Expert Systems With Applications 265 (2025) 125940

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

MRAGNN: Refining urban spatio-temporal prediction of crime occurrence with multi-type crime correlation learning

Shun Wang, Yong Zhang *, Xinglin Piao, Xuanqi Lin, Yongli Hu, Baocai Yin

Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Artificial Intelligence Institute, the Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China

ARTICLE INFO

Keywords:
Crime occurrence prediction
Graph neural network
Crime type correlations modeling

ABSTRACT

Predicting whether crime events will occur in different areas (framed as a classification task) is a typical spatio-temporal data mining problem, crucial for both urban management and public safety. Contemporary crime occurrence prediction models predominantly leverage deep learning techniques, focusing on capturing the spatio-temporal dependencies within crime data. Analysis of crime data reveals correlations among different crime types, indicating shared change patterns. Leveraging these correlations among crime types significantly enhances the accuracy of crime occurrence predictions. Nevertheless, existing crime occurrence prediction models frequently overlook the utilization of these type correlations. To solve this problem, we propose a new crime occurrence prediction model with multi-type crime correlation learning: the Multi-type Relations Aware Graph Neural Networks (MRAGNN). The model dynamically constructs a spatial/type graph structure of crime data and employs dynamic graph networks to capture both spatio-temporal and type-temporal dependencies within the data. We introduce a cross-modal gated fusion mechanism to fuse the representations of two dependencies. Furthermore, we develop an improved multi-label classification focal loss to address the challenges posed by the imbalance in crime occurrence data on classification results. Experimental results validate that our model outperforms state-of-the-art (SOTA) methods in crime occurrence prediction.

1. Introduction

Crime occurrence prediction is an important research issue, facilitating government agencies in understanding trends and patterns of criminal activities. Accurate prediction results are a valuable asset for formulating crime prevention and control strategies. Current research primarily focuses on predicting the overall incidence of crime or its likelihood of occurrence. Recent studies have introduced various spatio-temporal prediction models based on deep learning methodologies (Hu et al., 2021; Huang et al., 2018; Sun et al., 2021; Wang et al., 2022b; Yi et al., 2019; Zhao et al., 2022). These models chiefly concentrate on exploring the interplay of spatio-temporal crime patterns across diverse urban regions, employing techniques such as Convolutional Neural Networks (CNN) (Li et al., 2021), Graph Convolutional Networks (GCN) (Wu et al., 2020), and Recurrent Neural Networks (RNN) (Sutskever et al., 2011).

Crime occurrence prediction stands as a crucial research area within social computing (Parameswaran & Whinston, 2007) and urban computing (Zhao & Tang, 2018), garnering considerable attention from researchers. In our study, crime occurrence prediction is defined as a multi-label classification task. The goal is to predict whether different types of crimes will occur in a specific region, with each crime type prediction treated as a binary classification problem (0 for no occurrence, 1 for occurrence). Traditional crime occurrence prediction models often rely on statistical or machine learning methods, which may struggle to capture the intricate, nonlinear attributes of crime data. Deep learning techniques, renowned for their prowess in data modeling, have gained popularity in crime occurrence prediction. Many contemporary studies based on deep learning consider crime prediction as a spatio-temporal data mining challenge, primarily focusing on modeling the spatial and temporal correlations of crime data. For example, Boukabous and Azizi (2022) and Rajapakshe et al. (2019) introduced CNN-RNN-based methodologies for forecasting future crimes. Meanwhile, Li et al. (2022) and Liang et al. (2022) proposed GCN-based methods for this purpose. These approaches often partition the city into grids and establish a spatio-temporal graph to anticipate future criminal activities. Each node in this graph represents a grid, with each edge encapsulating the spatio-temporal relationship between two grids. These efforts primarily focus on spatio-temporal modeling of crime data, highlighting the

* Corresponding author.
E-mail addresses: wangshun@emails.bjut.edu.cn (S. Wang), zhangyong2010@bjut.edu.cn (Y. Zhang), piaoxl@bjut.edu.cn (X. Piao), sophialin@emails.bjut.edu.cn (X. Lin), huyongli@bjut.edu.cn (Y. Hu), ybc@bjut.edu.cn (B. Yin).

https://doi.org/10.1016/j.eswa.2024.125940
Received 21 August 2024; Received in revised form 25 October 2024; Accepted 24 November 2024
Available online 29 November 2024
0957-4174/

