



Ada-GCNLSTM: An adaptive urban crime spatiotemporal prediction model

Miaoxuan Shan^a, Chunlin Ye^b, Peng Chen^{a,*}, Shufan Peng^a

^a School of information network security, People's Public Security University of China, Beijing, 100038, China

^b School of Artificial Intelligence and Automation, Hohai University, Nanjing, 211100, China

ARTICLE INFO

Keywords:

Spatiotemporal prediction
Crime prediction
Graph convolutional networks
Deep learning

ABSTRACT

Accurate crime prediction is crucial for the proactive allocation of law enforcement resources and ensuring urban safety. A major challenge in achieving accurate predictions lies in identifying generalized patterns of criminal behavior from spatiotemporal features in crime data. Additionally, the inherent randomness and volatility of crime data at the spatiotemporal level introduce noise, which can mislead prediction models. While many effective spatiotemporal crime prediction methods have been proposed, most overlook this issue, reducing their ability to generalize. In this paper, we introduce a novel deep learning-based model, adaptive-GCNLSTM (Ada-GCNLSTM). Specifically, in the spatial feature extraction module, we enhance the model's ability to capture crime spatial distributions by leveraging graph convolutional networks to model spatial dependencies in conjunction with the maximum mean discrepancy to extract the universal features of crime data. We then incorporate a memory network based on long short-term memory network to capture the underlying relationships between temporal features. Through extensive experiments, our model demonstrates an average improvement of 11.7% in mean absolute error and 2.7% in root mean squared error across the three datasets, outperforming the best baseline model. These results underscore the effectiveness of our approach in enhancing crime prediction accuracy.

1 Introduction

Criminal activities are a persistent issue during social development. The rise in crime not only impacts the quality of life for citizens but also hinders socioeconomic progress [1]. Effective crime prevention not only enhances public safety but also reduces government expenditures. In the age of big data, the advancement of geographic information collection systems enables the accurate recording of crime data across regions. Moreover, the application of machine learning models in various domains offers new opportunities for crime prevention.

Currently, many researchers have researched crime spatiotemporal prediction [2,3]. Early methods for crime prediction based on spatiotemporal data relied heavily on location information and the concept of nearby repetition effects to model and forecast future crime activities [4]. These approaches analyze historical crime data to develop algorithmic models that predict the distribution of future crimes within specific spatiotemporal boundaries. One commonly used method is the kernel density estimation (KDE) model, a nonparametric technique that allows for the visualization of crime distributions. Using KDE, relevant authorities could identify high-incidence areas for targeted surveillance.

Bowers et al. [5] proposed the ProMap model based on KDE, which assessed the likelihood of crime occurring in a region. Xu et al. [6] extended the KDE model by incorporating a time factor to propose a spatiotemporal similarity-based crime prediction model, which was verified as effective. Some researchers have also observed that the spatiotemporal distribution of crimes is similar to that of aftershock phenomena following earthquakes. For example, Mohler et al. [7] adapted the aftershock model from seismology to predict residential burglaries, separating the time and space components of crime incidents and using kernel functions to assess crime likelihood.

While these methods have shown promise in improving crime prediction accuracy, they still face several limitations:

- (1) These methods are effective for assessing crime risks over large spatial areas.
- (2) Simple statistical methods struggle to account for the complexities and fluctuations in crime data across both short and long timeframes.

With the rise of machine learning, more flexible and accurate

* Corresponding author.

E-mail address: chenpeng@ppsc.edu.cn (P. Chen).

predictions are now possible, especially when dealing with complex crime data. Models such as random forest (RF) [8,9], XGBoost [10,11], and the autoregressive integrated moving average model [12] have gained traction in crime spatiotemporal prediction. Liu et al. [9] compared RF with spatiotemporal kernel density mapping in predicting crime hotspots, while Yan et al. [10] proposed a method based on XGBoost to predict various theft crimes. However, these models often struggle to capture intricate temporal changes in crime data. Conversely, deep learning models excel in extracting both short- and long-term features from increasingly complex and diverse datasets, leading to their widespread adoption across various domains [13–16]. For example, Yan and Hou [17] used long short-term memory (LSTM) networks to predict theft crime time series, while Cortez et al. [18] designed a special LSTM architecture for emergency event prediction. Further improvements have been achieved through ensemble models, which combine predictions from multiple algorithms to enhance accuracy. Duan et al. [19] proposed a novel spatiotemporal crime network using deep convolutional neural networks (CNNs) to automatically extract crime-relevant features, which were then fed into recurrent neural networks for prediction. Yi et al. [20] combined an improved continuous conditional random field with LSTM for enhanced crime prediction, whereas Dong et al. [21] refined the ST-3DNet model to achieve finer temporal scales in crime prediction. However, many of these methods fail to adequately capture spatial features within the data.

Recently, graph neural networks (GNNs) have gained widespread attention for their ability to model complex spatiotemporal relationships [22–25], including in crime prediction. Liang et al. [26] used multi-graph convolutional networks (GCNs) to capture spatial dependencies from multiple aspects, combining them with a gated recurrent unit for future crime prediction. Jin et al. [27] proposed an integrated model combining CNNs and adaptive GCNs to capture geographic correlations and latent region-wise dependencies. Despite the promising results of these methods, the complex and variable spatiotemporal patterns of crime data present several challenges that need to be addressed:

- (1) Data-level challenges: Crime data is inherently stochastic, exhibiting random fluctuations and uncertainties over time. This results in noise that can negatively affect the prediction accuracy. Effective strategies to mitigate this noise are crucial.
- (2) Label-level challenges: Multiple crime types can occur within the same time period, and there are often correlations between the quantities of different crime types. A model that can identify these latent relationships is essential for improving prediction performance.

To address these issues, we propose a new model called adaptive-GCNLSTM (Ada-GCNLSTM). Our model aims to extract more generalized features and fully explore the spatiotemporal correlations between different crime types to enhance prediction accuracy. Specifically, the model first employs GCN to extract hidden spatial features of crimes and uses maximum mean discrepancy (MMD) to reduce the distance from a Gaussian distribution. This minimizes the impact of outliers and improves generalizability. Next, LSTM is utilized to capture temporal features, and relational mechanism units (RMUs) are added to the LSTM to identify latent relationships between different crime types, further boosting predictive performance. The main contributions of this paper are as follows:

- (1) We improve GCN with an adaptive MMD penalty term to reduce the impact of noisy data. By measuring the distance between spatial features and a normal distribution, the model can effectively mitigate overfitting risks.
- (2) We introduce RMUs in LSTM to capture the interactive relationships between different temporal features during the propagation process.

- (3) We conduct experiments on real-world crime datasets to demonstrate the effectiveness of our proposed model. The results show that our model outperforms state-of-the-art techniques. We also perform fine-grained and convergence analysis to evaluate its robustness.

The remainder of this paper is organized as follows: Section 2 describes the computational process of the proposed model. Section 3 presents the experimental data and setup. Section 4 details comparative experiments to assess the model's performance. Section 5 summarizes the findings and discusses potential future directions for the prediction model.

2. Methods

This section presents the proposed Ada-GCNLSTM model, including its problem formulation and the details of the prediction framework.

2.1. Task formalization

Crime data is systematically recorded based on spatiotemporal attributes, such as crime type, time, longitude, and latitude. The input data, processed using a sliding window approach, can be represented as a three-dimensional tensor $X \in \mathbb{R}^{R \times T \times C}$, where R denotes the number of regions (indexed by r), T represents the length of the time window (indexed by t), and C corresponds to the number of crime types (indexed by c). Each element $x \in \mathbb{R}^{r \times t \times c}$ in X indicates the count of crime type c in region r during time window t . The task can be formulated as follows: Given the input $X \in \mathbb{R}^{R \times T \times C}$, the goal is to predict the counts of different crime types in each region at time $T + 1$. Thus, the output is $Y \in \mathbb{R}^{R \times C}$.

