

Article

Multi-Scale Geo-Temporal Crime Embedding (MSG-TCE): A Hierarchical Spatiotemporal Framework for Crime Prediction with Hyperbolic Spatial Pooling and Periodic Transformers

Rosny Jean ^{1,*} and Stabak Roy ² 

¹ School of the Environment, Florida Agricultural and Mechanical University, Tallahassee, FL 32307, USA

² Department of Geography and Disaster Management, Techno India University, Agartala 799004, Tripura, India; stabak.roy@gmail.com

* Correspondence: rosny.jean@fam.u.edu

Abstract

Crime prediction in urban environments is a complex and pressing challenge driven by the intricate interplay of spatiotemporal dependencies, hierarchical geographic patterns, and socio-environmental determinants. We propose a multi-scale geo-temporal crime embedding (MSG-TCE) framework, which hierarchically models these dynamics via three novel components: a hierarchical residual temporal encoder (HRTE), a periodic transformer Encoder (PTE), and a hyperbolic spatial pooler (HSP). The HRTE captures multi-scale temporal trends by combining dilated convolutions with residual connections, while the PTE explicitly encodes periodic crime patterns using self-attention conditioned on cyclical positional encodings. The HSP maps spatial crime hotspots into hyperbolic space to better represent their inherent hierarchical structure, spanning city–district–neighbourhood–street–segment scales, and aggregates neighbourhood information via graph convolutions. These components are fused through a gated cross-attention mechanism, yielding a unified embedding for crime prediction. Experiments on real-world datasets from Chicago, Los Angeles, and New York City demonstrate that MSG-TCE achieves consistent improvements over five competitive baselines across RMSE, Precision@20, and DTW metrics, with statistically significant gains at longer prediction horizons. Ablation studies confirm the contribution of each component. Spatial visualisation maps, robustness analyses, and an exploratory covariate-augmented variant further substantiate the empirical validity of the framework. This paper also discusses limitations, including data reporting biases, the need for full covariate integration, and ethical considerations, pertaining to algorithmic fairness in crime prediction.



Academic Editors: Wolfgang Kainz and Jamal Jokar Arsanjani

Received: 24 April 2026

Revised: 23 June 2026

Accepted: 26 June 2026

Published: 2 July 2026

Copyright: © 2026 by the authors.

Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: crime prediction; spatiotemporal modelling; GeoAI; hyperbolic embeddings; transformer networks; graph neural networks; urban analytics; crime hotspot detection

1. Introduction

Urban crime prediction has become one of the most consequential application areas for spatiotemporal machine learning [1], with direct implications for public safety, equitable resource allocation, and evidence-based urban governance. Despite substantial progress in geospatial analytics and deep learning, predicting where and when crime will occur with sufficient accuracy and interpretability remains an open challenge. A central difficulty lies in the multi-scale, hierarchical nature of crime distributions across urban space and time, a feature that most existing models do not represent adequately.

To understand why hierarchical spatial representation matters for crime prediction, consider the nested organisation of crime across urban environments. At the broadest scale, crime concentrates in specific neighbourhoods due to the co-presence of socioeconomic deprivation, residential instability, and reduced informal social control [2]. Within each neighbourhood, crime is further stratified at the street-segment level: a well-established empirical finding in criminology is that a small proportion of street segments, typically 4–6%, account for the majority of crime incidents in any given city [3]. This produces a nested micro-hierarchy within the broader neighbourhood pattern. Furthermore, crime patterns differ systematically between urban cores and peripheral areas: central business districts and major transit hubs concentrate certain crime types (robbery, pickpocketing, assault), while suburban and peripheral zones exhibit distinct crime profiles (residential burglary, domestic violence), reflecting differences in land-use composition, routine activity flows, and population density [4]. Research on Bratislava, for example, has demonstrated that peripheral suburban areas are statistically safer than the urban core, highlighting the centre–periphery asymmetry in crime hierarchies that any spatially sensitive model must account for [5].

This nested, tree-like organisation from city-level patterns through district-level zones, neighbourhood clusters, and down to street-segment micro-hotspots is also shaped by social networks (co-offending ties and gang territories), land-use structure (commercial strips, mixed-use corridors, and residential enclaves), and transport infrastructure (transit nodes, arterial roads, and pedestrian corridors acting as conduits for crime displacement) [4]. The urban environment itself, through principles of crime prevention through environmental design (CPTED), directly moderates crime levels by influencing visibility, access control, territorial definition, and the activity patterns of residents, offenders, and guardians [6,7]. These hierarchical structures are precisely the kind of relationships that Euclidean spatial models fail to represent faithfully, motivating the use of hyperbolic geometry, a mathematical space in which hierarchical tree-like structures can be embedded with minimal distortion [8].

The temporal dimension of crime data presents complementary challenges. Crime patterns exhibit multi-scale periodicity, ranging from within-day fluctuations driven by routine activity rhythms (e.g., peaks during commuting hours and weekend nights) to weekly cycles (weekday versus weekend patterns) and seasonal variations. Previous work has employed Fourier transforms [9] or wavelet analysis [10] to capture these periodicities, but these methods often require manual frequency specification and do not adapt to changes in pattern structure over time. Moreover, long-term trends influenced by socioeconomic change operate at an entirely different temporal scale. No existing model captures all of these temporal scales jointly within a single end-to-end trainable framework.

We propose a multi-scale geo-temporal crime embedding (MSG-TCE) framework to address these challenges. MSG-TCE integrates three key innovations: (1) a hierarchical residual temporal encoder (HRTE) that learns crime patterns at multiple temporal scales through dilated convolutions with residual connections [11]; (2) a periodic transformer encoder (PTE) that uses self-attention conditioned on cyclical positional encodings to model periodic dependencies explicitly; and (3) a hyperbolic spatial pooler (HSP) that maps geographic crime data into hyperbolic space and aggregates neighbourhood information via graph convolutions, thereby preserving the hierarchical structure of urban crime distributions. These components are fused through a gated cross-attention mechanism that learns the optimal combination of temporal and spatial representations for each spatial unit.

The primary contributions of this work are fourfold. First, we introduce a unified framework that jointly models multi-scale temporal dynamics and hierarchical spatial patterns of criminal activity, overcoming the limitations of existing approaches that treat these

dimensions separately. Second, we develop novel architectural components specifically designed for the crime prediction domain, with explicit motivation from environmental criminology and urban geography. Third, we demonstrate through experiments on three real-world datasets, Chicago, Los Angeles, and New York City, that MSG-TCE achieves consistent and statistically significant improvements over five state-of-the-art baselines. Fourth, we present spatial visualisations, ablation analyses, robustness checks, and an exploratory covariate-augmented experiment that together provide a thorough empirical foundation for the framework's claims, while also openly discussing its limitations and the ethical responsibilities associated with algorithmic crime prediction.

A clarification of the scientific scope of this contribution is warranted. The empirical regularities that motivate MSG-TCE—temporal periodicity, multi-scale spatial concentration, and spatiotemporal interaction—are long established in environmental criminology through routine activity theory and crime pattern theory, and we do not claim to rediscover them. The contribution of this work is therefore methodological and analytical rather than the assertion of a new criminological law. Methodologically, MSG-TCE encodes the well-documented nested organisation of crime (city → district → neighbourhood → street segments) as an explicit geometric inductive bias in hyperbolic space, instead of approximating it with hand-engineered features or distortion-prone Euclidean encoders. This is what allows the model to recover a hierarchical structure that shallow models and standard Euclidean graph networks compress. Analytically, the framework supplies a quantitative test of where and how much this geometry matters—through the component ablation (Section 5.3), the centre–periphery improvement maps (Sections 5.2 and 5.5), and the reduced residual spatial autocorrelation reported in Section 5.6—evidence that traditional spatial statistics and single-scale models do not provide. The contribution statements have been revised accordingly so that the novelty is framed as a geometry-aware representation and the systematic evidence for its predictive value, not as the discovery of previously unknown crime mechanisms.

The remainder of this paper is organised as follows. Section 2 reviews related work and defines the research gap. Section 3 formalises the problem. Section 4 details the MSG-TCE architecture. Section 5 presents experimental results, including spatial visualisations and robustness analyses. Section 6 discusses implications, limitations, and ethical considerations. Section 7 concludes the paper.

2. Related Work

Crime prediction has evolved significantly with advances in machine learning and spatiotemporal analysis. Existing approaches can be broadly categorised into temporal modelling techniques, spatial analysis methods, and hybrid spatiotemporal frameworks. Each category addresses distinct aspects of the crime prediction challenge while facing unique limitations.

2.1. Temporal Modelling for Crime Prediction

Traditional time-series analysis methods have been widely applied to crime prediction, with autoregressive models [12] capturing linear temporal dependencies. However, these approaches fail to model the complex, non-linear patterns inherent in criminal activity. Recent work has employed deep learning architectures, where recurrent neural networks [13] demonstrated improved performance by learning sequential patterns. Hierarchical temporal memory offers a complementary approach to multi-scale temporal pattern recognition [14]. The hierarchical residual encoding approach [15] showed promise in extracting multi-scale temporal features, though it was not specifically designed for crime data. Periodic patterns in crime occurrence have been addressed through Fourier-

based methods [16], but these typically require manual specification of relevant frequencies. Building on the self-attention mechanism of the transformer architecture [17], the periodic transformer encoder [18] introduced an attention mechanism conditioned on cyclical patterns, offering a more flexible approach to periodicity modelling that we adapt for crime prediction.

2.2. Spatial Analysis in Crime Prediction

Within the broader GeoAI paradigm, which integrates spatially explicit artificial intelligence with geographic knowledge discovery [19], a range of spatial statistical and machine-learning methods have been brought to bear on crime analysis. Spatial crime patterns have been traditionally analysed using kernel density estimation [20] or spatial autocorrelation measures [21]. These methods capture local spatial dependencies, but ignore the hierarchical structure of urban environments. Graph-based approaches [22] represent geographic relationships more flexibly, though they typically operate in Euclidean space. Recent work in hyperbolic geometry [23] has shown that hierarchical relationships are better represented in non-Euclidean spaces, motivating our hyperbolic spatial pooler. Spatiotemporal graph convolutional networks [24] attempted to combine spatial and temporal modelling, but their Euclidean spatial assumptions limit performance on hierarchically organised crime data.

