robbery and Larceny in NYC and Damage and Assault in CHI, across different locations within each city. For the purpose of our experiments, we divided each city into a spatial grid of 3 km × 3 km, resulting in 256 regions for NYC and 168 regions for CHI. Our prediction targets are set at a daily resolution.

The datasets were split into training and testing sets with a 7:1 ratio along the time dimension. Additionally, we used the last 30 days of the training set as a validation set to fine-tune the model parameters. Table I provides a summary of the dataset statistics.

TABLE I: Statistics of Experimented Urban Crime Datasets

Dataset	Time Span	Category	Count	Zero Rate
NYC-Crimes	Jan. 2014 to Dec. 2015	Burglary	31799	89
		Larceny	85899	81
		Robbery	33453	89
		Assault	40429	88
CHI-Crimes	Jan. 2016 to Dec. 2017	Theft	124630	68
		Battery	99389	71
		Assault	37972	81
		Damage	59886	75

2) Evaluation Metrics: We evaluate the performance of different models from three perspectives: 1. Point Estimation: We use Mean Absolute Error (MAE) to measure the accuracy of the mean value of the predicted distributions. A lower MAE indicates better accuracy. 2. Uncertainty Quantification: This includes Mean Prediction Interval Width (MPIW) and Prediction Interval Coverage Probability (PICP) within the 10%-90% confidence interval. MPIW calculates the average width of the confidence intervals, reflecting the extent of uncertainty in the predictions. PICP quantifies the percentage of actual data points that fall within these intervals, aiming for coverage as close to 90% as possible. Additionally, KL-Divergence is used to assess the similarity between the predicted and actual data distributions, with lower values indicating better model performance. 3. Discrete Metrics: We round our results to their closest integer to measure the true-zero rate and the F1-score. The true-zero rate evaluates the model's ability to accurately represent data sparsity, whereas the F1-score assesses the accuracy of discrete predictions. Higher values of the true-zero rate and F1 score signify superior performance.

3) Baseline Methods: To explore the advantages of STMGNN-ZINB, we compare the STMGNN-ZINB results against three other models: (1) Historical Value (HA) serves as the statistical baseline. It uses historical input directly to represent prediction results. (2) Spatial-Temporal Graph Convolutional Networks (STGCN) [1] is the state-of-the-art deep learning model for spatial-temporal prediction but it only produces point estimates; (3) Models with probabilistic assumptions to replace ZINB: Negative Binomial (STMGNN-NB), Gaussian (STMGNN-G), and Truncated Normal (STMGNN-TN).

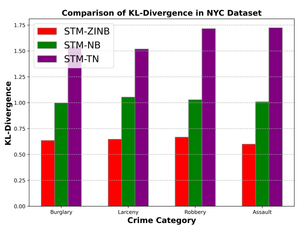
B. Experimental Results and Analysis

As shown in Table II, STMGNN-ZINB outperforms all baseline methods across the majority of evaluation metrics. Specifically, in point estimation, STMGNN-ZINB achieves the lowest Mean Absolute Error (MAE), indicating that its mean value accurately reflects the trends in real crime data. For uncertainty metrics, the lower Mean Prediction Interval Width (MPIW) of STMGNN-ZINB suggests that our model generates predictions with less uncertainty and greater precision. STMGNN-ZINB surpasses other methods on metrics such as the true zero rate and F1 Score in most cases, signifying STMGNN-ZINB's efficiency in capturing zero values.

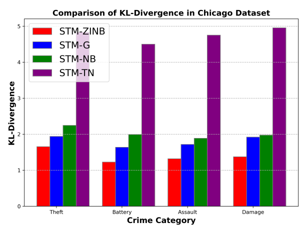
[1] https://github.com/FelixOpolka/STGCN-PyTorch

TABLE II: Experimental Results in NYC-Crimes and CHI-Crimes. The best and second-best scores are bold and underlined.