2.2. GCN

GCNs are a type of CNN that directly operates on graph structures, leveraging the inherent relationships between nodes for effective feature extraction from node data [28]. Compared to traditional statistical methods, GCNs are better equipped to capture complex data patterns and dependencies [29]. For this reason, we chose GCN to extract spatial features and explore underlying patterns in crime data. After processing through GCN, the graph data is transformed into a tensor. The computational process is outlined in Fig. 1, with the following formulas:

$$\tilde{A} = A + I, \quad (1)$$

$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}, \quad (2)$$

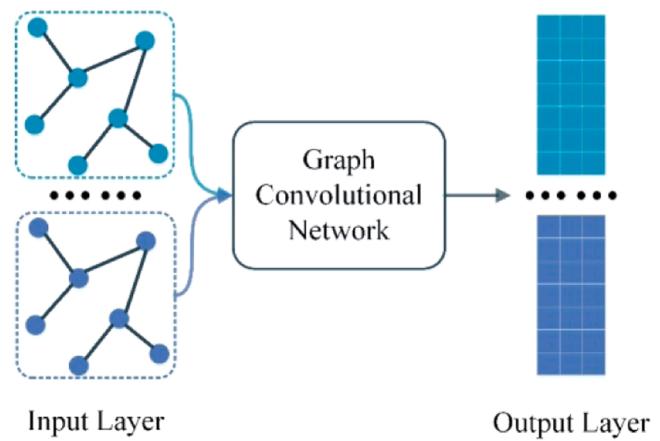


Fig. 1. GCN computation process.

$$H^{l+1} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l \omega^l \right). \quad (3)$$

where H^l denotes the input state at layer l , σ is the activation function (typically the sigmoid function), A represents the adjacency matrix that encodes relationships between nodes, ω^l represents the learnable weight parameters at layer l , and D is the diagonal degree matrix (often used in the Laplacian matrix computation).

2.3. Ada-GCNLSTM

The overall architecture of the Ada-GCNLSTM model is illustrated in Fig. 2. The training process is as follows: First, crime data from previous days is aggregated via the sliding window and a grid-based method, creating a matrix with daily time resolution as the model input. Next, the feature extraction module is employed to capture generalizable features, which are then input into the prediction module to generate predictions for the next day. Ada-GCNLSTM consists of two main components: the feature extraction module and the prediction module.

2.3.1. Feature extraction module

Initially, the crime spatiotemporal data is processed using a sliding window to convert it into supervised data. To account for spatial relationships within the crime data, we use GCNs to extract spatial features at a fine-grained level. The output of the feature extraction module, denoted as H_G , is given by:

$$H_G = GCN(X, A). \quad (4)$$

where X denotes the input data.

2.3.2. Prediction module

To model the spatiotemporal dynamics and the interdependence between crime types and geographical areas, we introduce a relational mechanism to update and calculate relationships within the LSTM framework. The full computational process is shown in Fig. 3.

First, we describe the LSTM's computational process.

$$F^i = \sigma(W_F \cdot [\hat{A}^{i-1}, H_G^i] + b_F), \quad (5)$$

$$I^i = \sigma(W_I \cdot [\hat{A}^{i-1}, H_G^i] + b_I), \quad (6)$$

$$\tilde{C}^i = \tanh(W_C \cdot [\hat{A}^{i-1}, H_G^i] + b_C), \quad (7)$$

$$C^i = F^i \times \tilde{C}^{i-1} + I^i \times \tilde{C}^i, \quad (8)$$

$$O^i = \sigma(W_O \cdot [\hat{A}^{i-1}, H_G^i] + b_O), \quad (9)$$

where H_G^i is the i -th element of H_G ; W and b denote the weight matrix and bias vector, respectively; and F , I , O , and C represent the forget gate, input gate, output gate, and long-term memory, respectively. \hat{A} is the short-term memory in the LSTM.

To improve the model's ability to capture the potential spatiotemporal relationships during the update process, we incorporate an RMU within the LSTM memory module. The introduction of matrices W_C and W_h , along with bias terms b_C and b_a , enhances the model's capacity to capture nuanced dependencies. Additionally, a nonlinear activation function σ is introduced to model complex relationships. The approach allows the LSTM to consider both long- and short-term temporal dependencies in the hidden features. The RMU formulas are as follows:

$$\hat{C}^i = \sigma(W_C \cdot C^i + b_C), \quad (10)$$

$$A^i = O^i \times \tanh(C^i), \quad (11)$$

$$\hat{A}^i = \sigma(W_h \cdot A^i + b_a). \quad (12)$$

Finally, the model produces the output result Y .

2.3.3. Loss function

Inspired by the work in [30], we incorporate an adaptive MMD loss term after GCN feature extraction to prevent spatial features from being influenced by noise and outliers. This approach ensures that the feature distribution during training is closer to a normal distribution, thus reducing the risk of overfitting. The MMD loss function is defined as follows:

$$MMD(H_G, P) = \left\| \frac{1}{h} \sum_{i=1}^h \varphi(H_G^i) - \frac{1}{h} \sum_{i=1}^h \varphi(P_i) \right\|_D^2, \quad (13)$$

where H_G represents the output of the GCN, P represents a normal distribution, and D denotes the distance between distributions, mapped by $\varphi(\cdot)$ into the reproducing kernel Hilbert space for metric computation. Batch size is commonly used in neural network computations, which makes the numbers of H_G and P equivalent. h represents the number of elements in H_G and P . Expanding the formula:

$$\begin{aligned} MMD(H_G, P) = & \left\| \frac{1}{h^2} \sum_{i=1}^h \sum_{j=1}^h k(H_G^i, H_G^j) - \frac{2}{h^2} \sum_{i=1}^h \sum_{j=1}^h k(H_G^i, P_j) + \frac{1}{h^2} \sum_{i=1}^h \right. \\ & \left. \times \sum_{j=1}^h k(P_i, P_j) \right\|. \end{aligned} \quad (14)$$

The Gaussian kernel function $k(u, v)$ is defined as follows:

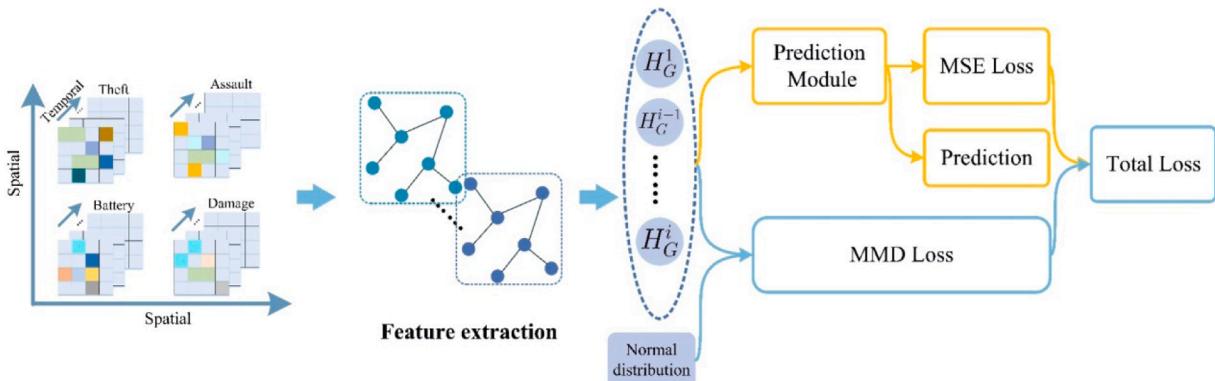


Fig. 2. Overall architecture of the proposed Ada-GCNLSTM model.