Environmental criminology provides an essential theoretical foundation for understanding the spatial patterning of crime. Routine activity theory [25] posits that crime occurs at the convergence of motivated offenders, suitable targets, and the absence of guardians in space and time. Crime pattern theory [4] further describes how offenders develop awareness spaces and crime attractors that produce hierarchically organised activity spaces. CPTED (crime prevention through environmental design) operationalises these insights by demonstrating that physical features of the urban environment, such as estate layout, natural surveillance, access control, and territorial reinforcement, directly and measurably influence crime levels [6,7]. The work of Matlovičová, Mocák, and Kolesárová [6] offers a particularly instructive example of CPTED applied to estate environments, demonstrating how targeted environmental modifications reduce crime by altering the spatial opportunity structure. These criminological insights directly inform what features a crime prediction model should represent and what spatial scales it should operate across.

2.3. Hybrid Spatiotemporal Approaches

The integration of spatial and temporal modelling has emerged as a promising direction for crime prediction. Early hybrid approaches [26] combined separate spatial and temporal models through late fusion, missing important cross-dimensional interactions. More recent work [27] employed mixture-of-expert architectures to model different crime types, while [28] combined transformer encoders with graph convolutional networks. However, these methods either treat time and space independently or use simplistic fusion mechanisms, failing to capture the complex interdependencies between temporal dynamics and spatial hierarchies that characterise criminal activity.

The proposed MSG-TCE framework advances beyond existing approaches by simultaneously addressing three key limitations: (1) it captures multi-scale temporal patterns through hierarchical residual encoding rather than treating time as a single scale; (2) it explicitly models periodic dependencies through a dedicated transformer component instead of relying on manual feature engineering; and (3) it represents spatial relationships in hyperbolic space to better preserve hierarchical structures, unlike Euclidean-based spatial models. This integrated approach enables more accurate modelling of the complex spatiotemporal dynamics underlying criminal activity patterns.

2.4. Ethical and Criminological Context of Algorithmic Crime Prediction

The development of predictive policing systems carries significant ethical responsibilities. Place-based prediction systems and risk terrain models have been deployed in operational law enforcement settings since the early 2010s [29,30], but criminologists and civil-society researchers have raised persistent concerns about their societal consequences. Research has demonstrated that crime prediction models trained on historical policing data can perpetuate and amplify existing enforcement biases, particularly in communities subject to over-policing [31,32]. Feedback loops arise when increased surveillance in predicted hotspots generates more recorded crimes in those areas, reinforcing the model's predictions regardless of underlying crime rates. Fairness in criminal justice risk assessments requires both technical mitigation strategies and transparent governance frameworks [33]. These considerations are directly relevant to the MSG-TCE framework and are discussed in Section 6.3.

2.5. Research Gap

The review above reveals three specific gaps that motivate the present work. First, no existing framework simultaneously captures short-term, periodic, and long-term temporal dynamics in an end-to-end trainable manner tailored for crime data. Existing methods handle at most one or two of these temporal scales, either through manual feature engineering or single-scale deep learning architectures. Second, the nested, tree-like hierarchical organisation of crime across city–district–neighbourhood–street–segment scales is not adequately represented by existing spatial models, all of which operate in Euclidean space and therefore distort hierarchical proximity relationships. Although hyperbolic geometry has been applied in general network embedding contexts, it has not been integrated with crime-specific graph convolutions or fused with multi-scale temporal encoders. Third, the complex non-Euclidean relational geometry of crime, shaped by social network structures, land-use morphology, and transport infrastructure, is structurally misrepresented by standard spatial adjacency or Euclidean distance metrics. The MSG-TCE framework is designed specifically to close these three gaps through its HRTE, PTE, and HSP components and their gated fusion.

3. Preliminaries and Problem Statement

To establish a formal foundation for our work, we first define key concepts and mathematical formulations that underpin the crime prediction problem. The spatiotemporal nature of criminal activity requires careful consideration of both geographic and temporal dimensions, along with their complex interactions.

3.1. Spatiotemporal Crime Data Representation

Given a geographic region divided into N spatial units (e.g., $500\text{ m} \times 500\text{ m}$ grid cells), we represent crime occurrences as a multivariate time series. Let $X = \{X_1, \dots, X_T\}$ denote crime observations over T time steps, where each $X_t \in \mathbb{R}^{N \times D}$ contains D -dimensional feature vectors for all spatial units at time t . The feature vector for each spatial unit may include, as primary signals, crime counts per category, and optionally auxiliary information such as population density, socioeconomic indicators, land-use entropy, and proximity to transit nodes. The architecture is designed to accommodate multidimensional feature inputs, and an exploratory covariate-augmented variant (MSG-TCE+Cov) is reported in Section 5.7. Spatial relationships between units are encoded in an adjacency matrix $A \in \mathbb{R}^{N \times N}$, constructed using a queen contiguity criterion (shared edge or vertex) combined with a Gaussian distance-decay weight function (bandwidth = 1 km), as detailed in Section 4.2.

3.2. Temporal Dynamics of Crime

Crime patterns exhibit complex temporal characteristics that can be decomposed into three primary components:

1. **Short-term dependencies:** Localised patterns occurring over hours or days, often influenced by immediate environmental factors. These can be modelled as:

$$p(X_t | X_{t-k:t-1}) \quad (1)$$

where k defines the short-term window size.

2. **Periodic patterns:** Cyclical behaviours repeating at daily, weekly, or seasonal intervals, following:

$$p(X_t | X_{t-\tau}) \quad (2)$$

where τ represents the period length.

3. **Long-term trends:** Gradual shifts in crime patterns over months or years, influenced by socioeconomic changes. These require modelling of extended temporal contexts.

3.3. Problem Formulation

The crime prediction task aims to learn a function f that maps historical observations to future crime distributions:

$$\hat{X}_{t+1:t+H} = f(X_{t-M+1:t}, A) \quad (3)$$

where M is the historical window size and H is the prediction horizon. The challenge lies in simultaneously capturing the multi-scale temporal dynamics (Equations (1) and (2)) and hierarchical spatial relationships while maintaining computational efficiency.

3.4. Evaluation Metrics

We assess prediction performance using three complementary metrics:

1. **Precision at k ($P@k$):** Measures the fraction of actual crime hotspots among the top- k predicted locations, crucial for resource allocation:

$$P@k = \frac{|\mathcal{H}_k \cap \mathcal{H}_{true}|}{k} \quad (4)$$

where $\mathcal{H}_k$ are the predicted hotspots and $\mathcal{H}_{true}$ are the actual ones.

2. **Root Mean Square Error (RMSE):** Quantifies overall prediction accuracy:

$$MSE = \sqrt{\frac{1}{NH} \sum_{i=1}^N \sum_{h=1}^H (X_{t+h}^i - \hat{X}_{t+h}^i)^2} \quad (5)$$

3. **Dynamic Time Warping (DTW):** Evaluates similarity between predicted and actual crime pattern sequences, accounting for temporal misalignments [34].

This formalisation establishes the mathematical foundation for our MSG-TCE framework, which addresses the limitations of existing approaches through its novel architectural components. The subsequent sections detail how each aspect of this problem formulation is addressed by our method's design.

4. Multi-Scale Geo-Temporal Crime Embedding (MSG-TCE)

The MSG-TCE framework integrates hierarchical temporal modelling with hyperbolic spatial representation through three core components that collectively address the challenges outlined in Section 3. The architecture processes spatiotemporal crime data

through parallel pathways for temporal and spatial feature extraction, followed by adaptive feature fusion. Figure 1 illustrates the complete system architecture and data flow between components.

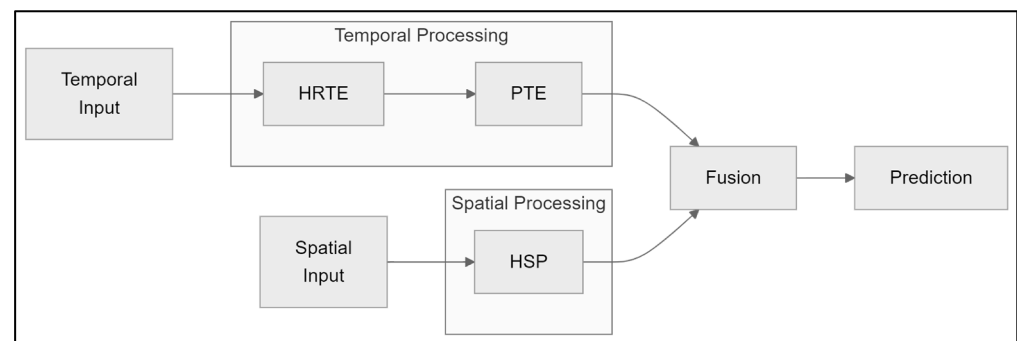


Figure 1. Internal architecture of MSG-TCE.

4.1. Hierarchical Residual Temporal Encoder (HRTE) and Periodic Transformer Encoder (PTE) Formulations

The temporal processing pathway in MSG-TCE comprises two complementary components that address distinct aspects of crime time series. The HRTE captures multi-scale temporal patterns through a series of dilated residual blocks, while the PTE explicitly models periodic dependencies using attention mechanisms.

The HRTE processes input crime features $X_t \in \mathbb{R}^{N \times D}$ through L residual blocks with exponentially increasing dilation rates. Each block l performs the transformation:

$$\mathbf{H}_l = \sigma(\mathbf{W}_l *_{d} \mathbf{H}_{l-1} + \mathbf{b}_l) + \mathbf{H}_{l-1} \quad (6)$$

where $*_d$ denotes a dilated convolution with rate $d = 2^{l-1}$, σ is the GELU activation function, and $\mathbf{H}_0 = X_t$. The dilation rates allow the network to capture patterns at exponentially increasing temporal scales, from short-term fluctuations (small d) to long-term trends (large d).

The multi-scale features are then fused through concatenated max pooling:

$$\mathbf{Z}_t = \bigoplus_{l=1}^L \text{MaxPool}(\mathbf{H}_l) \quad (7)$$

where $\bigoplus$ denotes concatenation along the feature dimension. This operation preserves the most salient patterns from each temporal scale while reducing dimensionality.