Datasets	Methods	MAE	KL Divergence	PICP	MPIW	F1 Score	True Zero Rate
		Point Estimation	Uncertainty Metrics			Discrete Metrics	
NYC-Crimes	STMGNN-ZINB	**0.2128**	0.6147	**0.9505%**	**1.028**	0.6556	96.57%
	STMGNN-G	0.2536	**0.1148**	0.9557%	1.30	¡0.1	**100.0%**
	STNGNN-NB	0.2300	1.0218	0.9690%	1.30	**0.6644**	93.01%
	STMGNN-TN	0.3206	1.0636	0.9682%	1.5148	0.5638	90.92%
	HV	0.3014	1.2783	/	/	0.5234	92.53%
	STGCN	0.2527	2.0259	/	/	0.6537	87.10%
CHI-Crimes	STMGNN-ZINB	**0.5242**	**1.3945**	**0.9141%**	**2.134**	0.7693	**90.73%**
	STMGNN-G	0.5587	1.8047	0.8938%	2.532	0.7259	86.50%
	STMGNN-NB	0.5575	2.0259	0.9444%	2.4164	0.7769	87.10%
	STMGNN-TN	0.7619	2.7606	0.8617%	2.906	0.6329	61.13%
	HV	0.6560	2.2312	/	/	0.6849	89.09%
	STGCN	0.5535	1.9512	/	/	**0.7858**	86.59%



(a) KL-Divergence in NYC.



(b) KL-Divergence in Chicago.

Fig. 2: KL-Divergence in (a) NYC and (b) Chicago.

During our experiments, we observed that STMGNN-G performed exceptionally well in terms of KL Divergence and True Zero Rate in the New York City dataset. However, its F1 Score was notably low. Further investigation revealed that STMGNN-G predominantly outputs near-zero mean values in this dataset. Given that the New York City dataset has a significantly higher zero rate compared to the Chicago dataset, the low F1 Score suggests that STMGNN-G struggles to effectively predict crime occurrences. Hence, we conclude that it is crucial to evaluate a model using diverse and multiple evaluation metrics. The performance of STMGNN-ZINB is more comprehensive than that of STMGNN-G.

When excluding STMGNN-G from the New York City dataset to evaluate KL divergence, as illustrated in Figures 2a and 2b, we can conclude that STMGNN-ZINB demonstrates the best performance in approximating the actual distribution.

C. Interpretation of Parameter π

The sparsity parameter π in the Zero-Inflated Negative Binomial (ZINB) distribution quantifies the likelihood that a specific zone is devoid of any crime activities. Our STMGNN-ZINB model outputs π for various types of crime activities across different spatial regions for successive future time steps.

Figure 3 displays a heatmap of the overall crime activities (parameter π) for New York City and Chicago. This visualization offers valuable insights for government agencies and public security entities. Notably, it reveals the presence of spatial locality, where changes in crime activities gradually vary across the spatial dimension. It is uncommon to find isolated small regions with low crime rates surrounded by areas with high crime rates. Furthermore, consistent with mainstream criminological research, our findings indicate a higher concentration of crime activities in the downtown areas of both cities compared to the marginal regions (rural areas) [30], [31]. This pattern underscores the heightened crime rates typically observed in urban centers as opposed to rural settings. Thus, the sparsity parameter π significantly enhances the interpretability of the STMGNN-ZINB model, offering a more nuanced understanding compared to models based on simpler distribution assumptions.

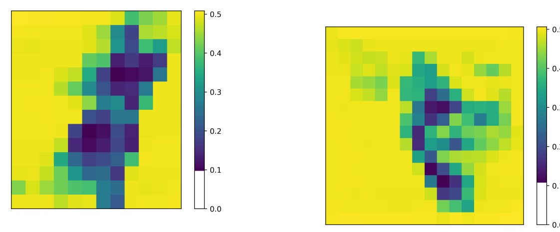


Fig. 3: Heatmap of π in the entire urban space of NYC (left) and Chicago (right). Results are averaged over the multivariate dimension and temporal dimension.