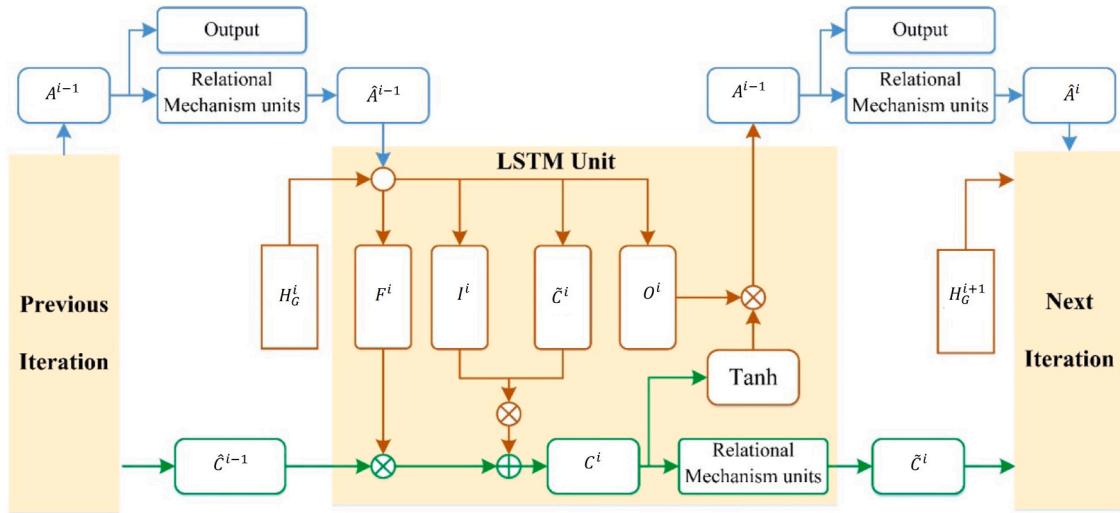


Fig. 3. Architecture of the prediction module.

$$k(u, v) = \exp\left(-\frac{\|u - v\|^2}{2 \cdot \text{bandwidth}^2}\right). \quad (15)$$

Typically, bandwidth is a constant. However, to achieve adaptive measurement of the differences between distributions and enhance the model's generalization capabilities, the specific calculation formula is as follows:

$$\text{bandwidth} = \frac{\sum_{i=1}^h \sum_{j=1}^h \|H_G^i - H_G^j\|^2 + \sum_{i=1}^h \sum_{j=1}^h \|H_G^i - P_j\|^2 + \sum_{i=1}^h \sum_{j=1}^h \|P_i - P_j\|^2}{h \cdot h + 2 \cdot h \cdot (h - 1)}. \quad (16)$$

Finally, the total loss function is defined as follows:

$$\text{Total loss} = L_{\text{MSE}} + \lambda \cdot L_{\text{MMD}} = L(Y, \hat{Y}) + \lambda \cdot \text{MMD}(H_G, P). \quad (17)$$

where Y represents the predicted output, \hat{Y} represents the true label, L denotes MSE loss, and $\lambda \in (0, 1)$ is the regularization term.

3. Experiments

This section describes the comparative experiment designed to validate the proposed method, including details on the crime datasets, evaluation metrics, baseline models, and experimental setup.

3.1. Crime datasets

To validate the effectiveness of the proposed method, we conducted experiments using real crime report data collected from New York City (NYC), Chicago (CHI), and a county-level city in China (CN-County)

over a 24-month period. NYC and Chicago, two of the most populous cities in the USA with complex social structures, provide a robust testbed for evaluating the generalizability of our prediction model. In addition, we included CN-County, a county-level city in eastern China, to assess the model's generalization ability in a lower-complexity region. CN-County, located in Zhejiang Province, has an area of 668 km² and a population of over 0.714 million registered residents. While the NYC

and CHI datasets were obtained from publicly accessible government websites, the CN-County data was provided by the local public security bureau in China and remains proprietary due to privacy considerations. The crime data statistics are summarized in Table 1.

To effectively represent spatial information and mitigate the sparsity issues caused by overly small grid sizes, we used a grid-based approach to divide NYC, CHI, and CN-County into 3 km × 3 km grid cells [31, 32]. This approach ensures the effective capture of spatial information while preventing data sparsity that can arise from using smaller grid units. The process followed these steps. First, we identified the minimum longitude and latitude values of the study area and used 3 km as the grid unit until the maximum longitude and latitude values were included. Next, we aggregated the crime data by day, counting the number of crimes within each grid cell on a daily basis. Areas that did not intersect with the grip map were removed from the dataset. Finally, the two-dimensional matrix, representing the number of crimes in each grid cell on a given day, was used as the input for the training set.

For the specific grid divisions, the NYC dataset was divided into 16 × 16 grid cells, the CHI dataset into 14 × 12 grid cells, and the CN-

Table 1
Statistics of the crime datasets.

NYC Crimes			CHI Crimes			CN-County Crimes		
Time range	Category	Cases	Time range	Category	Cases	Time range	Category	Cases
2014.1–2015.12	Burglary	31,799	2016.1–2017.12	Theft	124,630	2014.1–2015.12	Theft	10,599
	Robbery	33,453		Battery	99,389		Battery	453
	Assault	40,429		Damage	59,886		Robbery	55
	Larceny	85,899		Assault	37,972		-	-

County dataset into 6×7 grid cells. Crime data within each grid cell was recorded on a daily basis. The datasets were split in chronological order into a training set, a validation set, and a test set at a 7:2:1 ratio [31,32].

3.2. Comparison models

GNNs have become state-of-the-art models for spatiotemporal prediction, effectively capturing spatial dependencies and the flow of information between regions [31]. To validate the effectiveness of our proposed model, we compared it with several widely used spatiotemporal prediction models.

- EvolveGCN [33]: A model based on GCNs that can effectively capture the dynamic and temporal properties of data, spatiotemporal relationships, and node dependencies.
- TGCN [34]: A model designed to handle both linear and nonlinear time series data, learn node attributes and topological structures, and is capable of processing large-scale spatiotemporal data.
- STGCN [35]: This model extends graph convolution from spatial features to the temporal domain, allowing it to explore motion characteristics better and learn differentiated features of various nodes.
- AGCRN [36]: An attention-based spatiotemporal GCN model that captures fine-grained spatiotemporal correlations for specific nodes.
- DCRNN [37]: A model based on diffusion convolution and sequence-to-sequence learning, which is widely applied to various spatiotemporal prediction tasks.

The parameter settings for the models are as follows:

- For EvolveGCN, the hidden channels are set to 16.
- For TGCN, the output channels are set to 4.
- For STGCN, the hidden channels are set to 16.
- For AGCRN, the embedding dimension is set to 10, and the output channels are set to 4.
- For DCRNN, the filter size is set to 2, and the output channels are set to 4.
- For both Ada-GCNLSTM and GConvLSTM, the GCN and LSTM layers are set to 2.

The batch size is selected from {12, 24, 48, 64}, and the number of epochs for each training iteration is set to 100. The regularization term λ for Ada-GCNLSTM ranges from {0.0, 1.0}. All models use the Adam optimizer with a learning rate of 0.01. To ensure a fair comparison, all methods run 10 times, and the best-performing result was selected for evaluation during the testing phase.

3.3. Evaluation metrics

To evaluate the prediction accuracy of the proposed model, we used two common metrics: mean absolute error (MAE) and root mean squared error (RMSE).

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (18)$$

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (19)$$

where $y = \{y_1, \dots, y_i, \dots, y_n\}$ represents the true values, $\hat{y} = \{\hat{y}_1, \dots, \hat{y}_i, \dots, \hat{y}_n\}$ represents the predicted values, y_i is the true value for the i -th

Table 2
Performance of urban crime prediction on the NYC dataset (MAE and RMSE).