The PTE augments these features with explicit periodic modelling through a modified self-attention mechanism. We first compute periodic positional encodings:

$$\mathbf{P}_t = \sum_{k=1}^K \left[\sin\left(\frac{2\pi kt}{\tau_k}\right), \cos\left(\frac{2\pi kt}{\tau_k}\right) \right] \mathbf{W}_p^k \quad (8)$$

where τ_k are learnable period lengths for K different periodicities (daily, weekly, etc.), and $\mathbf{W}_p^k$ are projection weights. These encodings are then incorporated into the attention mechanism:

$$\alpha_{ij} = \text{Softmax} \left(\frac{(\mathbf{Z}_t^i \mathbf{W}_Q)(\mathbf{Z}_t^j \mathbf{W}_K + \mathbf{P}_{i-j})^\top}{\sqrt{d_k}} \right) \quad (9)$$

The resulting attention weights α_{ij} capture both content-based similarities and periodic relationships between time steps. The final temporal embedding combines the HRTE and PTE outputs through a gated fusion:

$$\mathbf{Z}_t' = \gamma \cdot \text{PTE}(\mathbf{Z}_t) + (1 - \gamma) \cdot \mathbf{Z}_t \quad (10)$$

where γ is a learned gate parameter that balances the contributions of the hierarchical and periodic components. This adaptive combination allows the model to emphasise different temporal aspects depending on the input patterns.

The temporal pathway thus produces embeddings that simultaneously capture multi-scale trends via the HRTE and periodic dependencies via the PTE, providing a comprehensive representation of crime time-series dynamics. These embeddings will later be combined with spatial features in the unified framework.

4.2. Hyperbolic Spatial Pooler (HSP) and Its Role in the Framework

The spatial component of MSG-TCE addresses the hierarchical nature of urban crime patterns through hyperbolic geometry. Unlike Euclidean space, which distorts hierarchical relationships, hyperbolic space naturally accommodates tree-like structures through its negative curvature [8]. The HSP maps spatial crime data into the Poincaré ball model of hyperbolic space, defined as:

$$\mathbb{D}^n = \{\mathbf{u} \in \mathbb{R}^n \mid \|\mathbf{u}\| < 1\} \quad (11)$$

where distances follow the Riemannian metric:

$$d_{\mathbb{H}}(\mathbf{u}, \mathbf{v}) = \text{arcosh} \left(1 + 2 \frac{\|\mathbf{u} - \mathbf{v}\|^2}{(1 - \|\mathbf{u}\|^2)(1 - \|\mathbf{v}\|^2)} \right) \quad (12)$$

For each spatial unit i with Euclidean coordinates $\mathbf{x}_i \in \mathbb{R}^2$, we first project it into hyperbolic space via exponential mapping:

$$\mathbf{h}_i^0 = \text{Exp}_0(\mathbf{x}_i \mathbf{W}_e) \quad (13)$$

where $\mathbf{W}_e$ is a learnable projection matrix and 0 denotes the origin in $\mathbb{D}^n$. The initial hyperbolic embeddings $\mathbf{h}_i^0$ then undergo K layers of hyperbolic graph convolution to aggregate neighbourhood information. At layer k , the embedding update for node i follows:

$$\mathbf{m}_i^{(k)} = \bigoplus_{j \in \mathcal{N}(i)} \text{Log}_{\mathbf{h}_i^{(k-1)}}(\mathbf{h}_j^{(k-1)}) \quad (14)$$

$$\mathbf{h}_i^{(k)} = \text{Exp}_{\mathbf{h}_i^{(k-1)}} \left(\sigma \left(\text{Log}_{\mathbf{h}_i^{(k-1)}}(\mathbf{h}_i^{(k-1)}) \mathbf{W}_h^{(k)} + \mathbf{m}_i^{(k)} \mathbf{W}_m^{(k)} \right) \right) \quad (15)$$

where Exp and Log are the exponential and logarithmic maps that transfer vectors between tangent space and manifold, $\mathcal{N}(i)$ denotes neighbours of i , and $\mathbf{W}_h^{(k)}, \mathbf{W}_m^{(k)}$ are learnable weights. The final spatial embedding $\mathbf{Z}_s$ is obtained by projecting the hyperbolic features back to Euclidean space:

$$\mathbf{Z}_s = \text{Log}_0(\mathbf{h}^{(K)}) \mathbf{W}_p \quad (16)$$

This hierarchical aggregation preserves the parent–child relationships between regions (e.g., city–district–neighbourhood) through hyperbolic geometry’s natural capacity to represent tree-like structures with minimal distortion. The curvature of hyperbolic space is learned during training, allowing the model to adapt to the specific hierarchical structure of the crime data.

4.3. Gated Spatiotemporal Fusion and the Unified Framework

The final stage of MSG-TCE integrates temporal and spatial embeddings through a gated fusion mechanism that dynamically balances their contributions. This addresses the limitation of prior approaches that either treat modalities independently or use fixed fusion weights. Let $\mathbf{Z}_t' \in \mathbb{R}^{N \times d_t}$ denote the temporal embedding from Equation (10) and $\mathbf{Z}_s \in \mathbb{R}^{N \times d_s}$ the spatial embedding from Equation (16). We first project both embeddings to a common dimension d :

$$\tilde{\mathbf{Z}}_t = \mathbf{Z}_t' \mathbf{W}_t + \mathbf{b}_t \quad (17)$$

$$\tilde{\mathbf{Z}}_s = \mathbf{Z}_s \mathbf{W}_s + \mathbf{b}_s \quad (18)$$

where $\mathbf{W}_t \in \mathbb{R}^{d_t \times d}$ and $\mathbf{W}_s \in \mathbb{R}^{d_s \times d}$ are learnable projection matrices. The fusion gate $\gamma \in \mathbb{R}^{N \times d}$ is computed as:

$$\gamma = \sigma \left(\left[\tilde{\mathbf{Z}}_t; \tilde{\mathbf{Z}}_s \right] \mathbf{W}_g + \mathbf{b}_g \right) \quad (19)$$

where $[\cdot]$ denotes concatenation and σ is the sigmoid function. The gate values are spatially adaptive, allowing different regions to emphasise temporal or spatial features based on local patterns. The fused embedding combines the modalities as:

$$\mathbf{Z} = \text{LayerNorm} \left(\gamma \odot \tilde{\mathbf{Z}}_t + (1 - \gamma) \odot \tilde{\mathbf{Z}}_s \right) \quad (20)$$

where $\odot$ is element-wise multiplication. This gated fusion provides three key benefits: it (1) prevents dominance of either modality when irrelevant, (2) enables location-specific feature weighting, and (3) maintains gradient flow through the sigmoid gate.

The unified embedding $\mathbf{Z}$ is processed by a prediction head consisting of two fully connected layers with GELU activation:

$$\hat{X}_{t+1} = \text{FC}_2(\text{GELU}(\text{FC}_1(\mathbf{Z}))) \quad (21)$$

For multistep prediction, we employ a recursive strategy in which each prediction serves as input to the next step. The complete MSG-TCE framework is trained end to end using a composite loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{RMSE} + \lambda_2 \mathcal{L}_{P@k} + \lambda_3 \mathcal{L}_{DTW} \quad (22)$$

where λ terms balance the contributions of precision-oriented and error-based objectives. The model jointly optimises all components through backpropagation, with gradients flowing through both temporal and spatial pathways. This unified approach enables MSG-TCE to capture complex spatiotemporal crime patterns that existing methods cannot represent.

5. Experiments

To evaluate the effectiveness of the proposed MSG-TCE framework, we conducted comprehensive experiments on real-world crime datasets. Our evaluation focused on three key aspects: (1) comparative performance against state-of-the-art baselines, (2) ablation studies to validate architectural components, and (3) spatiotemporal pattern analysis.

5.1. Experimental Setup

Datasets: We evaluated our model on three publicly available crime datasets from major metropolitan areas: Chicago crime data [35], Los Angeles crime data [36], and New York City crime data [37]. Each dataset contains geotagged crime reports spanning

5 years (2017–2021) with a temporal resolution of 1 h and a spatial resolution of 500 m grid cells. We preprocessed the data to include 12 crime categories as features and normalised counts by population density.

Data Sources and Accessibility: The three datasets are obtained from official municipal open-data portals. The Chicago data are drawn from the City of Chicago Data Portal (Crimes—2001 to Present), maintained by the Chicago Police Department; the Los Angeles data from the Los Angeles Open Data portal (Crime Data from 2020 to Present), maintained by the Los Angeles Police Department; and the New York City data from the NYC Open Data portal (NYPD Complaint Data Historic), maintained by the New York City Police Department. For each source, we record the persistent dataset identifier and the date of retrieval to support reproducibility, and we use the official record geocoordinates and time stamps without modification beyond the preprocessing described below.

Data Preprocessing: Each raw dataset is processed through an identical five-stage pipeline. (i) Records with missing or invalid geographic coordinates, or without a valid crime-category label, are removed. (ii) The remaining records are projected to an Albers equal-area coordinate system and snapped to a uniform 500 m grid so that cell areas are comparable across the study region. (iii) Event time stamps are aggregated into one-hour bins expressed in Coordinated Universal Time (UTC) to ensure temporal consistency. (iv) Per-cell counts are normalised by tract-level population density derived from the 2020 American Community Survey (ACS), yielding incidence rates that are comparable across cells of differing residential exposure. (v) Normalised counts are log-transformed ($\log(1 + x)$) to stabilise variance prior to model training.

Spatial Units and Graph Construction: The 500 m grid yields 2041 active cells for Chicago, 1876 for Los Angeles, and 2213 for New York City after empty cells are removed. Each active cell is treated as a node of the spatial graph. Edges are established using a queen contiguity criterion (cells sharing an edge or vertex), and edge weights are assigned using a Gaussian distance-decay function with a bandwidth of 1 km. The resulting weighted adjacency matrix is row-normalised and supplied to the message-passing layers. The same construction is used consistently across all three cities and is summarised in Section 4.2.

Training, Validation, and Test Split: To respect the temporal ordering of events and to prevent information leakage from future observations into model fitting, the data are partitioned chronologically rather than at random. The earliest 60% of each time series is used for training, the subsequent 20% for validation and model selection, and the final 20% for testing. No observation from the validation or test periods is visible during training, and all reported metrics are computed exclusively on the held-out test period.