IV. Conclusion

In this work, we introduce a novel combination of Spatial-Temporal Multivariate Graph Neural Networks and Zero-Inflated Negative Binomial distribution to quantify uncertainty in the urban crime prediction problem. We validated our model's performance through extensive experiments across five representative scenarios, focusing particularly on point estimation and uncertainty measurement. Notably, there is a substantial body of existing research focused on designing spatial-temporal graph neural networks specifically for crime prediction purposes [29], [32], [33]. To further advance this work, it is essential to consider the unique characteristics of crime data in designing more accurate neural network architectures. Additionally, exploring more complex distributional assumptions may help capture the intricate crime patterns.

REFERENCES

[1] C. Wang, Z. Lin, X. Yang, J. Sun, M. Yue, and C. Shahabi, "Hagen: Homophily-aware graph convolutional recurrent network for crime forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, 2022, pp. 4193–4200.

[2] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla, "Deepcrime: Attentive hierarchical recurrent networks for crime prediction," in *Proceedings of the 27th ACM international conference on information and knowledge management*, 2018, pp. 1423–1432.

[3] C. Huang, C. Zhang, J. Zhao, X. Wu, D. Yin, and N. Chawla, "Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting," in *The world wide web conference*, 2019, pp. 717–728.

[4] Y. Liu, D. Yang, Y. Wang, J. Liu, J. Liu, A. Boukerche, P. Sun, and L. Song, "Generalized video anomaly event detection: Systematic taxonomy and comparison of deep models," *ACM Computing Surveys*, vol. 56, no. 7, pp. 1–38, 2024.

[5] F. Lin, S. Crawford, K. Guillot, Y. Zhang, Y. Chen, X. Yuan, L. Chen, S. Williams, R. Minvielle, X. Xiao *et al.*, "Mmst-vit: Climate change-aware crop yield prediction via multi-modal spatial-temporal vision transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 5774–5784.

[6] Z. Ke, Z. Li, Z. Cao, and P. Liu, "Enhancing transferability of deep reinforcement learning-based variable speed limit control using transfer learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4684–4695, 2020.

[7] Y. Liu, X. Dong, W. Lyu, R. N. Rosenthal, R. Wong, T. Ma, J. Kong, and F. Wang, "Enhancing clinical predictive modeling through model complexity-driven class proportion tuning for class imbalanced data: An empirical study on opioid overdose prediction," *AMIA Summits on Translational Science Proceedings*, vol. 2024, p. 334, 2024.

[8] R. Bautista-Montesano, R. Galluzzi, K. Ruan, Y. Fu, and X. Di, "Autonomous navigation at unsignalized intersections: A coupled reinforcement learning and model predictive control approach," *Transportation research part C: emerging technologies*, vol. 139, p. 103662, 2022.

[9] Z. Wang, P. Sun, and A. Boukerche, "A novel time efficient machine learning-based traffic flow prediction method for large scale road network," in *Proceedings of the 2022 IEEE International Conference on Communications*. IEEE, 2022, pp. 3532–3537.

[10] D. Wei, Y. Liu, X. Zhu, J. Liu, and X. Zeng, "Msaf: Multimodal supervise-attention enhanced fusion for video anomaly detection," *IEEE Signal Processing Letters*, vol. 29, pp. 2178–2182, 2022.

[11] W. Safat, S. Asghar, and S. A. Gillani, "Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques," *IEEE access*, vol. 9, pp. 70 080–70 094, 2021.

[12] J. Sun, M. Yue, Z. Lin, X. Yang, L. Nocera, G. Kahn, and C. Shahabi, "Crimeforecaster: Crime prediction by exploiting the geographical neighborhoods' spatiotemporal dependencies," in *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part V*. Springer, 2021, pp. 52–67.

[13] Z. Mo, H. Xiang, and X. Di, "Cross-and context-aware attention based spatial-temporal graph convolutional networks for human mobility prediction," *ACM Transactions on Spatial Algorithms and Systems*, 2024.

[14] K. Ruan and X. Di, "Infostgcan: An information-maximizing spatial-temporal graph convolutional attention network for heterogeneous human trajectory prediction," *Computers*, vol. 13, no. 6, p. 151, 2024.