Model	NYC Data							
	Burglary		Larceny		Robbery		Assault	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
EvolveGCN	0.216	0.479	0.341	0.818	0.218	0.503	0.228	0.527
TGCN	0.206	0.475	<u>0.329</u>	0.810	<u>0.207</u>	0.496	<u>0.219</u>	0.522
STGCN	0.249	0.506	0.513	1.184	0.253	0.555	0.301	0.598
AGCRN	0.368	0.573	0.607	1.459	0.405	0.655	0.416	0.678
DCRNN	0.251	0.481	0.377	0.822	0.252	0.505	0.270	0.538
GConvLSTM	0.274	<u>0.472</u>	0.390	<u>0.765</u>	0.265	<u>0.487</u>	0.285	<u>0.508</u>
Our model	0.190	<u>0.446</u>	0.309	<u>0.749</u>	0.183	<u>0.463</u>	0.205	<u>0.486</u>
Model	CHI Data							
	Theft		Battery		Assault		Damage	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
EvolveGCN	0.656	<u>1.328</u>	0.561	1.013	0.372	0.606	0.458	0.807
TGCN	0.660	<u>1.460</u>	<u>0.540</u>	1.005	0.352	0.607	<u>0.440</u>	0.807
STGCN	0.896	1.935	0.688	1.219	0.360	0.635	0.467	0.855
AGCRN	1.048	2.292	0.882	1.307	0.696	0.952	0.736	1.051
DCRNN	<u>0.648</u>	1.348	0.542	1.045	<u>0.351</u>	0.625	<u>0.440</u>	0.818
GConvLSTM	0.650	1.354	0.569	<u>0.938</u>	0.374	<u>0.559</u>	0.476	<u>0.747</u>
Our model	0.545	<u>1.221</u>	0.455	<u>0.917</u>	0.271	<u>0.547</u>	0.366	<u>0.738</u>
Model	CN-County Data							
	Theft		Battery		Robbery			
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
EvolveGCN	0.435	0.870	<u>0.133</u>		0.199	<u>0.121</u>	0.142	
TGCN	0.435	<u>0.856</u>	0.135		0.196	0.124	0.142	
STGCN	0.471	1.016	0.170		0.173	0.136	0.149	
AGCRN	<u>0.418</u>	1.087	0.193		<u>0.134</u>	0.190	0.115	
DCRNN	0.447	0.886	0.134		0.197	0.123	0.141	
GConvLSTM	0.473	0.919	0.188		0.151	0.173	0.195	
Our model	0.401	<u>0.822</u>	0.114		<u>0.144</u>	0.101	<u>0.107</u>	

Note: The bold and underlined values indicate the best and second-best results, respectively

sample, \hat{y}_i is the predicted value for the i -th sample, and n is the total number of samples. Smaller values of MAE and RMSE indicate better prediction performance by the model.

4. Results and discussion

This section is structured as follows: First, we compare the prediction results across all the models to assess the performance of the proposed model. Next, we analyze the errors in different regions. We then examine the convergence behavior of the prediction model during training. Finally, we discuss the sensitivity of the model to hyperparameters and present an ablation analysis.

4.1. Model performance

Table 2 presents the predictive performance of our proposed model, Ada-GCNLSTM, alongside other baseline models, on crime data from NYC, CHI, and CN-County. For the NYC dataset, Ada-GCNLSTM achieves an average improvement of 8% in MAE and 4% in RMSE across the four crime types compared to the best-performing baseline models. For the CHI dataset, Ada-GCNLSTM demonstrates an average improvement of 18% in MAE and 3% in RMSE across the four crime types. In the case of the CN-County dataset, Ada-GCNLSTM achieves an average improvement of 9% in MAE across all crime types. While the RMSE improvement is slightly smaller on some datasets, Ada-GCNLSTM consistently demonstrates strong overall performance. These results indicate that Ada-GCNLSTM is effective in adapting to crime prediction tasks, allowing for in-depth exploration of features across different crime types. This, in turn, provides valuable insights for relevant personnel. For larger cities such as NYC and CHI, which have complex social structures, Ada-GCNLSTM successfully captures the intricate spatiotemporal dynamics of crime patterns. Conversely, in a smaller region such as CN-County, where the complexity is lower, Ada-GCNLSTM is capable of mitigating prediction errors caused by the sparsity of crime data, thereby

enhancing the model's overall generalization capability in spatial crime prediction.

To further demonstrate the superiority of our proposed model, we compare its predictive performance with GConvLSTM and other baseline models. **Table 2** shows that Ada-GCNLSTM significantly outperforms TGCN and STGCN in predictive accuracy. While the improvement in RMSE compared to GConvLSTM is modest, Ada-GCNLSTM shows a notable enhancement in MAE across various crime types. Additionally, Ada-GCNLSTM outperforms GConvLSTM in both MAE and RMSE, highlighting its ability to handle noise and skewed distributions in crime spatiotemporal data more effectively than other GNN-based methods.

To visually compare the spatial prediction errors, **Figs. 4, 5**, and **6** illustrate the MAE for Ada-GCNLSTM and GConvLSTM in each region for NYC, CHI, and CN-County. The intensity of the colors increases with higher prediction errors.

From **Fig. 4**, it is evident that our proposed method consistently exhibits significantly lower prediction errors for larceny and robbery compared to GConvLSTM. This indicates that Ada-GCNLSTM is better at capturing the spatial distribution patterns of these crimes in NYC. Because larceny is the most frequent crime type, our model is particularly effective at identifying its underlying patterns, leading to more accurate predictions for this crime category.

Figs. 5 and **6** further validate the advantages of our model in spatiotemporal crime prediction. In contrast, traditional GConvLSTM struggles to capture complex spatial crime patterns due to the dynamic and noisy nature of crime data. However, Ada-GCNLSTM effectively addresses this challenge by leveraging RMUs and MMD-based constraints, which help mitigate the impact of noise and improve prediction accuracy.

Additionally, in the predicted spatial distribution of crimes, certain grid areas exhibit higher error rates. This can be attributed to regional complexities, such as varying population densities, community structures, and economic disparities across different areas. Moreover, in