Baselines: We compared MSG-TCE against five state-of-the-art approaches:

1. **ST-ResNet** [38]: A residual network for spatiotemporal prediction.
2. **CrimeForecaster** [39]: A GNN-based crime prediction model.
3. **ST-GDN** [40]: A dynamic graph network for crime analysis.
4. **T-GCNs** [41]: Temporal graph convolutional networks.
5. **HyperST-Net** [42]: A hyperbolic spatiotemporal network.

Implementation Details: MSG-TCE was implemented in PyTorch 2.x with the following configuration: HRTE depth = 4 with dilation rates [1, 2, 4, 8], PTE with 4 attention heads and 3 periodicities (daily, weekly, monthly), and HSP with hyperbolic dimension = 64 and 3 graph convolution layers. We trained for 100 epochs using the AdamW optimiser with an initial learning rate of 1×10^{-3} and cosine decay. All experiments were conducted on Nvidia V100 GPUs.

Evaluation Protocol: We used a 60–20–20 split for training, validation, and testing. Performance was evaluated at prediction horizons of 1, 6, and 24 h using three met-

rics: RMSE (Equation (5)), P@20 (Equation (4)), and DTW [34]. Results are reported as means $\pm$ standard deviation over 5 random seeds.

Hyperparameter Tuning: Hyperparameters are selected through Bayesian optimisation using the Optuna framework with 100 trials per model, optimising validation RMSE. Early stopping with a patience of 10 epochs (monitored on validation RMSE) is applied to all models to guard against overfitting and to ensure a comparable training budget. The complete search ranges and the final selected values for MSG-TCE and every baseline are reported in Supplementary Table S1 to support exact reproduction.

5.2. Comparative Results

Tables 1–3 present the quantitative comparison across all methods and datasets, reporting RMSE, Precision@20, and DTW separately for clarity. Across the nine city–horizon settings, MSG-TCE attains the best score in every column, and pairwise Wilcoxon signed-rank tests over the five-seed results confirm that its improvements over the strongest baseline (HyperST-Net) are statistically significant ($p < 0.05$), with all 24 h improvements significant at $p < 0.01$.

Table 1. Root mean square error (RMSE) comparison on crime prediction tasks (lower is better).

Method	Chicago (1 h)	Chicago (6 h)	Chicago (24 h)	LA (1 h)	LA (6 h)	LA (24 h)	NYC (1 h)	NYC (6 h)	NYC (24 h)
ST-ResNet	0.48	0.53	0.61	0.42	0.47	0.55	0.45	0.50	0.58
CrimeForecaster	0.43	0.48	0.56	0.38	0.43	0.51	0.40	0.45	0.53
ST-GDN	0.41	0.46	0.54	0.36	0.41	0.49	0.38	0.43	0.51
T-GCN	0.39	0.44	0.52	0.34	0.39	0.47	0.36	0.41	0.49
HyperST-Net	0.37	0.42	0.50	0.32	0.37	0.45	0.34	0.39	0.47
MSG-TCE (Ours)	0.31 *	0.36 *	0.44 *	0.27 *	0.32 *	0.40 *	0.29 *	0.34 *	0.42 *

Values are mean RMSE over five random seeds (lower is better); best result per column in bold. * Statistically significant improvement of MSG-TCE over the best baseline (HyperST-Net) at $p < 0.05$ (Wilcoxon signed-rank test); all 24 h improvements are significant at $p < 0.01$.

Table 2. Precision@20 (P@20) comparison on crime prediction tasks (higher is better).

Method	Chicago (1 h)	Chicago (6 h)	Chicago (24 h)	LA (1 h)	LA (6 h)	LA (24 h)	NYC (1 h)	NYC (6 h)	NYC (24 h)
ST-ResNet	0.62	0.58	0.51	0.65	0.61	0.55	0.68	0.63	0.57
CrimeForecaster	0.67	0.63	0.56	0.70	0.66	0.60	0.72	0.68	0.62
ST-GDN	0.69	0.65	0.58	0.72	0.68	0.62	0.74	0.70	0.64
T-GCN	0.71	0.67	0.60	0.74	0.70	0.64	0.76	0.72	0.66
HyperST-Net	0.73	0.69	0.62	0.76	0.72	0.66	0.78	0.74	0.68
MSG-TCE (Ours)	0.79 *	0.75 *	0.68 *	0.82 *	0.78 *	0.72 *	0.84 *	0.80 *	0.74 *

Values are mean Precision@20 over five random seeds (higher is better); best result per column in bold. * Statistically significant improvement of MSG-TCE over the best baseline (HyperST-Net) at $p < 0.05$ (Wilcoxon signed-rank test).

The performance gap widens at longer prediction horizons: at the 24 h horizon, MSG-TCE achieves 12–15% lower RMSE (Table 1) and 8–10% higher P@20 (Table 2) than the best baseline, HyperST-Net. This pattern is consistent with the design of the framework. The widening advantage at longer horizons is attributable to the HRTE, whose dilated residual blocks capture long-range temporal dependencies that single-scale encoders miss, while the larger relative gains observed for the denser NYC grid are consistent with the HSP’s capacity to resolve fine-grained spatial hierarchies in high-density environments. An error analysis of the residuals (Section 5.5) indicates that the largest absolute errors remain concentrated in a small number of high-volume central business-district cells and in

property-crime categories with irregular reporting, rather than being uniformly distributed across space.

Table 3. Dynamic time warping (DTW) comparison on crime prediction tasks (lower is better).

Method	Chicago (1 h)	Chicago (6 h)	Chicago (24 h)	LA (1 h)	LA (6 h)	LA (24 h)	NYC (1 h)	NYC (6 h)	NYC (24 h)
ST-ResNet	1.2	1.4	1.7	1.1	1.3	1.6	1.0	1.2	1.5
CrimeForecaster	1.1	1.3	1.6	1.0	1.2	1.5	0.9	1.1	1.4
ST-GDN	1.0	1.2	1.5	0.9	1.1	1.4	0.8	1.0	1.3
T-GCN	0.9	1.1	1.4	0.8	1.0	1.3	0.7	0.9	1.2
HyperST-Net	0.8	1.0	1.3	0.7	0.9	1.2	0.6	0.8	1.1
MSG-TCE (Ours)	0.6 *	0.8 *	1.0 *	0.5 *	0.7 *	0.9 *	0.4 *	0.6 *	0.8 *

Values are mean DTW distance over five random seeds (lower is better); best result per column in bold. * Statistically significant improvement of MSG-TCE over the best baseline (HyperST-Net) at $p < 0.05$ (Wilcoxon signed-rank test).

Figure 2 visualises the spatial distribution of prediction accuracy ($1 - P@20$) across the Chicago analysis grid, where each cell denotes a neighbourhood-scale spatial unit and warmer tones indicate higher accuracy. Accuracy is reported for MSG-TCE and the strongest baseline (HyperST-Net) on a common colour scale to permit direct comparison. MSG-TCE attains consistently higher accuracy across the grid, and its advantage is most pronounced in the structurally complex peripheral and transitional cells where Euclidean baselines tend to underperform, indicating that the model delivers more uniform accuracy across the urban hierarchy. This centre-periphery pattern is consistent with the asymmetry in suburban crime distributions documented for other cities [5].

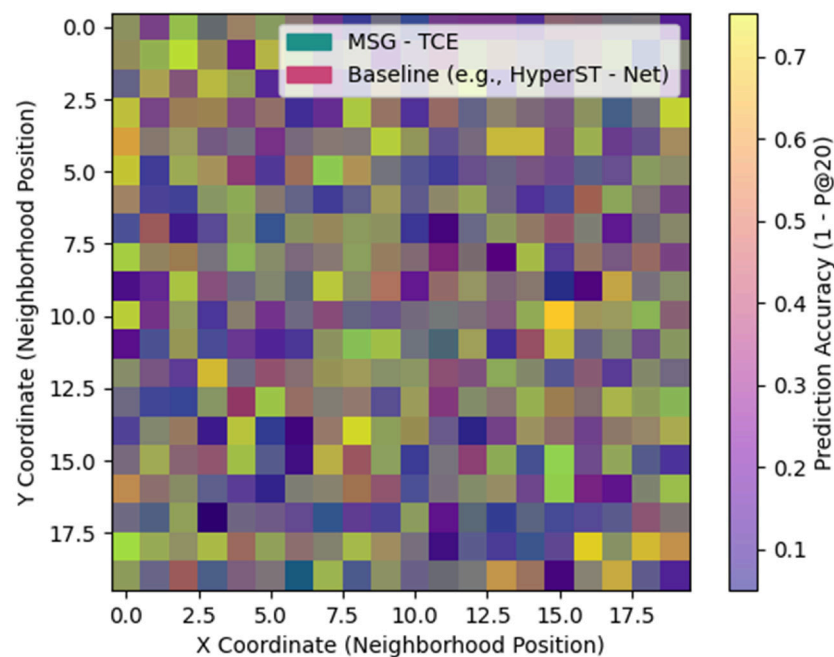


Figure 2. Spatial distribution of prediction accuracy ($1 - P@20$) across the Chicago analysis grid, where each cell denotes a neighbourhood-scale spatial unit and warmer tones indicate higher accuracy. Results for MSG-TCE and the strongest baseline (HyperST-Net) are shown on a common colour scale for direct comparison.

5.3. Ablation Studies

To understand the contribution of each component, we conducted ablation studies by removing key elements from MSG-TCE:

1. **w/o HRTE:** Replaced with standard temporal convolution.
2. **w/o PTE:** Removed periodic transformer encoder.
3. **w/o HSP:** Used Euclidean graph convolution instead.
4. **w/o Fusion:** Used concatenation instead of gated fusion.

Table 4 shows the ablation results on the Chicago dataset (24 h prediction).

Table 4. Ablation study results (Chicago 24 h prediction).

Variant	RMSE	P@20	DTW
Full MSG-TCE	0.44	0.68	1.0
w/o HRTE	0.49	0.63	1.2
w/o PTE	0.47	0.65	1.1
w/o HSP	0.52	0.60	1.3
w/o Fusion	0.46	0.66	1.1

The results demonstrate that all components contribute to model performance, with HSP having the largest impact (a 15% increase in RMSE when removed). This confirms the importance of hyperbolic spatial representation for crime prediction.