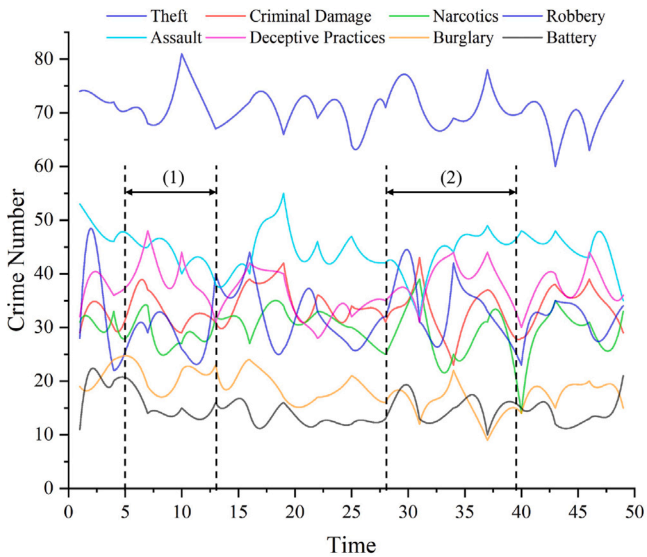


Fig. 1. Number of different types of crimes. The X-axis represents time, with each tick marking a single day. The Y-axis represents the number of crimes. The figure showcases the number of different categories of crimes within the Los Angeles dataset for the first 50 days of 2018. Using the green and yellow lines as an example, these two types of crimes show two different patterns of correlation between the (1) and (2) time periods. The two share a similar trend in time period (1) and the opposite in time period (2).

crucial role of urban big data in enhancing the accuracy of crime occurrence prediction. Therefore, effectively utilizing existing crime data for accurate crime occurrence prediction has become a crucial issue within crime forecasting.

However, as shown in Fig. 1, it is noteworthy that dynamic correlations exist among different crime types, a factor often overlooked by most current prediction models. Crime type correlation refers to the interconnectedness between various categories of crimes, often manifesting as parallel or inverse trends within the data. These correlations provide a more nuanced understanding of crime patterns by revealing data characteristics from a perspective different from the traditional spatio-temporal analyses. Each crime type has its unique characteristics. For instance, burglary primarily centers around property-related offenses, whereas murder and rape involve physical harm to victims. Sociological studies (Quick et al., 2018; Steffensmeier et al., 1989) have highlighted the inherent connections among these crimes. For example, the social disorganization theory (Sampson & Groves, 2017) suggests that the aggregation of minor crimes leads to an increase in more serious offenses. Some endeavors have sought to incorporate crime-type correlations into prediction models. For instance, CCC (Zhao et al., 2022) introduced a prediction framework capturing cross-type and spatio-temporal correlations in crime data. Mohler (2014) presented a marked point process to incorporate both crime types and crime data into crime hotspot maps. However, it did not explicitly leverage crime-type correlation but employed it as constraint information to optimize spatio-temporal prediction models. Consequently, our research emphasizes the integration of these crime-type correlations into crime occurrence prediction models, presenting it as a key innovation in our efforts to comprehend and predict criminal activities more effectively. Despite significant progress in previous studies, two major challenges remain:

Challenge 1: Insufficient utilization of the interrelationships between different types of crime. The intricate correlations between different crime types carry significant implications for the accuracy of crime occurrence prediction. Although existing crime occurrence prediction models forecast the number of crimes occurring or the likelihood of crime events within specific locales, they typically overlook the intrinsic correlations between diverse crime types. Consequently, invaluable information hidden within crime data remains untapped.

Challenge 2: The prediction bias stems from the imbalanced distribution of crime incidents. Crime data distribution exhibits imbalances across both temporal and spatial dimensions. A substantial proportion of crime data records are dominated by zero values, undermining the model's ability to predict future criminal incidents. Mitigating the influence of imbalanced crime data distribution on predictions constitutes a pivotal component of addressing the crime occurrence prediction challenge.

Our research delves deeper into integrating crime-type correlation information into crime occurrence prediction models. Our research aims to model crime-type correlations and rectify the prediction bias resulting from the imbalanced distribution of crime data, ultimately leading to more precise crime occurrence prediction. To tackle the aforementioned challenges in crime forecasting, we introduce the Multi-type Relations Aware Graph Neural Network (MRAGNN) model. This model integrates two graph learning modules, enabling the dynamic learning of spatial and type correlations across different regions and crime types. Building upon traditional spatio-temporal prediction frameworks, we develop two specialized modules: the Dynamic Spatial–Temporal Graph Convolutional Network (DSGCN) and the Dynamic Type-Temporal Graph Convolutional Network (DTGCN). These modules are dedicated to capturing spatial and type dependencies within crime patterns, respectively. Additionally, we devise a Gated Multi-modal Feature Fusion (GMFF) Module to effectively fuse the spatial and type dependencies of crime data. Furthermore, given the prevalence of zero-valued entries in crime datasets, indicating instances of no reported crimes in certain times and areas, we introduce an improved multi-label classification focal loss to address data distribution imbalances. In summary, our study makes four primary contributions:

- We propose a Multi-type Relations Aware Graph Neural Networks (MRAGNN) model consisting of two branches: dynamic spatial–temporal and dynamic type-temporal modeling network. This model can effectively utilize the type correlations of crime data to enhance crime occurrence prediction.
- We design the DSGCN and DTGCN modules to learn dynamic spatial dependencies and dynamic type dependencies. The GMFF module is proposed to fuse these multimodal features to improve the accuracy of crime occurrence prediction.
- We design an improved multi-label classification focal loss to address the imbalanced distribution of crime data, which gives more importance to samples with high crime risk to reduce the bias of model predictions.
- Extensive experiments on crime datasets from two cities show that the MRAGNN model has superior performance to state-of-the-art methods.

The main content of the rest of the paper: Section 2 reviews relevant research about crime occurrence prediction and spatio-temporal prediction in urban computing. Section 3 introduces the background knowledge of using graphs for crime occurrence prediction. Section 4 shows the specific content of the different modules in the MRAGNN model. In Section 5, we evaluate the predictive performance of the proposed model on real-world datasets. Finally, we conclude the paper in Section 6.

2. Related work

Crime occurrence prediction has become one of the hot research issues in urban computing. In the related work section, we investigate different approaches to crime occurrence prediction research and show some recent predictive models. Besides that, we also introduce related research on Spatial–Temporal Prediction in Urban Computing.

2.1. Crime occurrence prediction

In this section, we divide current crime occurrence prediction methods into two categories: traditional predictive models and deep learning-based predictive models. Traditional models mainly use statistical and machine learning methods such as Bayesian learning (Liao et al., 2010), decision trees (Sharma, 2014), and support vector machine (Noble, 2006). These methods accomplish classification, regression, and cluster analysis of crime data. The advantages of traditional models are that they are simple, interpretable, and easy to implement. The disadvantages are that they require manual feature extraction and are challenging to capture the complex nonlinear relationships and spatio-temporal dependencies in crime data.

In recent years, deep learning-based crime occurrence prediction methods have attracted increasing attention because they can effectively handle crime data's high-dimensional, multi-source, and dynamic characteristics, thus improving the accuracy and efficiency of crime occurrence prediction. These methods can be broadly categorized into two main groups: CNN-based approaches (Huang et al., 2019; Quick et al., 2018), which employ spatial and temporal convolutions to capture the spatial and temporal characteristics of crime data; and methods based on Graph Neural Networks (GNN) (Li et al., 2022; Wu et al., 2020; Zhao et al., 2022), which leverage graph convolution and graph attention mechanisms to model the complex network structures in crime data. The main difference between these two categories is that CNN-based methods assume that the spatial relationship between regions is fixed and regular. In contrast, GNN-based methods can model the spatial relationship between regions as a flexible and irregular graph.or example, Boukabous and Azizi (2022) and Rajapakshe et al. (2019) proposed CNN-RNN-based methods for predicting future crimes. Li et al. (2022) and Liang et al. (2022) proposed a GCN-based method for predicting future crimes. These methods divide the city into grids and construct a spatio-temporal graph for predicting future crimes. These methods have achieved better results in predicting crime hotspots in several cities.

While spatial correlations between different regions have been extensively explored in crime occurrence prediction research, the correlations between different types of crimes have not been fully exploited. Existing approaches generally use crime-type correlation as a constraint on the spatial correlation representation, which limits its effective utilization. Therefore, we will further explore the dynamic correlations between different types of crimes to improve the performance of future crime occurrence prediction.

2.2. Spatial–temporal prediction in urban computing

Effective prediction of crime is an important foundation for maintaining urban safety. Crime occurrence prediction is closely related to other typical spatio-temporal prediction tasks in urban computing, including traffic prediction, weather prediction, and social event prediction. Given the benefits of graph networks in representing non-Euclidean data, extensive research has been dedicated to leveraging Graph Neural Networks (GNNs) for modeling spatio-temporal correlations. Auto-DSTSGN (Jin et al., 2022) proposed a graph structure search method to automatically construct spatio-temporal synchronization graphs for different data scenarios. DGCRN (Li et al., 2023) designed a dynamic graph convolutional recurrent network that uses a generative approach to simulate the fine-grained topological structure of dynamic graphs at each time. KFGNN (Wang et al., 2023) proposed multiple graph convolutional networks to fully explore the implicit semantic relationship from traffic knowledge. CLCRN (Lin et al., 2022) formed a conditional local convolutional recurrent network that uses graph convolution to simulate meteorological flows and capture local spatial patterns of weather. These models explore the utilization of GNN in learning the spatial correlation of spatio-temporal data. In spatio-temporal event prediction, data imbalance or zero inflation often