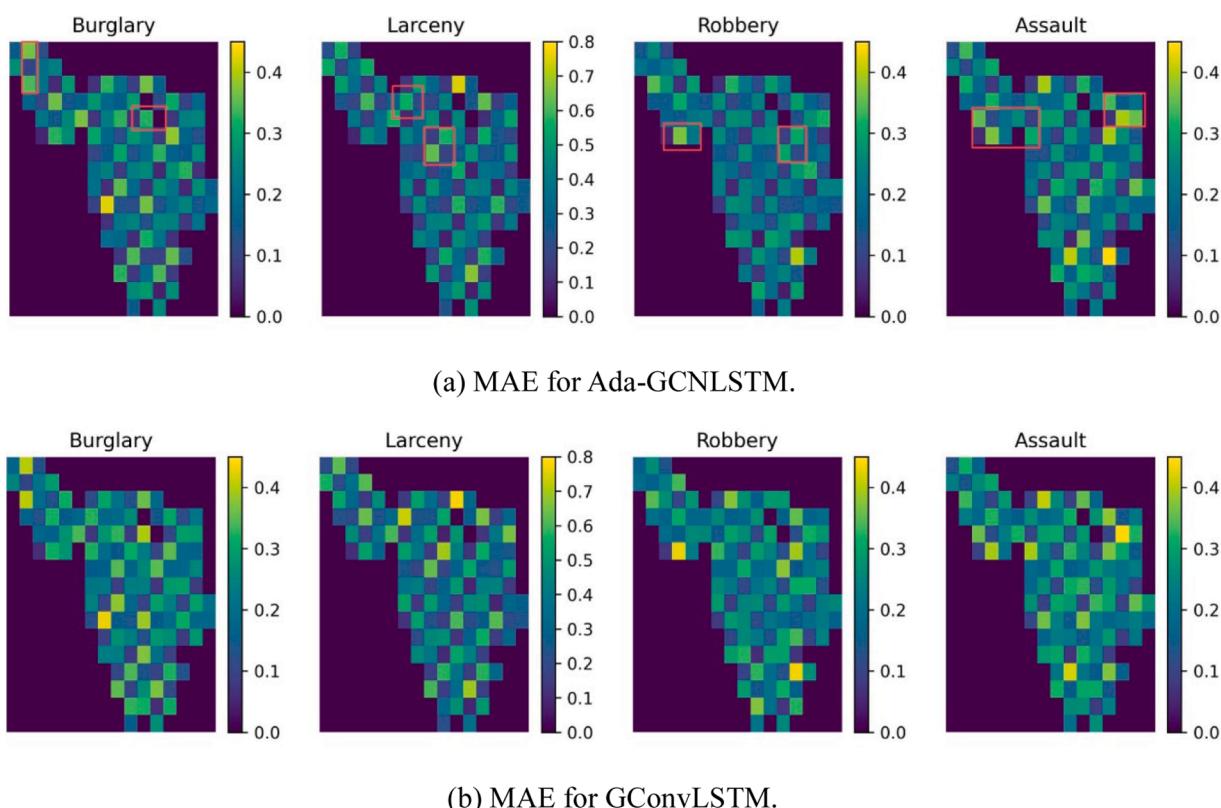
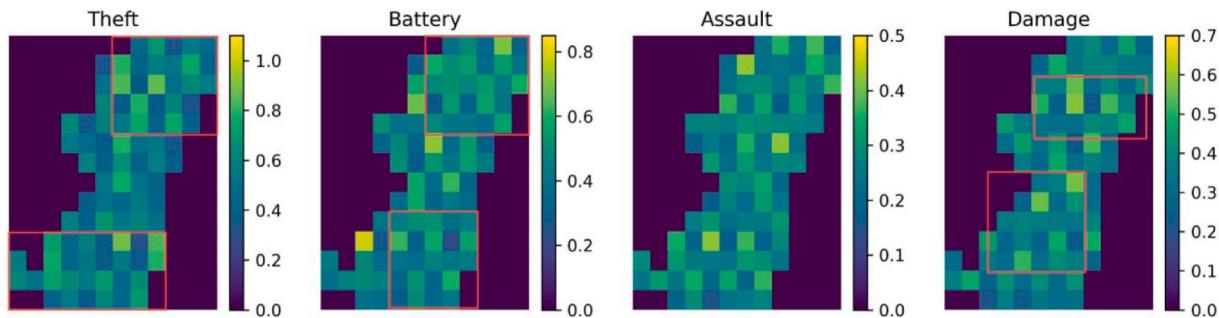
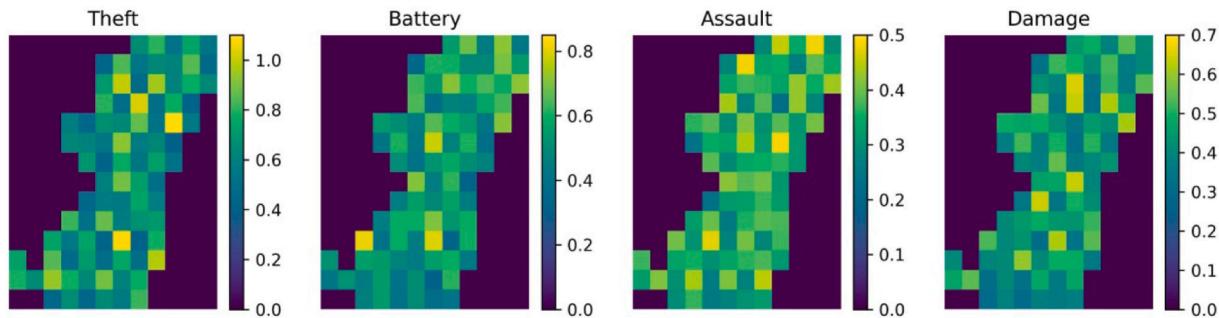


Fig. 4. Visualization of MAE for Ada-GCNLSTM and GConvLSTM for NYC dataset.

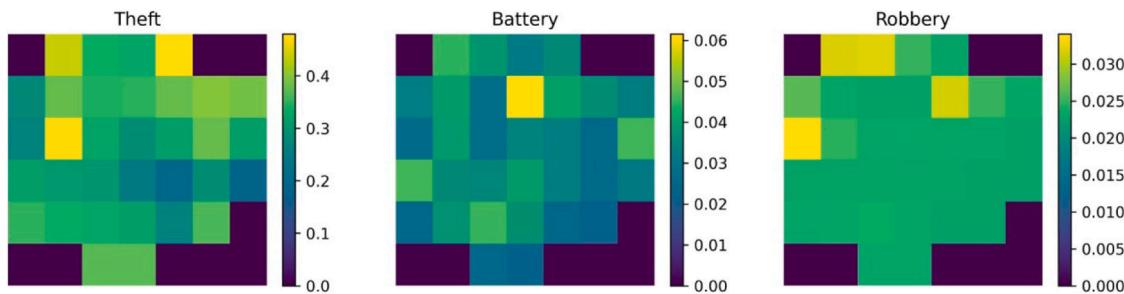


(a) MAE for Ada-GCNLSTM.

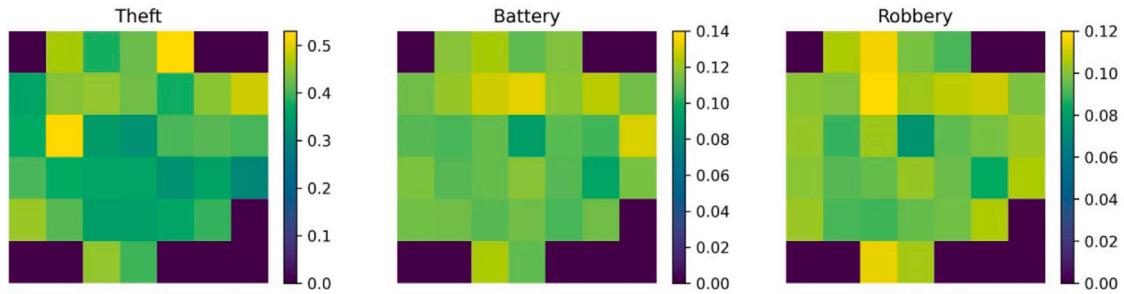


(b) MAE of GConvLSTM.

Fig. 5. Visualization of MAE for Ada-GCNLSTM and GConvLSTM for CHI dataset.



(a) MAE for Ada-GCNLSTM.



(b) MAE for GConvLSTM.

Fig. 6. Visualization of MAE for Ada-GCNLSTM and GConvLSTM for CN-County dataset.

sparserly populated regions with limited crime reporting, the lack of comprehensive data further exacerbates prediction errors, making it more challenging to achieve accurate results.

4.2. Convergence analysis

We analyzed the convergence of the training process by recording both the total loss and MMD loss at each epoch. Fig. 7 illustrates the loss