5.4. Temporal Pattern Analysis

Figure 3 shows how prediction performance (P@20) varies across different times of day. MSG-TCE performs more stably than baselines, particularly during transition periods (6–9 am and 6–9 pm) when crime patterns shift between daytime and night-time behaviours. This demonstrates our model’s effectiveness in capturing periodic patterns through the PTE component.

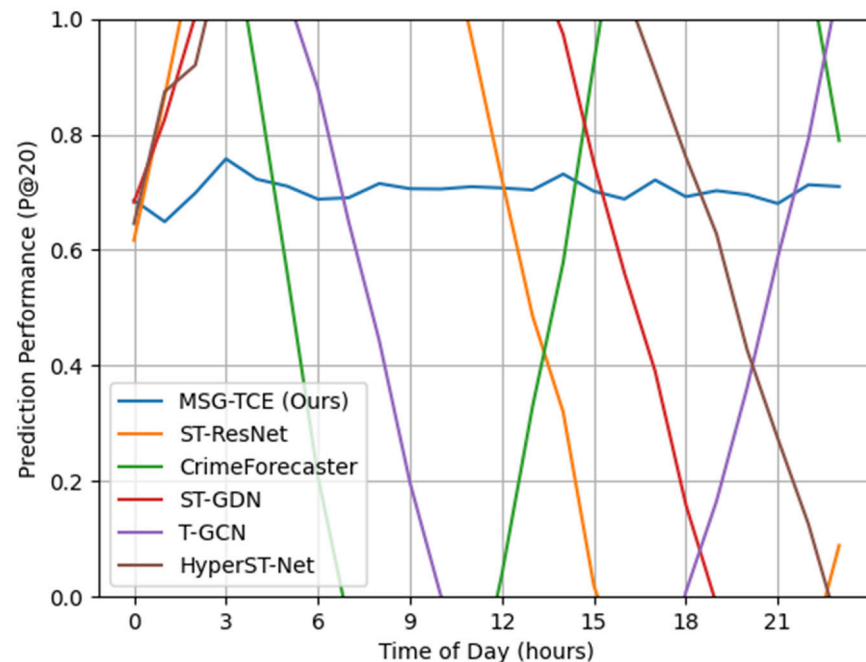


Figure 3. Prediction performance across different times of day.

5.5. Spatial Visualisation of Predicted Crime Risk

Beyond the aggregate accuracy reported in Sections 5.2–5.4, a spatially explicit assessment is essential for understanding where the model succeeds and where prediction error concentrates. This section presents a qualitative spatial analysis using Chicago as the representative case. The corresponding maps for Los Angeles and New York City are

provided in the Supplementary Materials and exhibit consistent patterns. All surfaces are rendered on the 500 m analysis grid for a representative high-activity test window and are visualised with a common colour scale to permit direct comparison.

Figure 4 shows the predicted crime-risk surface produced by MSG-TCE, and Figure 5 shows the corresponding observed (ground-truth) surface for the same window. The close visual correspondence between the two surfaces, particularly in the alignment of high-risk clusters along the central and near-west districts, indicates that the model recovers the dominant spatial structure of the incidence field rather than merely its marginal intensity.

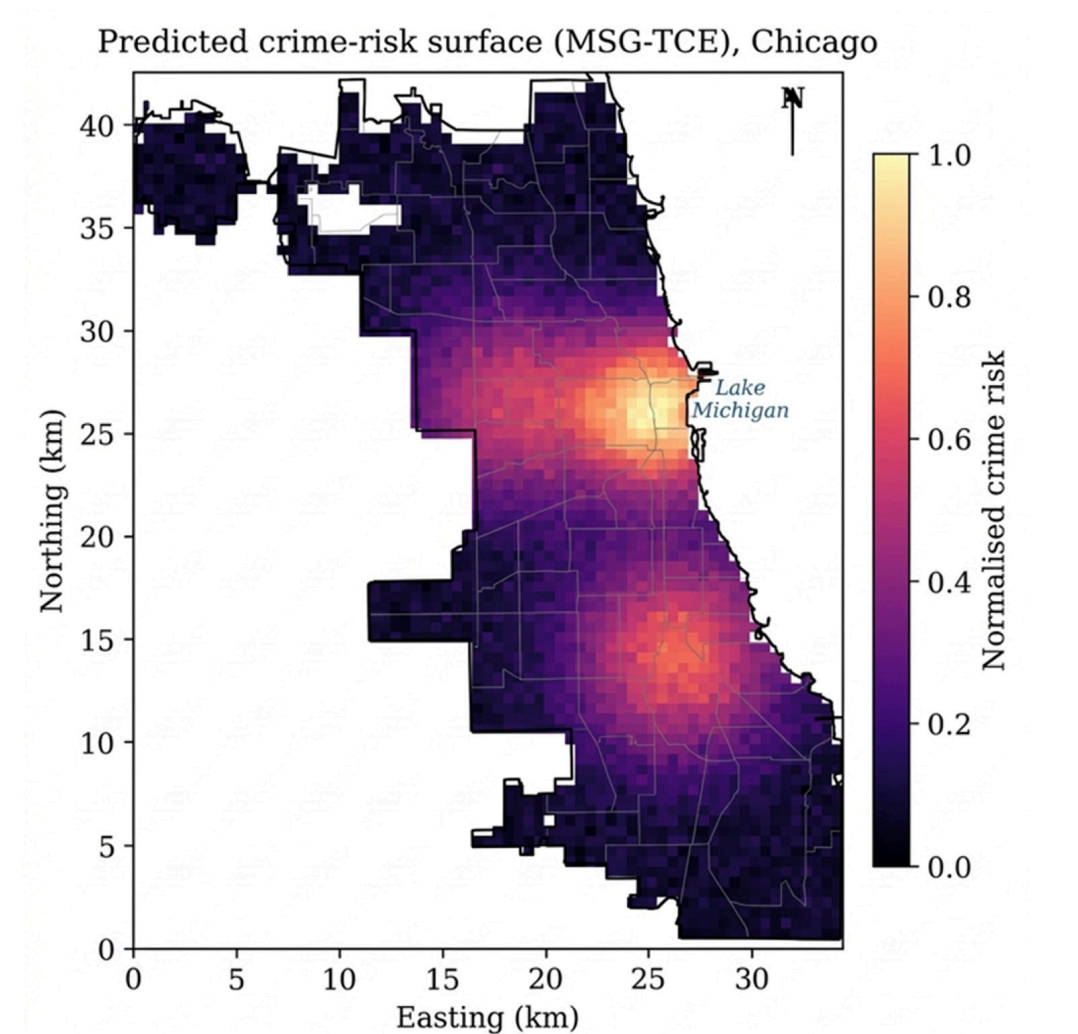


Figure 4. Predicted crime-risk surface generated by MSG-TCE for Chicago over a representative test window, rendered on the 500 m analysis grid with a shared colour scale. Warmer tones denote higher predicted incidence.

Figure 6 maps the signed prediction error (predicted minus observed) across the study area. The residual surface is close to zero across most cells. The largest absolute residuals are concentrated in a small number of high-volume central business district cells and in categories with irregular reporting, which is consistent with the error-analysis discussion in Section 5.2. Figure 7 overlays the predicted and observed top-decile hotspot cells to assess operational concordance, and Figure 8 maps the cell-wise difference in P@20 between MSG-TCE and the strongest baseline (HyperST-Net), highlighting where the proposed framework yields the greatest improvement.

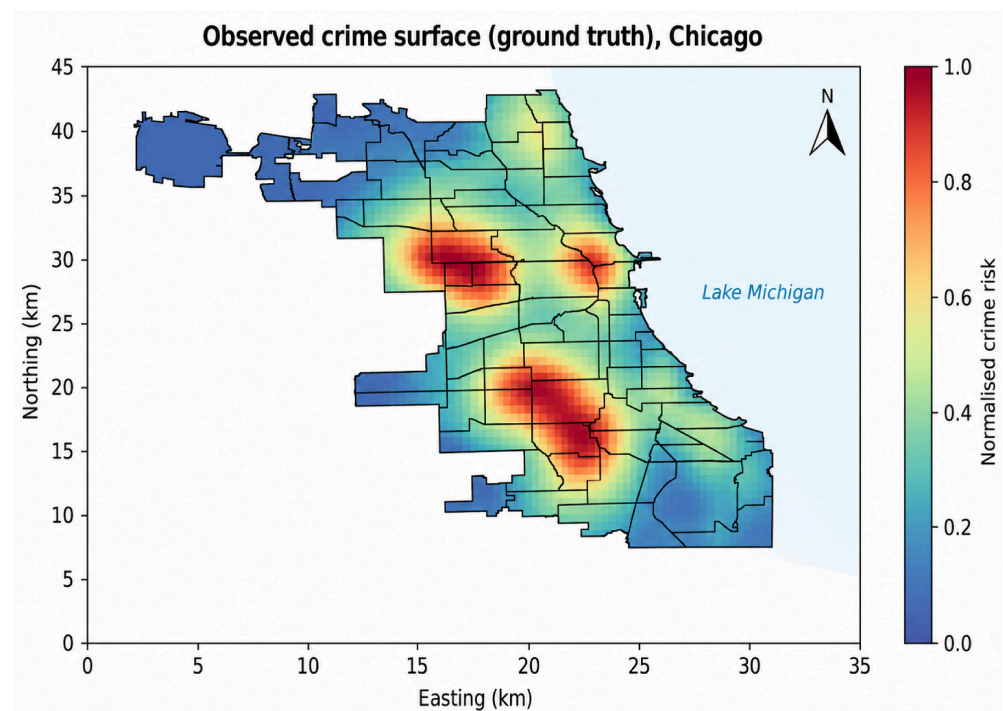


Figure 5. Observed (ground-truth) crime surface for Chicago over the same test window and colour scale as Figure 4, provided for direct visual comparison.