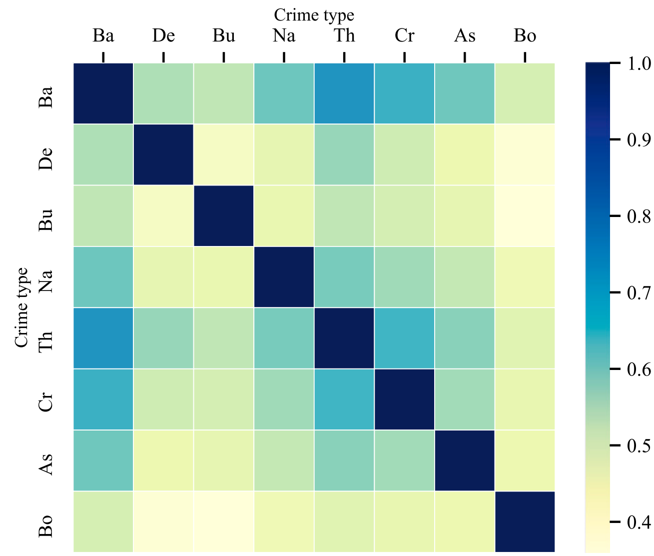


Fig. 2. Heatmap of the correlations between different crime types in Chicago city(2015). The X and Y axes in the heatmap represent eight crime types: Battery(Ba), Deceptive Practices(De), Burglary(Bu), Narcotics(Na), Theft(Th), Criminal Damage(Cr), Assault(As), Robbery(Ro). In this heatmap, darker squares represent a stronger correlation between the two types of crimes.

occurs. That is, there is a large difference in the number of events of different categories, or zero values dominate the spatio-temporal states of data values. Several studies have explored the issue of data imbalance or zero inflation in crime occurrence prediction. NAHC (Liang et al., 2022) adopted a data augmentation strategy based on prior knowledge to address the zero-inflation problem. STZINB-GNN (Zhuang et al., 2022) designed a spatio-temporal zero-inflated negative binomial graph neural network to quantify the uncertainty of sparse travel demand.

Graph-based spatio-temporal prediction models have garnered significant attention in the field of urban computing. In this paper, we leverage graph structures to represent the correlations between different types of crime in crime occurrence prediction.

3. Preliminaries

In this section, we elaborate on the settings for crime occurrence prediction. The task of crime occurrence prediction is to predict the occurrence of future crime events based on historical crime data. In the proposed MRAGNN model, crime occurrence prediction is defined as a classification task. The purpose of the prediction model is to learn the characteristics of crime data from the past period in order to predict the occurrence of different types of crimes in the future. Below are some of the main definitions used in this paper:

Crime Data. In order to effectively process crime data, the entire city can be divided into N zones based on latitude and longitude. Crime data records the occurrences of C different crime categories in these N regions at different time intervals. For a given time period of length T, the crime data for the entire city during this period can be represented as $X \in \mathbb{R}^{T\times N\times C}$. We follow the general setup of crime occurrence prediction research (citing from Huang et al., 2019; Sun et al., 2021). If a crime of type c occurs in region i at time t, the corresponding crime data is recorded as $X^t_{i,c} = 1$. If no crime occurs, the corresponding value is recorded as 0.

Type Graph. The type graph illustrates the relationships between different crime types. As shown in Fig. 2, there is a correlation between different crime types. We use an undirected graph to represent the correlations between types. The structure of the type graph can be described as $G_c = (V_c, E_c, A_c)$. V_c denotes the set of nodes of type graph, i.e., different crime types, E_c denotes the set of relationships between