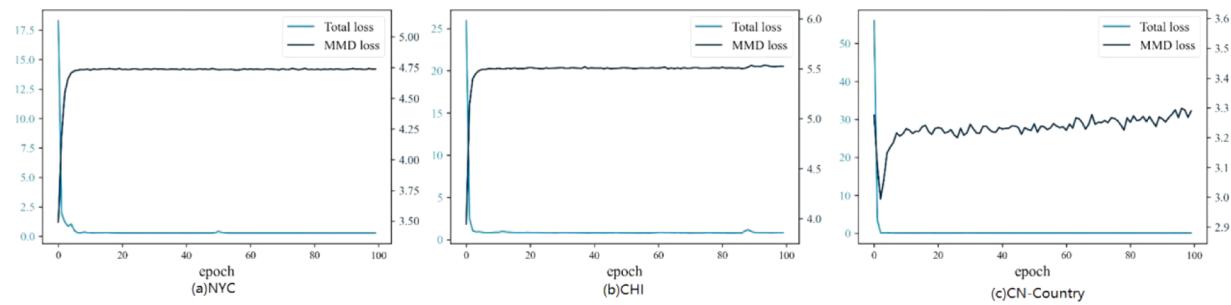
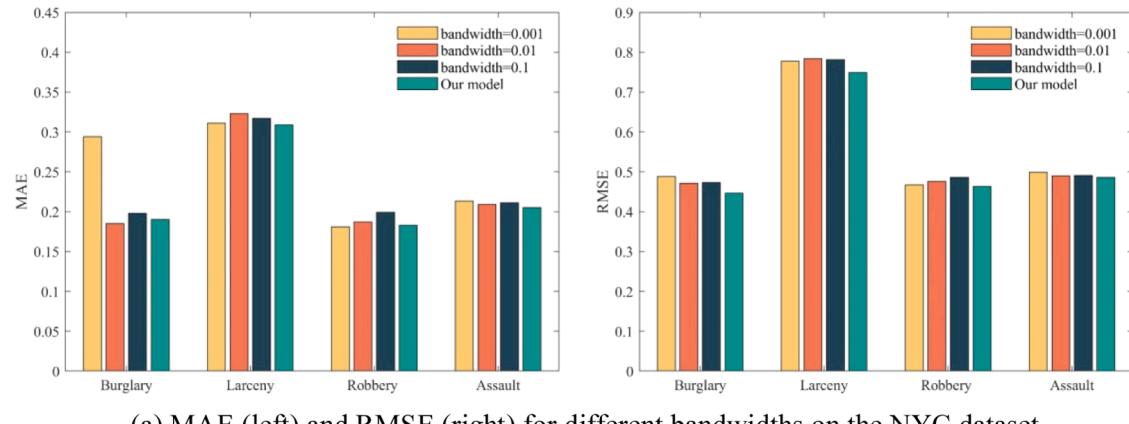
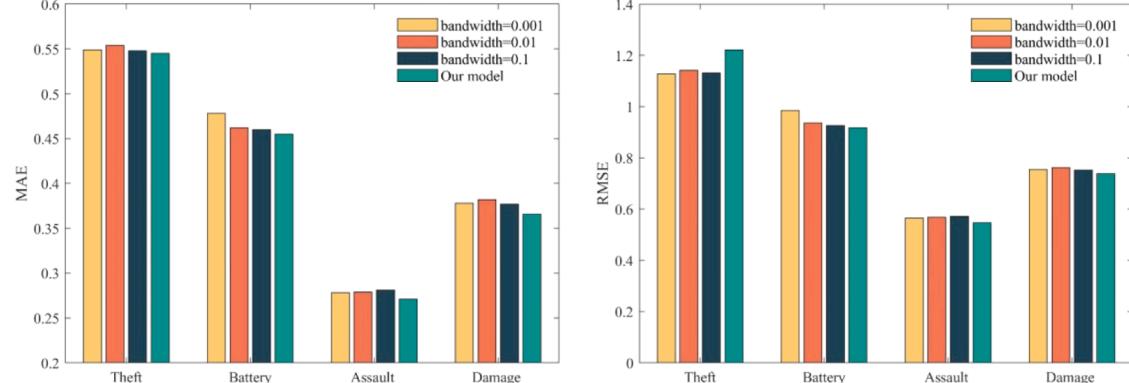


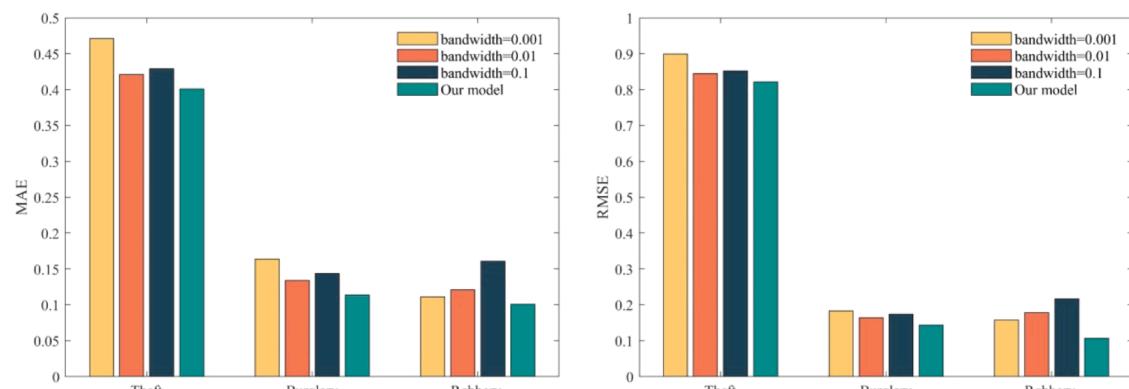
Fig. 7. Convergence analysis of the model.



(a) MAE (left) and RMSE (right) for different bandwidths on the NYC dataset.



(b) MAE (left) and RMSE (right) for different bandwidths on the CHI dataset.



(c) MAE (left) and RMSE (right) for different bandwidths on the CN-County dataset.

Fig. 8. Predictive performance on different datasets with varying MMD functions.

variations for our model across the three crime datasets. As shown, the total loss of our model converges quickly. In Fig. 7(c), while the MMD loss experiences some initial oscillations, it ultimately stabilizes within a specific range. This behavior can be attributed to the smaller spatial scale of the crime data in China, which leads to fluctuations of a certain magnitude.

4.3. Ablation study

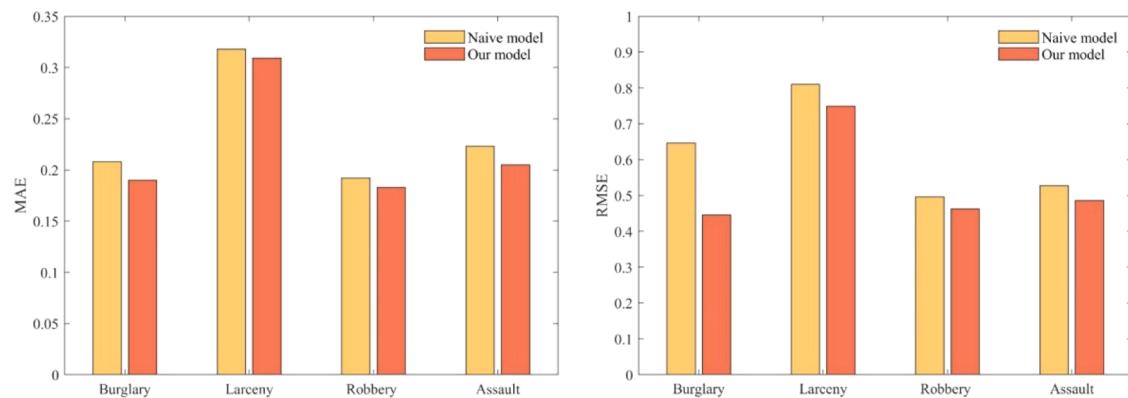
To evaluate the effectiveness of our model using the adaptive MMD distance function, Fig. 8 presents a comparative analysis of predictive performance across three crime datasets. For the naive function, bandwidths of 0.001, 0.01, and 0.1 were considered.

As shown in Fig. 8(a), our proposed model consistently achieves a lower overall RMSE compared to the naive method, with only marginal

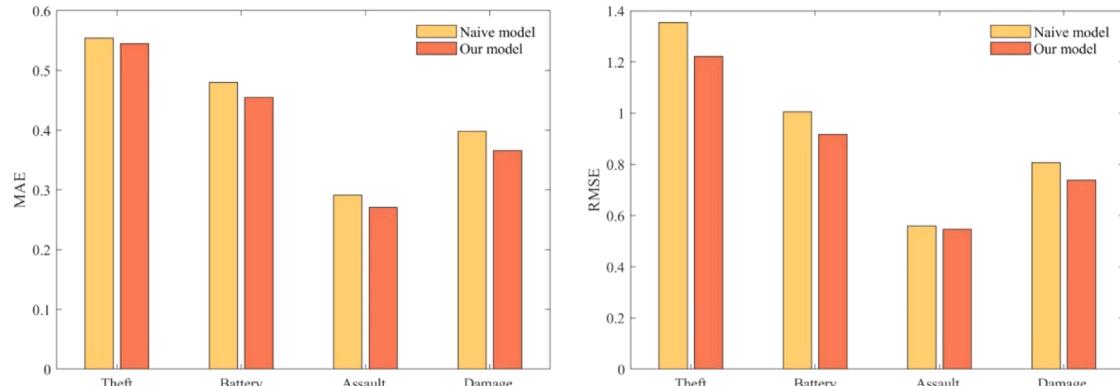
differences in MAE. Fig. 8(b) and (c) further demonstrate that our model achieves slightly superior MAE and lower RMSE. This indicates that our model can dynamically adjust to the distribution differences based on data characteristics, thereby improving the prediction accuracy. In contrast, the traditional fixed bandwidth approach struggles to capture the dynamic changes in crime data, leading to a reduction in predictive performance.