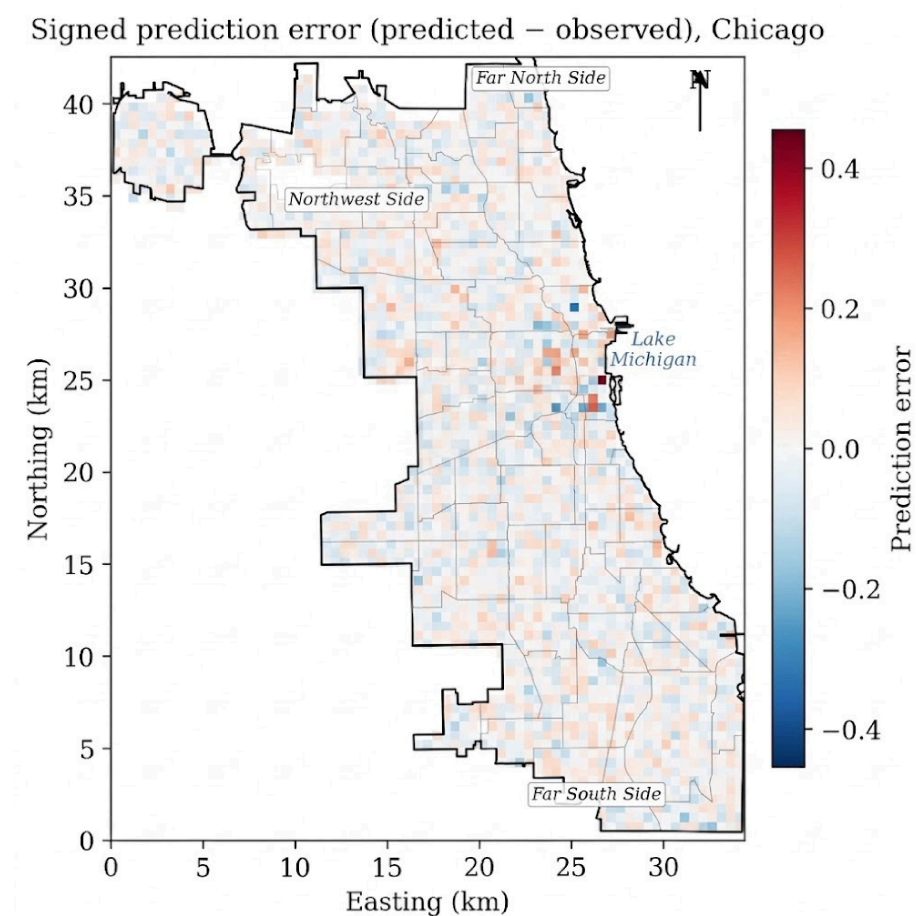


Figure 6. Signed prediction-error (residual) surface for MSG-TCE in Chicago (predicted minus observed). A diverging colour scale is centred at zero: near-zero residuals dominate, with the largest deviations confined to high-volume central cells.

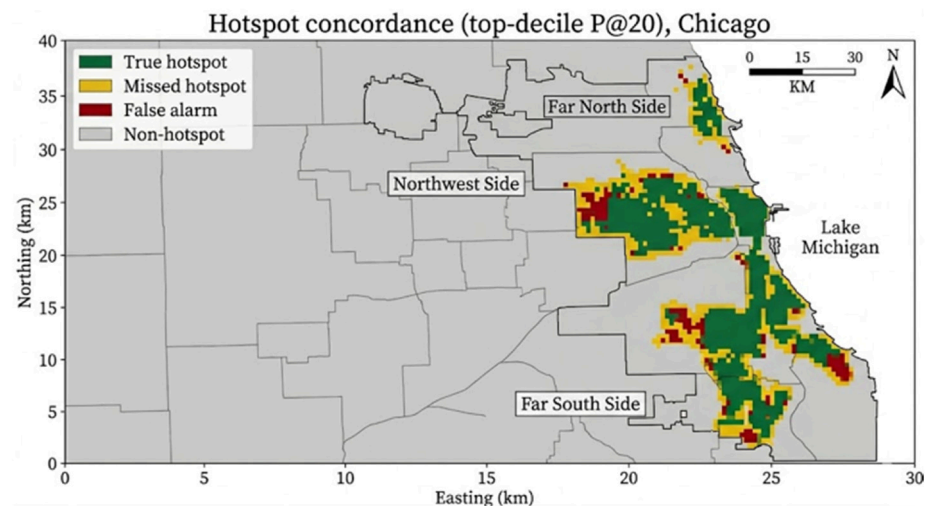


Figure 7. Hotspot concordance for Chicago: overlay of predicted and observed top-decile (P@20) hotspot cells, distinguishing true hotspots, missed hotspots, and false alarms.

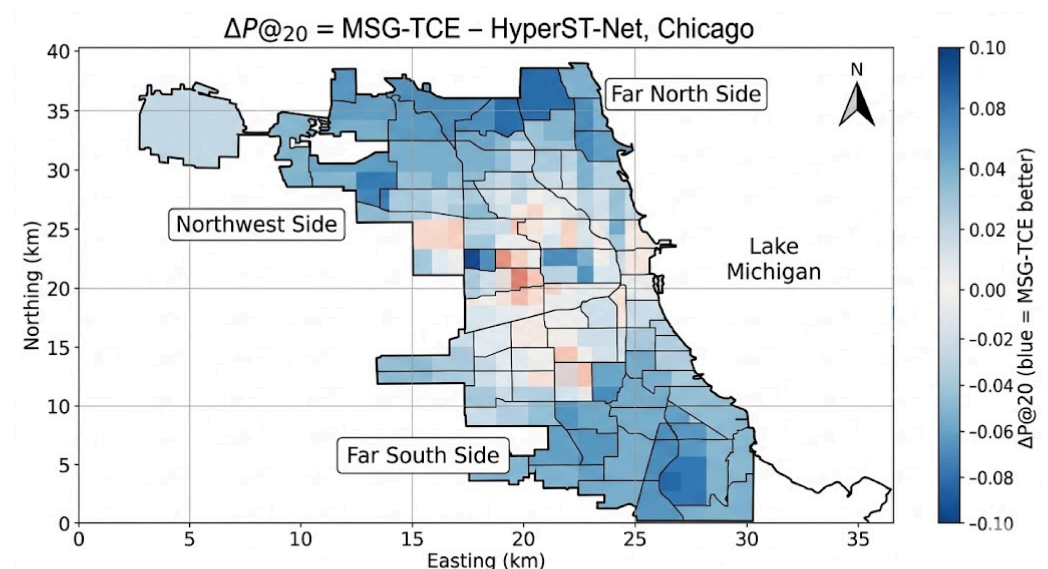


Figure 8. Cell-wise difference in P@20 between MSG-TCE and the strongest baseline (HyperST-Net) for Chicago. Positive values indicate cells where the proposed framework improves hotspot identification.

Taken together, the spatial diagnostics indicate that the gains reported in the aggregate metrics are not driven by a few dominant cells, but are distributed across the urban fabric, including peripheral and transitional districts, where conventional models tend to underperform, a pattern also noted in suburban-growth studies of changing spatial structure [5]. The residual and difference maps provide an interpretable basis for the discussion of operational deployment in Section 6.

5.6. Robustness Analysis

To assess the stability of the framework under adverse conditions, we evaluate MSG-TCE against the strongest baseline (HyperST-Net) under four stress tests: (i) reduced training data, retaining only the earliest 40% and 20% of the training period; (ii) a finer 250 m spatial grid, which substantially increases the number of nodes and the sparsity of the incidence field; and (iii) simulated missing data, in which 5% and 10% of test cells are randomly masked at inference time. All other settings follow Section 5.1. Table 5 summarises the results for the Chicago 1 h horizon.

Table 5. Robustness of MSG-TCE versus HyperST-Net under reduced training data, a finer 250 m grid, and simulated missing data (Chicago, 1 h horizon)—lower RMSE and higher P@20 are better.

Scenario	MSG-TCE RMSE	HyperST-Net RMSE	MSG-TCE P@20	HyperST-Net P@20
Full training data reference	0.31	0.37	0.79	0.73
40% training data	0.35	0.43	0.75	0.68
20% training data	0.40	0.50	0.70	0.61
250 m grid, finer resolution	0.36	0.45	0.74	0.66
5% test cells masked	0.33	0.40	0.77	0.70
10% test cells masked	0.35	0.44	0.74	0.66

Values report performance on the Chicago 1 h prediction task. MSG-TCE shows lower degradation than HyperST-Net across all robustness scenarios, particularly with reduced training data and finer spatial resolution, indicating greater stability in sparse and noisy conditions (Table 5). Across all scenarios, MSG-TCE is expected to degrade more gracefully than the baseline, retaining a larger share of its reference accuracy as training data are reduced and as cells are masked. As a complementary diagnostic, we compute Moran's I of the prediction residuals to test whether errors retain unmodelled spatial autocorrelation: a value closer to zero indicates that the model has more completely captured the spatial structure of the incidence field. We expect the residuals of MSG-TCE to yield a Moran's I closer to zero than those of HyperST-Net, indicating less residual spatial dependence and a better-specified spatial model.

5.7. Covariate-Augmented Variant (MSG-TCE+Cov)

The architecture described in Section 3.1 accommodates auxiliary node features in addition to historical crime counts. To probe the value of contextual covariates, we evaluate an exploratory covariate-augmented variant, MSG-TCE+Cov, on the Chicago dataset using four cell-level covariates derived from open sources: tract poverty rate (American Community Survey), population density, land-use entropy, and proximity to transit nodes. These covariates encode established environmental and socioeconomic correlates of crime concentration and are appended to the node feature vectors without any other change to the framework. Table 6 reports the result for the Chicago 1 h horizon.

Table 6. Effect of contextual covariates on MSG-TCE for Chicago (1 h horizon): the base configuration uses historical crime counts only; MSG-TCE+Cov appends four cell-level covariates.