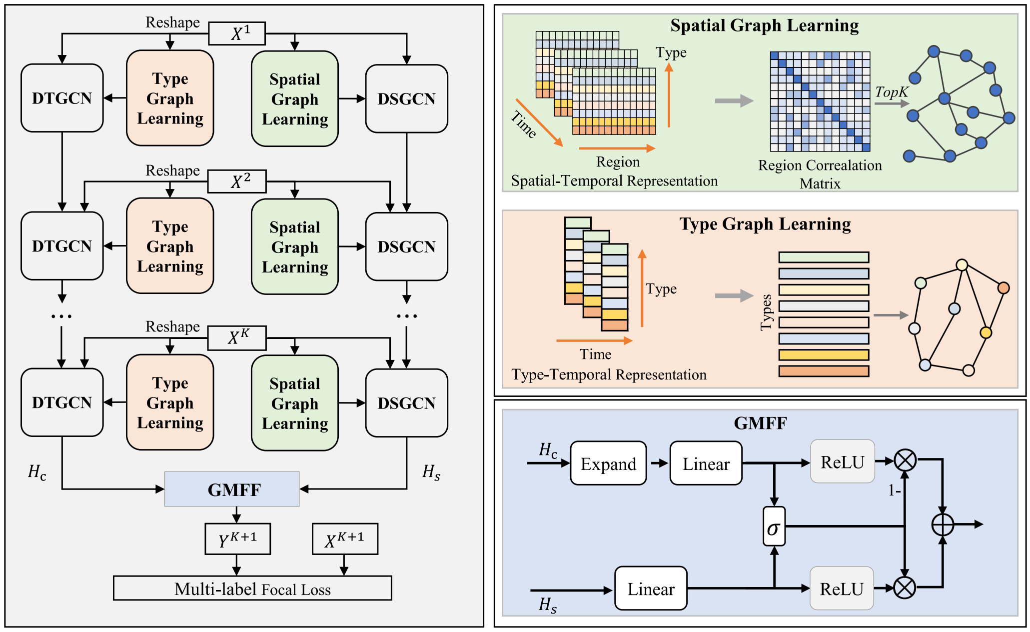


Fig. 3. The overall framework of MRAGNN. The crime data is passed through the spatial map learning and type map learning modules to get the spatial and type graphs. These two types of graph structures serve as inputs to the DSGCN and DTGCN modules to learn the spatial and type dependencies of the representation dynamics, respectively. Finally, the GMFF module integrates these two types of representation features to produce the prediction results.

crime types, and A_c signifies the adjacency matrix of the crime type graph.

Spatial Graph. The spatial graph illustrates the crime-related spatial correlations among different regions in the city. To effectively model the relationships of crime flow, we employ a spatial graph to represent the spatial correlations between regions. The spatial graph structure can be denoted as $G_s = (V, E_s, A_s)$, where V represents the nodes set corresponding to N regions, E_s denotes the set of edges between regions, and A_s signifies the adjacency matrix of the spatial graph.

Problem definition. The crime occurrence prediction model takes crime records from time slot T_1 to T_K as input data, which can be expressed as: $\{X^1, X^2, \dots, X^k, \dots, X^K\}$. In our model, the spatial correlation graph for crime regions is denoted as G_s, and the type correlation graph is represented as G_c. The goal of the proposed crime occurrence prediction model is to develop a method that can capture both spatio-temporal and type-temporal dependencies. This approach aims to adaptively learn the graph structures G_s and G_c from the data, allowing for accurate predictions of future crime occurrences. Set the time length of history crime data is K, the prediction result Y^{K+1} denotes the crime occurrence status of different crime types for different areas in the entire city at time $K+1$. Then the crime occurrence prediction model can be summarized as the following equation:

$$Y^{K+1} = f([X^1, X^2, \dots, X^k, \dots, X^K] : G_s, G_c) \tag{1}$$

where $X^k \in \mathbb{R}^{N\times C}$ represents the crime data in the k time period.