To evaluate the effectiveness of the proposed RMUs in the prediction module, Fig. 9 provides a comparative analysis of predictive performance against the naive LSTM model.

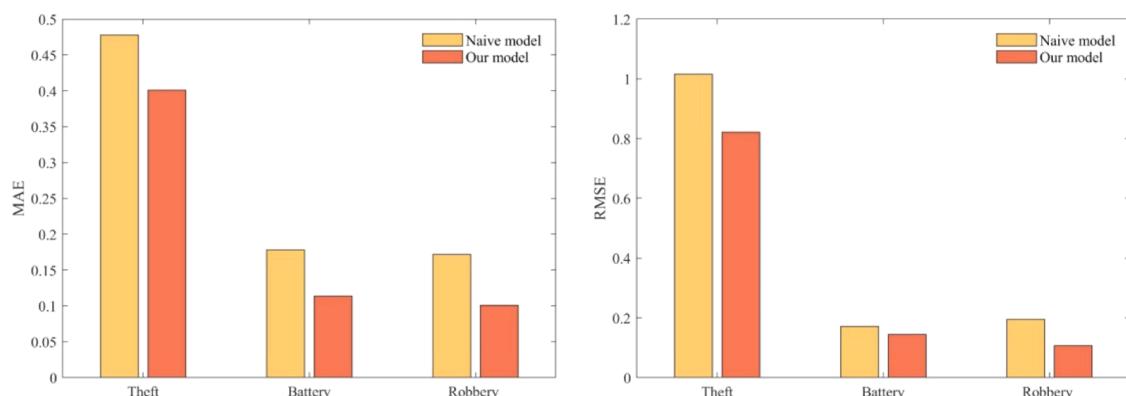
In Fig. 9(a), our method consistently demonstrates lower MAE and RMSE values across various crime types compared to the naive LSTM. In Fig. 9(b), the performance of our model and the naive LSTM is comparable for theft and battery predictions, whereas our model outperforms it in predicting assault and damage. In Fig. 9(c), while our model shows



(a) MAE (left) and RMSE (right) for different prediction modules on the NYC dataset.



(b) MAE (left) and RMSE (right) for different prediction modules on the CHI dataset.



(c) MAE (left) and RMSE (right) for different prediction modules on the CN-County dataset.

Fig. 9. Predictive performance on different datasets with varying prediction modules.

a slightly higher RMSE for theft, it performs better overall than the naive LSTM. These results validate the effectiveness of our approach in improving predictive accuracy. The internal iterations within the LSTM enhance the model's adaptability in handling correlated data, highlighting its improved capabilities for spatiotemporal crime prediction.

5. Conclusion

In this paper, we propose Ada-GCNLSTM for spatiotemporal crime prediction, which is designed to mitigate the noise introduced by the randomness and volatility of crime data across spatiotemporal scales. By incorporating an MMD-based constraint term, we reduce the impact of noise and improve the robustness of the predictions. Additionally, we introduce RMUs to capture the complex relationships between different crime types and geographic regions. The experimental results demonstrate that Ada-GCNLSTM effectively captures spatial relationships in temporal modeling, significantly enhancing prediction accuracy.

Despite its advantages, Ada-GCNLSTM has certain limitations. First, we assume a normal distribution of crime data, which may not fully reflect the complexity of crime patterns in urban environments, especially in areas with extreme crime rates. This assumption could lead to increased prediction errors in such regions. In future work, we aim to analyze crime data distributions more thoroughly to improve prediction accuracy. Another limitation is the use of static grids for spatial analysis. The current approach assumes that a single, static grid adequately captures spatial crime patterns, potentially overlooking finer-grained crime dynamics. Future research will explore dynamic grid systems to better represent spatial crime patterns at varying scales, improving the model's generalizability.

Our research has significant implications for real-world policing and policy-making. By enhancing the accuracy of crime predictions, law enforcement agencies can optimize resource allocation, potentially reducing crime rates and improving public safety.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Statements and declarations

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Availability of data and materials

The datasets of NYC and CHI used in this paper can be downloaded from <https://opendata.cityofnewyork.us>, <https://data.cityofchicago.org>.

CRediT authorship contribution statement

Miaoxuan Shan: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Chunlin Ye:** Software, Data curation. **Peng Chen:** Writing – review & editing, Methodology. **Shufan Peng:** Methodology, Data curation.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements

This work was partially supported by the Top-Notch Innovative Talent Cultivation Project at the People's Public Security University of China (2022yjsky012) and the Fundamental Research Business Funding Project of the People's Public Security University of China (2024JKF04).