[15] Z. Wang, P. Sun, Y. Hu, and A. Boukerche, "A novel mixed method of machine learning based models in vehicular traffic flow prediction," in *Proceedings of the 25th International ACM Conference on Modeling Analysis And Simulation of Wireless And Mobile Systems*, 2022, pp. 95–101.

[16] Z. Wang, Y. Nie, P. Sun, N. H. Nguyen, J. Mulvey, and H. V. Poor, "St-mlp: A cascaded spatio-temporal linear framework with channel-independence strategy for traffic forecasting," *arXiv preprint arXiv:2308.07496*, 2023.

[17] D. Zhuang, S. Wang, H. Koutsopoulos, and J. Zhao, "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*, 2022, pp. 4639–4647.

[18] T. Almanie, R. Mirza, and E. Lor, "Crime prediction based on crime types and using spatial and temporal criminal hotspots," *arXiv preprint arXiv:1508.02050*, 2015.

[19] W. He and Z. Jiang, "Uncertainty quantification of deep learning for spatiotemporal data: Challenges and opportunities," *arXiv preprint arXiv:2311.02485*, 2023.

[20] X. Wu, C. Huang, C. Zhang, and N. V. Chawla, "Hierarchically structured transformer networks for fine-grained spatial event forecasting," in *Proceedings of the web conference 2020*, 2020, pp. 2320–2330.

[21] L. Zhao, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Multi-resolution spatial event forecasting in social media," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 2016, pp. 689–698.

[22] X. Gao, X. Jiang, D. Zhuang, H. Chen, S. Wang, and J. Haworth, "Spatiotemporal graph neural networks with uncertainty quantification for traffic incident risk prediction," *arXiv preprint arXiv:2309.05072*, 2023.

[23] X. Jiang, D. Zhuang, X. Zhang, H. Chen, J. Luo, and X. Gao, "Uncertainty quantification via spatial-temporal tweedie model for zero-inflated and long-tail travel demand prediction," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 3983–3987.

[24] D. Zhuang, Y. Bu, G. Wang, S. Wang, and J. Zhao, "Sauc: Sparsity-aware uncertainty calibration for spatiotemporal prediction with graph neural networks," in *Temporal Graph Learning Workshop@ NeurIPS 2023*, 2023.

[25] X. Gao, X. Jiang, J. Haworth, D. Zhuang, S. Wang, H. Chen, and S. Law, "Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction," *Accident Analysis & Prevention*, vol. 208, p. 107801, 2024.

[26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint: 1609.02907*, 2023.

[27] M. Jiang and H. Zhang, "Sparse estimation in high-dimensional zero-inflated poisson regression model," in *Journal of Physics: Conference Series*, vol. 1053, no. 1. IOP Publishing, 2018, p. 012128.

[28] Z. Wang, D. Zhuang, Y. Li, J. Zhao, P. Sun, S. Wang, and Y. Hu, "St-gin: An uncertainty quantification approach in traffic data imputation with spatio-temporal graph attention and bidirectional recurrent united neural networks," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 1454–1459.

[29] Z. Li, C. Huang, L. Xia, Y. Xu, and J. Pei, "Spatial-temporal hypergraph self-supervised learning for crime prediction," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 2022, pp. 2984–2996.

[30] M. Lan, L. Liu, and J. E. Eck, "A spatial analytical approach to assess the impact of a casino on crime: An example of jack casino in downtown cincinnati," *Cities*, vol. 111, p. 103003, 2021.

[31] R. Rosenfeld and E. Lopez, "Pandemic, social unrest, and crime in us cities," *Federal Sentencing Reporter*, vol. 33, no. 1/2, pp. 72–82, 2020.

[32] L. Xia, C. Huang, Y. Xu, P. Dai, L. Bo, X. Zhang, and T. Chen, "Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning," *arXiv preprint arXiv:2201.02435*, 2022.

[33] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 1720–1730.