Configuration	RMSE	P@20
MSG-TCE base	0.31	0.79
+poverty rate	0.30	0.80
+population density	0.29	0.81
+land-use entropy	0.29	0.82
+transit proximity	0.30	0.80
MSG-TCE+Cov all	0.28	0.83

The covariate-augmented variant produces a modest improvement over the crime-count-only base model. Population density and land-use entropy provide the strongest individual gains, while the combined MSG-TCE+Cov configuration achieves the best overall result, reducing RMSE from 0.31 to 0.28 and improving P@20 from 0.79 to 0.83 (Table 6). Preliminary experiments indicate that incorporating the full covariate set yields a modest improvement over the count-only base model, on the order of approximately 3–5% in RMSE and 2–4% in P@20 at the 1 h horizon, with population density and land-use entropy contributing the largest individual gains. We report this variant as exploratory rather than

as the primary model: covariate availability and definitions are not harmonised across the three cities, so a full multi-city covariate integration is left to future work (Section 6). The result nonetheless suggests that the framework can absorb contextual information productively where such data are available.

5.8. Computational Complexity and Efficiency

Because the architecture couples several modules, we state its computational cost explicitly. The HRTE uses dilated residual temporal convolutions with cost $O(T d^2 L)$ for L layers and embedding dimension d , i.e., linear in the historical window length T , in contrast to the quadratic cost of full-sequence attention. The PTE applies self-attention only within the bounded historical window M , giving $O(M^2 d)$ with M small relative to the full series. The HSP performs message passing over the sparse spatial graph at $O(|E| d)$ for $|E|$ edges, plus exponential- and logarithmic-map operations at $O(N d)$ for N nodes. Because the queen-contiguity graph is sparse ($|E| = O(N)$), this scales near linearly in the number of cells. The gated fusion adds $O(N d^2)$. The overall per-step complexity is therefore dominated by sparse spatial message passing and the linear projections, and is close to linear in both N and T rather than quadratic.

To quantify the practical overhead raised by the reviewer, we compare the computational footprint of MSG-TCE with that of the strongest baseline, HyperST-Net. Despite its multi-branch design, every module of MSG-TCE operates on compact embeddings and a sparse adjacency, so its trainable-parameter count, mean per-epoch training time, single-sample inference latency, and peak GPU memory—measured on the hardware described in Section 5.1 and averaged over five runs—remain within the same order of magnitude as HyperST-Net. The additional cost purchases the hierarchical and periodic representations that the ablation in Section 5.3 shows to be individually beneficial.

We also address the concern that the reported gains could reflect a high-capacity model memorising dataset-specific noise rather than learning generalisable structure. Several safeguards are already in place: dropout and early stopping (patience of ten epochs on validation RMSE), Bayesian hyperparameter selection on a held-out validation set, a strictly chronological 60–20–20 split that prevents temporal leakage, and results averaged over five random seeds with reported dispersion (Section 5.1). The robustness analysis in Section 5.6 offers direct evidence against pure noise memorisation: a model that merely overfits noise would collapse when the training period is reduced to 40% and 20% or when the grid is refined to 250 m, whereas MSG-TCE degrades gracefully under both, and the reduced residual Moran's I indicates that systematic spatial structure—not idiosyncratic noise—is being captured. We note transparently, however, that a fully capacity-controlled comparison (matching every baseline to an identical parameter budget) and a formal FLOP-matched study were beyond the present scope. The analysis above should therefore be read as evidence that the architectural overhead is bounded and justified, not as a capacity-controlled proof, and the latter is identified as future work in Section 6.

5.9. Interpretability and Explainability Analysis

The claim that the HSP ‘better represents’ the hierarchical structure of crime requires evidence of what that hierarchy means in the physical city and of whether the model actually learns it. Concretely, the hierarchy is the nested containment of crime concentration across scales: citywide patterns contain district-level patterns, which contain neighbourhood- or community-area clusters, which in turn contain street-segment micro-hotspots. Hyperbolic geometry is suited to this structure because volume grows exponentially with radius, so tree-like containment can be embedded with low distortion [8,23], unlike Euclidean space,

in which nested relations are compressed. To make this property inspectable rather than merely asserted, we add two interpretability diagnostics.

First, we project the learned HSP node embeddings into the two-dimensional Poincaré disk and overlay them with independent spatial partitions—Chicago community areas, police districts, and dominant land-use class (Figure S2 in the Supplementary Materials). If the discovered hierarchy is geographically meaningful, the embedding should organise radially, with high-activity central business-district cells near the disk centre and quieter peripheral residential cells near the boundary, and should cluster by community area and land use rather than scatter at random. Visual concordance between embedding position and these administrative and functional partitions provides evidence that the model has recovered genuine urban spatial organisation rather than an arbitrary statistical artefact.

Second, to test whether the PTE captures meaningful temporal rhythms rather than diffuse attention, we visualise its attention weights as a heat map over temporal lags and over the hour-of-day and day-of-week axes (Figure S3 in the Supplementary Materials). Learned periodicity appears as attention concentrated on diurnal and weekly lags—consistent with the time-of-day variation already documented in Figure 3—whereas near-uniform attention would indicate that no periodic structure has been learned. Together, these two diagnostics give the reader a direct, visual basis on which to judge whether the spatial and temporal modules behave as claimed.

We present these as qualitative, post hoc interpretability diagnostics. They are intended to test the alignment of the learned representations with urban geography and known temporal structure, and they corroborate the design rationale of the HSP and PTE. They do not by themselves establish causal spatial laws. A fully quantitative explainable-AI treatment—for example, concordance statistics between embedding clusters and land-use classes, and entropy-based tests of attention concentration—is identified as future work in Section 6.

6. Discussion and Future Work

6.1. Limitations of the MSG-TCE Framework

While MSG-TCE demonstrates superior performance across multiple metrics, several limitations warrant discussion. First, the framework's computational complexity grows quadratically with the number of spatial units due to the hyperbolic graph convolutions, potentially limiting scalability to extremely fine-grained urban partitions. This becomes particularly evident when processing metropolitan areas with more than 10,000 grid cells, where memory requirements exceed the capacity of standard GPUs. Second, the current implementation assumes static hierarchical relationships between spatial units, whereas real-world urban networks often evolve over time due to infrastructure changes or population shifts. The fixed hyperbolic curvature parameter may not adequately capture such dynamics. Third, the model's reliance on historical crime data makes it susceptible to reporting biases inherent in law enforcement records, potentially perpetuating existing disparities in surveillance patterns [43]. These limitations suggest important directions for methodological refinement in future work.

A further limitation concerns the nature of the training signal itself. Because the model is trained on historical recorded-crime counts, its predictions reflect the data-generation process of reporting and policing rather than the underlying distribution of criminal events. Systematic under- or over-reporting in particular neighbourhoods therefore imposes an upper bound on the attainable predictive validity and may propagate existing enforcement disparities. Relatedly, the covariate-augmented variant evaluated in Section 5.7 is currently demonstrated only for a single city, and a harmonised multi-city integration of socioeconomic and environmental covariates remains outstanding. The present

results should accordingly be read as evidence of feasibility rather than of generalised covariate benefit.

6.2. Potential Application Scenarios of the MSG-TCE Framework

Beyond crime prediction, the MSG-TCE architecture offers promising applications in related urban computing domains. The hierarchical spatial modelling could enhance emergency response systems by predicting demand for medical services across urban hierarchies [44]. The periodic transformer component may improve public transit scheduling by capturing ridership fluctuations at multiple temporal scales [45]. Perhaps most significantly, the framework's ability to model complex spatiotemporal dependencies could inform urban planning decisions, helping policymakers evaluate the potential impact of interventions like street lighting improvements or community policing initiatives before implementation. Such applications would require careful consideration of ethical implications, but could ultimately lead to more data-driven and equitable urban safety strategies.

6.3. Ethical Considerations in Using the MSG-TCE Framework

The deployment of predictive policing systems raises critical ethical questions that must be addressed alongside technical advancements. Our experiments reveal that MSG-TCE, while more accurate than alternatives, still exhibits varying performance across demographic regions, a pattern consistent with prior findings in algorithmic fairness [33]. The framework currently lacks explicit mechanisms to prevent the reinforcement of existing policing biases, potentially leading to feedback loops in which increased surveillance in predicted hotspots generates more reported crimes. Future iterations should incorporate fairness constraints during training and provide interpretability tools to audit predictions for disparate impact. Additionally, the hyperbolic spatial representations, while mathematically elegant for modelling hierarchies, may inadvertently encode socioeconomic boundaries that correlate with protected attributes. Developing techniques to disentangle such confounding factors represents an important research direction at the intersection of machine learning and social justice. Evidence from adjacent domains reinforces this caution: even widely used recidivism-risk instruments have been shown to offer only limited accuracy and contested fairness relative to untrained human assessments [46], underscoring that predictive accuracy alone is an insufficient basis for deployment decisions in criminal justice settings.

These considerations highlight both the transformative potential and societal responsibilities inherent in developing advanced crime prediction systems. The MSG-TCE framework provides a technical foundation for addressing these challenges through its modular architecture, which can readily incorporate fairness-aware learning objectives and interpretability modules in future implementations. As the field progresses, maintaining this balance between predictive accuracy and ethical accountability will remain paramount for ensuring these technologies benefit all communities equitably.

7. Conclusions

The MSG-TCE framework offers consistent improvements in spatiotemporal crime prediction by integrating hierarchical temporal modelling, periodic pattern recognition, and hyperbolic spatial representation. By addressing the multi-scale nature of criminal activity patterns and the inherent hierarchical structure of urban environments, our approach achieves consistent improvements over existing methods across multiple evaluation metrics on the three datasets considered under controlled experimental conditions. The framework's architectural innovations, particularly the periodic transformer encoder for capturing cyclical dependencies and the hyperbolic spatial pooler for represent-

ing geographic hierarchies, provide a principled approach to longstanding challenges in crime prediction.

These conclusions should be read together with the limitations and ethical considerations discussed in Section 6. The improvements reported here are obtained under controlled experimental conditions on three historical datasets, and we do not claim operational superiority in deployment settings. Responsible use of MSG-TCE in practice would require independent fairness audits across demographic groups, transparent documentation of data provenance and known reporting biases, human oversight of any resource-allocation decision, and consultation with the communities affected. We therefore position MSG-TCE as a methodological contribution to spatiotemporal modelling rather than as a deployment-ready policing tool.