Reference

- [1] H. Chen, W. Chung, J.J. Xu, G. Wang, Y. Qin, M. Chau, Crime data mining: a general framework and some examples, *Comput* 37 (4) (2004) 50–56, <https://doi.org/10.1109/MC.2004.1297301>.
- [2] N. Polvi, T. Looman, C. Humphries, K. Pease, The time course of repeat burglary victimization, *Br. J. Criminol.* 31 (4) (1991) 411–414, <https://doi.org/10.1093/oxfordjournals.bjc.a048138>.
- [3] L.W. Sherman, P.R. Gartin, M.E. Buerger, Hot spots of predatory crime: Routine activities and the criminology of place, *Criminol* 27 (1) (1989) 27–56, <https://doi.org/10.1111/j.1745-9125.1989.tb00862.x>.
- [4] G. Farrell, K. Pease, Once bitten, Twice bitten: Repeat victimisation and Its Implications For Crime Prevention, Home Off. Police Res. Group, London, 1993. <https://api.semanticscholar.org/CorpusID:142407021>.
- [5] K.J. Bowers, S.D. Johnson, K. Pease, Prospective hot-spotting: the future of crime mapping, *Br. J. Criminol.* 44 (5) (2004) 641–658, <https://doi.org/10.1093/bjc/azh036>.
- [6] C. Xu, L. Liu, S.H. Zhou, The comparison of predictive accuracy of crime hotspot density maps with the consideration of the near similarity: a case study of robberies at DP peninsula, *Sci. Geogr. Sin.* 36 (1) (2016) 55–62, <https://doi.org/10.13249/j.cnki.sgs.2016.01.007>.
- [7] G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoenberg, G.E. Tita, Self-exciting point process modeling of crime, *J. Am. Stat. Assoc.* 106 (493) (2011) 100–108, <https://doi.org/10.1198/jasa.2011.ap09546>.
- [8] L.G.A. Alves, H.V. Ribeiro, F.A. Rodrigues, Crime prediction through urban metrics and statistical learning, *Physica A* 505 (2018) 435–443, <https://doi.org/10.1016/j.physa.2018.03.084>.
- [9] L. Liu, W.J. Liu, W.W. Liao, H.J. Yu, C. Jiang, R.P. Lin, J.K. Ji, Z. Zhang, Comparison of random forest algorithm and space-time kernel density mapping for crime hotspot prediction, *Prog. Phys. Geogr.* 37 (6) (2018) 761–771, <https://doi.org/10.18306/dlkxjz.2018.06.003>.
- [10] Z. Yan, H. Chen, X. Dong, K. Zhou, Z. Xu, Research on prediction of multi-class theft crimes by an optimized decomposition and fusion method based on XGBoost, *Expert Syst. Appl.* 207 (2022) 117943, <https://doi.org/10.1016/j.eswa.2022.117943>.
- [11] B.Han Wang, B. Patel, C. Rudin, In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction, *J. Quant. Criminol.* 39 (2) (2022) 519–558, <https://doi.org/10.1007/s10940-022-09545-w>.
- [12] Rummens, W. Hardyns, L. Pauwels, The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context, *Appl. Geogr.* 86 (2017) 255–261, <https://doi.org/10.1016/j.apgeog.2017.06.011>.
- [13] T. Shi, J. Fu, X. Hu, TSE-Tran: Prediction method of telecommunication-network fraud crime based on time series representation and transformer, *J. Saf. Sci. Resilience.* 4 (4) (2023) 340–347, <https://doi.org/10.1016/j.jnlssr.2023.07.001>.
- [14] X. Lu, C. Ye, M. Shan, B. Qin, Y. Wang, H. Xing, X. Xie, Z. Liu, The prediction of PM2.5 concentration using transfer learning based on ADGRU, *Water Air Soil Pollut.* 234 (4) (2023) 58, <https://doi.org/10.1007/s11270-023-06271-2>.
- [15] S. Lv, M. Shan, W. Wang, J. Ding, H. Zhang, A rolling-EEMD method for transformer oil level prediction, in: Proceedings of the Annual Conference of China Electrotechnical Society, 2022, pp. 228–237, https://doi.org/10.1007/978-981-99-0408-2_25.
- [16] X. Hu, J. Hu, M. Hou, A two-step machine learning method for casualty prediction under emergencies, *J. Saf. Sci. Resilience.* 3 (3) (2022) 243–251, <https://doi.org/10.1016/j.jnlssr.2022.03.001>.
- [17] J. Yan, M. Hou, Predicting time series of theft crimes based on LSTM network, *Data Anal. Knowl. Discov.* 4 (11) (2020) 84–91, <https://doi.org/10.11925/infotech.2096-3467.2020.0536>.
- [18] B. Cortez, B. Carrera, Y.J. Kim, J.Y. Jung, An architecture for emergency event prediction using LSTM recurrent neural networks, *Expert Syst. Appl.* 97 (2018) 315–324, <https://doi.org/10.1016/j.eswa.2017.12.037>.
- [19] L. Duan, T. Hu, E. Cheng, J. Zhu, C. Gao, Deep convolutional neural networks for spatiotemporal crime prediction, in: Proceedings of the International Conference on Information and Knowledge Engineering (IKE), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2017, pp. 61–67. <https://api.semanticscholar.org/CorpusID:43937446>.
- [20] F. Yi, Z. Yu, F. Zhuang, B. Guo, Neural network based continuous conditional random field for fine-grained crime prediction, in: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2019, pp. 4157–4163. <https://dl.acm.org/doi/10.5555/3367471.3367619>.
- [21] Q. Dong, Y. Li, Z. Zheng, X. Wang, G. Li, ST3DNetCrime: Improved ST-3DNet Model for crime prediction at fine spatial temporal scales, *ISPRS Int. J. Geoinf.* 11 (10) (2022) 529, <https://doi.org/10.3390/ijgi11100529>.
- [22] H. Jiang, P. Cao, M. Xu, J. Yang, O. Zaiane, Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain

- disorders prediction, *Comput. Biol. Med.* 127 (2020) 104096, <https://doi.org/10.1016/j.combiomed.2020.104096>.
- [23] Y. Liu, S. Rasouli, M. Wong, T. Feng, T. Huang, RT-GCN: Gaussian-based spatiotemporal graph convolutional network for robust traffic prediction, *Inf. Fusion.* 102 (2024) 102078, <https://doi.org/10.1016/j.inffus.2023.102078>.
- [24] Z. Li, K. Jiang, S. Qin, Y. Zhong, A. Elofsson, GCSENet: A GCN, CNN and SENet ensemble model for microRNA-disease association prediction, *PLoS Comput. Biol.* 17 (6) (2021) e1009048, <https://doi.org/10.1371/journal.pcbi.1009048>.
- [25] J. Chen, X. Wang, X. Xu, GC-LSTM: Graph convolution embedded LSTM for dynamic network link prediction, *Appl. Intell.* (2022) 1–16, <https://doi.org/10.1007/s10489-021-02518-9>.
- [26] W. Liang, Y. Wang, H. Tao, J. Gao, Towards hour-level crime prediction: A neural attentive framework with spatial-temporal-categorical fusion, *Neurocomputing* 486 (2022) 286–297, <https://doi.org/10.1016/j.neucom.2021.11.052>.
- [27] G. Jin, C. Liu, Z. Xi, H. Sha, Y. Liu, J. Huang, Adaptive dual-view WaveNet for urban spatial-temporal event prediction, *Inf. Sci.* 588 (2022) 315–330, <https://doi.org/10.1016/j.ins.2021.12.085>.
- [28] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu, A comprehensive survey on graph neural networks, *IEEE Trans. Neural Netw. Learn. Syst.* 32 (1) (2020) 4–24, <https://doi.org/10.1109/TNNLS.2020.2978386>.
- [29] Y. Wang, C. Chen, Y. Zhang, M. Zhang, Y. Gao, Deep temporal multi-graph convolutional network for crime prediction, in: *Conceptual Modeling: Proceedings of the 39th International Conference, ER*, Springer International Publishing, Vienna, Austria, 2020, pp. 678–689, https://doi.org/10.1007/978-3-030-62522-1_39. November 3–6, 2020.
- [30] Y. Zhang, Y. Chen, J. Wang, Z. Pan, Unsupervised deep anomaly detection for multi-sensor time-series signals, *IEEE Trans. Knowl. Data Eng.* (2021), <https://doi.org/10.1109/TKDE.2021.3102110>.
- [31] Z. Li, C. Huang, L. Xia, Y. Xu, J. Pei, Spatial-temporal hypergraph self-supervised learning for crime prediction, in: *Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE)*, IEEE, 2022, pp. 2984–2996, <https://doi.org/10.1109/ICDE53745.2022.00269>.
- [32] L. Xia, C. Huang, Y. Xu, P. Dai, L. Bo, X. Zhang, T. Chen, Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning, in: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2021, pp. 1631–1637, <https://doi.org/10.24963/ijcai.2021/225>.
- [33] G. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, C. Leiserson, Evolvegcn: Evolving graph convolutional networks for dynamic graphs, *Proced. AAAI Confer. Art. Intell.* 34 (04) (2020) 5363–5370, <https://doi.org/10.1609/aaai.v34i04.5984>.
- [34] W. Chen, R. Guo, X. Tang, X. Xin, Y. Ding, X.Q. He, D. Wang, TGNC: Tag graph convolutional network for tag-aware recommendation, in: *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, ACM, 2020, pp. 155–164, <https://doi.org/10.1145/3340531.3411927>.
- [35] H. Han, M. Zhang, M. Hou, F. Zhang, Z. Zhang, E. Chen, H. Wang, J. Ma, Q. Liu, STGCN: A spatial-temporal aware graph learning method for POI recommendation, in: *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*, IEEE, 2020, pp. 1052–1057, <https://doi.org/10.1109/ICDM50108.2020.00124>.
- [36] L. Bai, L. Yao, C. Li, X. Wang, C. Wang, Adaptive graph convolutional recurrent network for traffic forecasting, *Adv. Neural Inf. Process. Syst.* 33 (2020) 17804–17815, <https://doi.org/10.48550/arXiv.2007.02842>.
- [37] Y. Li, R. Yu, C. Shahabi, Y. Liu, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in: *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018, <https://doi.org/10.48550/arXiv.1707.01926>.