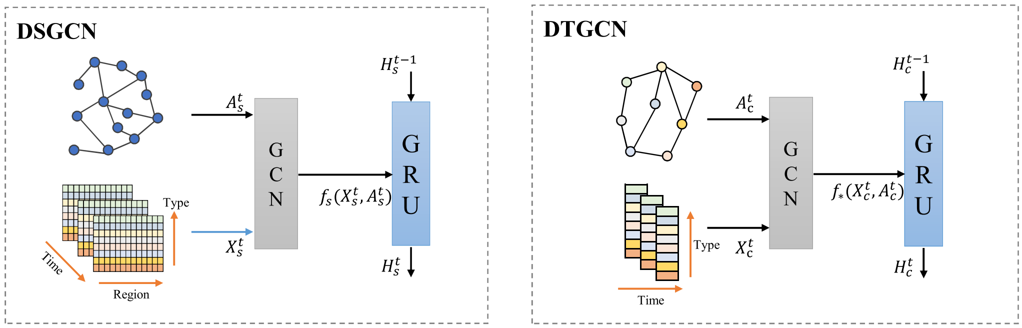
4. Methodology

As shown in Fig. 3, the overall framework of the MRAGNN model consists of five types of modules: spatial graph learning, type graph learning, dynamic spatial graph convolutional network (DSGCN), dynamic type graph convolutional network (DTGCN), and gated multi-modal feature fusion (GMFF). Firstly, the crime data is passed through the spatial graph learning module and the type graph learning module to generate two key graph structures: the spatial graph and the type graph. The spatial graph represents the spatial relationships between different locations, while the type graph represents the relationships between different crime types. Then, these two graph structures serve as inputs to the DSGCN and DTGCN modules, respectively. DSGCN focuses on learning dynamic spatial dependencies to capture spatially relevant features over time, while DTGCN focuses on learning dynamic type dependencies to analyze the changing patterns of different types of crimes. The DSGCN combined with the spatial graph learning layer forms the Dynamic Spatial–Temporal Graph Neural Network branch, while the DTGCN combined with the type graph learning layer forms the Dynamic Type-Temporal Graph Neural Network branch. Each branch consists of a graph learning layer and a dynamic graph convolutional network, aimed at learning feature representations from crime data. Finally, we fuse the spatial–temporal representation and the type-temporal representation learned from these two branches to form a unified feature representation. This fused feature will be used as an input to the decoder to generate predictions about the occurrence of crimes in the future time.

4.1. Dynamic type-temporal graph neural network

Type Graph Learning: Existing methodologies focus on modeling the spatial correlations between different regions within a city based on crime data. However, upon examining the crime data from a different perspective, it becomes apparent that it encompasses a wide range of criminal activities, including robbery, burglary, and others. Fig. 1 highlights correlations among specific types of criminal incidents. We employ a type graph learning method to learn the graph A_c that derives the correlation between different crime types. The data input for the prediction model is represented as $X \in \mathbb{R}^{T\times N\times C}$, where T represents the temporal duration, N signifies the count of regions, and C stands for the variety of types. We aggregate the input data in the spatial dimension, obtaining the total count of each crime type within any given time step. The crime type data is represented as $X_c \in \mathbb{R}^{T\times C}$. The representation of crime types at time step t is obtained by inputting X_c^t into a Fully Connected Layer. Let $M_1 \in \mathbb{R}^{C\times d'}$ and $M_2 \in \mathbb{R}^{C\times d'}$ represent the embedding of the source types and target types, respectively. The correlation matrix A_c for crime types is computed using the following equation:

$$A_c = SoftMax(ReLU(M_1 M_2^T)), \tag{2}$$



(a) DSGCN (b) DTGCN

Fig. 4. *a* and *b* denote the structure of a DSGCN module and a DTGCN module, respectively. Each module mainly consists of GCN layers and a GRU layer.

where $SoftMax$ and $ReLU$ denote nonlinear activation functions. In contrast to the directional nature of spatial graph structures, the connections within the type graph structure are undirected. Similar to spatial graph learning, we perform a sparsification operation on the type graph structure to ensure that the information regarding the connection relations in the type graph structure is effective for prediction.

Dynamic Type Graph Convolutional Network: Based on the analysis above, it can be inferred that different types of crimes are interconnected. Moreover, this correlation changes over time. Using the type graph learning module, we learn a dynamic type graph structure A_c^t for each time step t. To effectively capture the dependency among crime types, we propose a Dynamic Type Graph Convolutional Network (DTGCN) module. As shown in Fig. 4(a), DTGCN mainly comprises a GCN layer and a Gated Recurrent Unit (GRU). The graph convolutional network is formulated as follows:

$$f_*(X_c, A_c, \theta_c) = \sigma(D_c^{-\frac{1}{2}} A_c D_c^{-\frac{1}{2}} X_c \theta_c) \tag{3}$$

where D_c represents the diagonal matrix obtained from the adjacency matrix A_c, and θ_c denotes the weight matrix of the GCN layer. To model the temporal patterns of crime data, we use GRU to learn temporal dynamic features in the predictive model.

$$r_c^t = \sigma(W_c^r f_*(X_c^t, A_c^t, \theta_c) + U_c^r H_c^{t-1} + b_c^r) \tag{4}$$

$$z_c^t = \sigma(W_c^z f_*(X_c^t, A_c^t, \theta_c) + U_c^z H_c^{t-1} + b_c^z) \tag{5}$$

$$\tilde{h}_c^t = tanh(W_c X_c^t + U_c(r_c^t \odot H_c^{t-1})) \tag{6}$$

$$H_c^t = z_c^t \odot H_c^{t-1} + (1 - z_c^t) \odot \tilde{h}_c^t \tag{7}$$

where X_c^t and H_c^t represent the input data and output features of the DTGCN module at moment t, respectively. H_c^{t-1} denotes the output at the moment $t-1$ of the DTGCN module. r_c^t and z_c^t denote the reset and update gates, respectively. b_c^r and b_c^z represent the bias terms. W_c^r, W_c^z, U_c^r, U_c^z, W_c and U_c represent the weight matrices. σ denotes the sigmoid activation function. $\odot$ represents the Hadamard product. With the DGCN module, the model learns the dynamic correlation dependencies between different crime types.