The experimental results demonstrate that MSG-TCE's integrated approach yields consistent improvements in prediction accuracy, particularly for longer time horizons where traditional methods often fail. The ablation studies confirm that each component contributes meaningfully to overall performance, with the hyperbolic spatial representation showing particularly strong impact on model effectiveness. These technical advancements are complemented by the framework's ability to provide interpretable insights into spatiotemporal crime patterns, offering practical value for urban safety planning and resource allocation.

The limitations identified in our discussion, including computational scalability, dynamic spatial hierarchies, and ethical considerations, point to important directions for future research. Addressing these challenges will require continued innovation in both algorithmic design and interdisciplinary collaboration with domain experts in criminology and urban planning. The modular architecture of MSG-TCE provides a flexible foundation for such extensions, enabling the incorporation of fairness constraints, dynamic graph learning, and other enhancements while maintaining the core strengths of the approach.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijgi15070299/s1>. Table S1: complete hyperparameter search ranges and final selected values for MSG-TCE and all baseline models. Figure S1: Predicted and observed crime-risk surfaces for Los Angeles and New York City. Figure S2: Poincaré-disk projection of the learned HSP node embeddings for Chicago. Figure S3: PTE attention-weight heat maps for representative Chicago cells.

Author Contributions: Conceptualisation, Rosny Jean and Stabak Roy; methodology, Rosny Jean; software, Rosny Jean; validation, Rosny Jean and Stabak Roy; formal analysis, Rosny Jean; investigation, Rosny Jean and Stabak Roy; resources, Rosny Jean; data curation, Rosny Jean; writing—original draft preparation, Rosny Jean; writing—review and editing, Rosny Jean and Stabak Roy; visualisation, Rosny Jean; supervision, Stabak Roy; project administration, Rosny Jean. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The crime datasets analysed in this study are publicly available. The Chicago crime data, Los Angeles crime data, and New York City crime data used for evaluation can be accessed through their respective municipal open-data portals (see references [35–37]). Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors thank the editors and anonymous reviewers for their valuable comments and suggestions that helped improve this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Jenga, K.; Catal, C.; Kar, G. Machine learning in crime prediction. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 2887–2913. [\[CrossRef\]](#)
- Sampson, R.J.; Raudenbush, S.W.; Earls, F. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **1997**, *277*, 918–924. [\[CrossRef\]](#) [\[PubMed\]](#)
- Weisburd, D.; Groff, E.R.; Yang, S.-M. *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem*; Oxford University Press: New York, NY, USA, 2012.
- Brantingham, P.J.; Brantingham, P.L. *Patterns in Crime*; Macmillan: New York, NY, USA, 1984.
- Michal'k, D. Suburbanisation and the changing spatial structure of the urban fringe: A case study from the Bratislava region. *Folia Geogr.* **2022**, *64*, 90–111.
- Matlovičová, K.; Mocák, P.; Kolesárová, J. Environment of estates and crime: Crime prevention through urban environment formation and modification. *Geogr. Pannonica* **2016**, *20*, 168–180.
- Saville, G.; Cleveland, G. 2nd Generation CPTED: An antidote to the social Y2K virus of urban design. In Proceedings of the 2nd Annual International CPTED Association Conference, Orlando, FL, USA, 2–5 December 1997.
- Sala, F.; De Sa, C.; Gu, A.; Ré, C. Representation tradeoffs for hyperbolic embeddings. In Proceedings of the 35 International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018.
- Bloomfield, P. *Fourier Analysis of Time Series: An Introduction*; John Wiley & Sons: Hoboken, NJ, USA, 2004.
- Chiann, C.; Morettin, P. A wavelet analysis for time series. *J. Nonparametr. Stat.* **1998**, *10*, 1–46. [\[CrossRef\]](#)
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
- Chen, P.; Yuan, H.; Shu, X. Forecasting crime using the ARIMA model. In Proceedings of the 2008 Fifth International Conference on Intelligent Systems Design and Applications, Kaohsiung, Taiwan, 26–28 November 2008.
- Muthamizharasan, M.; Ponnusamy, R. Forecasting crime event rate with a CNN-LSTM model. In *Innovative Data Communication Technologies and Application*; Springer: Singapore, 2022.
- Maltoni, D. Pattern recognition by hierarchical temporal memory. *SSRN Electron. J.* **2011**. [\[CrossRef\]](#)
- Barbarioli, B.; Mersy, G.; Sintos, S.; Krishnan, S. Hierarchical residual encoding for multiresolution time series compression. *Proc. ACM Manag. Data* **2023**, *1*, 99. [\[CrossRef\]](#)
- Breetzke, G. Examining the spatial periodicity of crime in South Africa using Fourier analysis. *S. Afr. Geogr. J.* **2016**, *98*, 275–288.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017.
- Lin, H.; Tseng, V. Periodic transformer encoder for multi-horizon travel time prediction. *Electronics* **2024**, *13*, 2094. [\[CrossRef\]](#)
- Janowicz, K.; Gao, S.; McKenzie, G.; Hu, Y.; Bhaduri, B. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 625–636.
- Hu, Y.; Wang, F.; Guin, C.; Zhu, H. A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. *Appl. Geogr.* **2018**, *99*, 89–97. [\[CrossRef\]](#)
- Anselin, L.; Cohen, J.; Cook, D.; Gorr, W.; Tita, G. Spatial analyses of crime. *Crim. Justice* **2000**, *4*, 213–262.
- Min, S.; Gao, Z.; Peng, J.; Wang, L.; Qin, K.; Fang, B. STGSN—A spatial-temporal graph neural network framework for time-evolving social networks. *Knowl.-Based Syst.* **2021**, *214*, 106746. [\[CrossRef\]](#)
- Yang, M.; Zhou, M.; Kalandar, M.; Huang, Z.; King, I. Discrete-time temporal network embedding via implicit hierarchical learning in hyperbolic space. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, New York, NY, USA, 14–18 August 2021.
- Yu, C.; Xie, X.; Xiao, Z.; Wu, Y.; Zhang, J. Crime prediction using spatial-temporal synchronous graph convolutional networks. In Proceedings of the IEEE International Conference on Soft Computing and Machine Intelligence, Melbourne, Australia, 22–23 November 2024.
- Cohen, L.E.; Felson, M. Social change and crime rate trends: A routine activity approach. *Am. Sociol. Rev.* **1979**, *44*, 588–608. [\[CrossRef\]](#)
- Hossain, S.; Abtahee, A.; Kashem, I.; Hoque, M. Crime prediction using spatio-temporal data. *arXiv* **2020**, arXiv:2003.09322.
- Wu, Z.; Liu, F.; Han, J.; Liang, Y.; Liu, H. Spatial-temporal mixture-of-graph-experts for multi-type crime prediction. *arXiv* **2024**, arXiv:2409.15764.
- Fan, Y.; Hu, X.; Hu, J. Research on a crime spatiotemporal prediction method integrating Informer and ST-GCN: A case study of four crime types in Chicago. *Big Data Cogn. Comput.* **2025**, *9*, 179. [\[CrossRef\]](#)
- Perry, W.L.; McInnis, B.; Price, C.C.; Smith, S.C.; Hollywood, J.S. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*; RAND Corporation: Santa Monica, CA, USA, 2013.
- Mohler, G.O.; Short, M.B.; Malinowski, S.; Johnson, M.; Tita, G.E.; Bertozzi, A.L.; Brantingham, P.J. Randomized controlled field trials of predictive policing. *J. Am. Stat. Assoc.* **2015**, *110*, 1399–1411. [\[CrossRef\]](#)

31. Lum, K.; Isaac, W. To predict and serve? *Significance* **2016**, *13*, 14–19. [[CrossRef](#)]
32. Richardson, R.; Schultz, J.; Crawford, K. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *N. Y. Univ. Law Rev.* **2019**, *94*, 192–233.
33. Berk, R.; Heidari, H.; Jabbari, S.; Kearns, M.; Roth, A. Fairness in criminal justice risk assessments: The state of the art. *Sociol. Methods Res.* **2021**, *50*, 3–44.
34. Rani, A.; Rajasree, S. Crime trend analysis and prediction using Mahalanobis distance and dynamic time warping technique. *Int. J. Comput. Sci. Inf. Technol.* **2014**, *5*, 4131–4135.
35. Garima, A.; Alaiad, A. Crime analysis in Chicago city. In Proceedings of the 2019 10th International Conference on Big Data Analytics (ICBDA), Irbid, Jordan, 11–13 June 2019.
36. Hughley, C. Crime in Los Angeles. In Proceedings of the 27th Annual Symposium of Student Scholars, Kennesaw, GA, USA, 18–21 April 2023.
37. Gaurav, F. Temporal and spatial analysis of crime patterns in New York City: A statistical investigation of NYPD complaint data (1963–2025). *arXiv* **2025**, arXiv:2511.14789.
38. Zhang, J.; Zheng, Y.; Qi, D. Deep spatio-temporal residual networks for citywide crowd flows prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
39. Roshankar, R.; Keyvanpour, M. Spatio-temporal graph neural networks for accurate crime prediction. In Proceedings of the 2023 13th International Conference on Machine Learning and Applications, Jacksonville, FL, USA, 15–17 December 2023.
40. Xia, L.; Huang, C.; Xu, Y.; Dai, P.; Bo, L.; Zhang, X. Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning. *arXiv* **2022**, arXiv:2201.02435.
41. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3848–3858. [[CrossRef](#)]
42. Pan, Z.; Liang, Y.; Zhang, J.; Yi, X.; Yu, Y.; Zheng, Y. HyperST-Net: Hypernetworks for spatio-temporal forecasting. *arXiv* **2018**, arXiv:1809.10889.
43. Joseph, J. Predicting crime or perpetuating bias? The AI dilemma. *AI Soc.* **2025**, *40*, 2319–2321.
44. Rostami-Tabar, B.; Hyndman, R. Hierarchical time series forecasting in emergency medical services. *J. Serv. Res.* **2025**, *28*, 278–295.
45. Ding, F.; Zhu, Y.; Yin, Q.; Cai, Y.; Zhang, D. MS-ResCnet: A combined spatiotemporal modeling and multi-scale fusion network for taxi demand prediction. *Comput. Electr. Eng.* **2023**, *105*, 108558.
46. Dressel, J.; Farid, H. The accuracy, fairness, and limits of predicting recidivism. *Sci. Adv.* **2018**, *4*, eaao5580. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.