4.2. Dynamic spatial–temporal graph neural network

Spatial Graph Learning: In graph-based prediction models, the quality of the graph structure is crucial to the prediction results. However, existing models usually construct graph structures based on factors such as Euclidean distance and spatial connectivity. This graph construction method has limitations in representing the spatial correlation of crime data. For example, two distant places that are both commercial areas may have similar crime patterns, even if these two areas are not directly connected in space. This is known as implicit correlation. Therefore, when making crime occurrence predictions, we need to consider this potential implicit correlation. To solve this problem, we can learn from the data to obtain a clearly defined graph structure to accurately capture the spatial correlation of crime situations in different areas of the city. In this way, we can not only overcome the limitations of constructing graph structures based on Euclidean distance and spatial connectivity but also better utilize implicit correlations to improve the accuracy of predictions.

In traffic flow prediction models such as Graph WaveNet (Wu et al., 2019) and AGCRN (Bai et al., 2020), the automatically learned adjacency matrices are usually in pairs. Previous studies have shown that crime can move between different areas, suggesting that the occurrence of crime in a particular region can be affected by other nearby regions (Morenoff & Sampson, 1997; Song et al., 2019). Therefore, we use the spatial graph structure to represent the correlation of crimes between urban areas. To explicitly model the interdependencies between crime sequences in different regions, we design a spatial graph learning layer. We propose to de-adaptively learn a latent adjacency matrix A_s, where nodes denote different regions in the city and edges denote correlations between regions. Set $E_1, E_2 \in \mathbb{R}^{N \times d}$ as the initialized embedding matrix of two sets of nodes selected in random order. N represents the number of grid regions obtained by dividing the whole city, and d denotes the dimension of node embedding. Therefore, we compute the adjacency matrix A_s representing the region correlations based on the pairwise region node embeddings. The computational formula is shown below:

$$A_s = ReLU(tanh(\alpha(E_1 E_2^T))) + A_p, \tag{8}$$

where A_s denotes the spatial correlation matrix of the region, A_p denotes the spatial adjacency matrix between different city regions.

The graph structure A_s obtained after initial learning usually contains a large number of nonzero values. If the learned A_s is directly used as the input to the graph network, the computational cost of the model becomes quite large. To alleviate the computational burden on the model, we sparsify the learned spatial graph structure to mask a large number of non-zero elements present in A_s:

$$A_s[i, -argmax(A_s[i, :], k)] = 0, \tag{9}$$

where k denotes the hyperparameter that controls the sparsity of the graph structure A_s. $argmax(A_s[i, :], k)$ denotes the index of the largest k value in row i of the adjacency matrix.

Dynamic Spatial Graph Convolutional Network: The occurrence of crime in urban areas is highly correlated with that of their surrounding regions. The inter-regional crime relationships also undergo continuous changes over time. For instance, the patterns of crime occurrence differ between daytime and nighttime conditions. Therefore, there is a need to dynamically explore spatial dependencies among regions concerning crime data. To model the evolving spatial correlations of crime across regions, we design a Dynamic Spatial Graph Convolutional Network (DSGCN) module. The DSGCN module is illustrated in Fig. 4(b). The following equation shows the computation of the GCN layer:

$$f_*(X, A_s, \theta_s) = \sigma(D_s^{-\frac{1}{2}} A_s D_s^{-\frac{1}{2}} X\theta_s) \tag{10}$$

where D_s represents the diagonal matrix obtained from the adjacency matrix A_s, and θ_s denotes the parameter matrix. To effectively utilize information embedded in the spatial graphs, we also adopt a graph convolution network to model the spatial dependency. Similar to the previous branch that models dynamic type correlation, we employ GRU layers to learn the dynamic temporal features in the predictive model.

$$f_s(X, A_s) = f_*(X, A_s, \theta_s), \tag{11}$$

$$r_s^t = \sigma(W_s^r f_s(X_s^t, A_s^t) + U_s^r H_s^{t-1} + b_s^r) \tag{12}$$

$$z_s^t = \sigma(W_s^z f_s(X_s^t, A_s^t) + U_s^z H_s^{t-1} + b_s^z) \tag{13}$$

$$\tilde{h}_s^t = tanh(W_s X_s^t + U_s(r_s^t \odot H_s^{t-1})) \tag{14}$$

$$H_s^t = z_s^t \odot H_s^{t-1} + (1 - z_s^t) \odot \tilde{h}_s^t \tag{15}$$

where X_s^t and H_s^t denote the inputs and outputs of the DGCN module at moment t. r_s^t and z_s^t represent the reset and update gates. b_s^r and b_s^u denote the bias units. W_s^r, W_s^z, U_s^r, U_s^z, W_s and U_s all represent the weight matrix. σ denotes the nonlinear activation function.

4.3. Gated multimodal feature fusion

After feature extraction, we fuse the feature representations obtained from the spatial–temporal branch and the type-temporal branch. Specifically, H_s and H_c represent the crime feature representations learned from the spatio-temporal graph neural network and the type-temporal graph neural network, respectively. These representations capture relationships among different crime types and spatial patterns across different regions.

To effectively integrate information from two modalities, crime category-dependent, and crime spatial-dependent, we propose a gated multimodal feature fusion module. Due to the inconsistent dimensions of crime category features $H_c^t \in \mathbb{R}^{C\times d_c}$ and spatial features $H_s^t \in \mathbb{R}^{N\times C\times d_s}$, we employ a feature expansion mechanism to extend the dimensions of features in the region dimension for H_c. The transformed feature representation is obtained. Subsequently, we use two linear layers to transform $\hat{H}_c^t$ and H_s^t into the same feature space.

$$\tilde{H}_s^t = Linear(H_s^t) \tag{16}$$

$$\tilde{H}_c^t = Linear(\hat{H}_c^t) \tag{17}$$

To extract the most valuable information for crime occurrence prediction from the features of these two modalities, we adopt a gated feature fusion mechanism to fuse the feature representations of the two modalities. The computational formula for the gated multimodal feature fusion module is as follows:

$$g = \sigma(W_g \cdot [\tilde{H}_s^t, \tilde{H}_c^t]) \tag{18}$$

$$H = g * ReLU(\tilde{H}_s^t) + (1 - g) * ReLU(\tilde{H}_c^t) \tag{19}$$

where σ and $ReLU$ denote the activation functions, and H represents the output of the GMFF module. Through the fusion strategy, our model becomes more effective in learning and comprehending patterns and correlations in urban crime data.

4.4. Multi-label focal loss

Compared to the normal state where no crime occurs, crime events typically exhibit a considerably sparse distribution in both time and space. Additionally, the occurrence frequencies of certain crime types are relatively low. For instance, robbery incidents are generally much rarer than theft cases. This imbalanced distribution of data poses a challenge for the construction of crime occurrence prediction models. To address this problem, we design an enhanced loss function called multi-label classification focal loss in reference to Ridnik et al. (2021), which is different from the commonly used cross-entropy loss function. The distinguishing feature of the multi-label classification focal loss lies in its introduction of weight factors during the training process, with a specific emphasis on relatively rare and hard-to-classify crime event samples within the dataset. By increasing the weights of these samples, the loss function can effectively enhance the predictive accuracy for these challenging samples. The skillful application of this strategy underscores the outstanding performance of our model in handling imbalanced data, thereby contributing to improved identification and prediction accuracy for rare crime events. The calculation formula for the multi-label classification focal loss is as follows:

$$\begin{aligned}\mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{C} \; & (y_{ij}(1-\hat{y}_{ij})^{\lambda} log(\hat{y}_{ij}) \\ & - (1-y_{ij})(\hat{y}_{ij})^{\lambda} log(1-\hat{y}_{ij})),\end{aligned} \tag{20}$$

where $\hat{y}_{ij}$ represents the probability value of prediction type, λ denotes parameters that control the imbalance sampling weights. We adopt the multi-label classification focal loss as the predicted loss to further alleviate the difficulty in predicting crime due to imbalanced data distribution.

Table 1
The overall information for datasets.

Datasets	Nodes	Timesteps	Time range	Types
Los Angeles	113	16,992	01/2018 - 12/2018	8
Chicago	77	17,856	01/2015 - 12/2016	8

5. Experiments

In this section, we demonstrate the superiority of our model through experiments. We conducted experiments on a workstation equipped with a 3090 GPU (24 GB memory).

5.1. Datasets and setup

We conducted experiments on the crime record datasets of two cities to validate the effectiveness of the MRAGNN model. As shown in Table Table 1, detailed information on these two datasets can be found in CrimeForecaster (Sun et al., 2021).

Los Angeles (LA) (Sun et al., 2021): crime data for Los Angeles city in the year 2018. Crime types can be categorized into eight, including Theft, Vehicle Theft, Burglary, Fraud, Assault, Sexual offenses, Robbery, and Vandalism.

Chicago (CHI) (Sun et al., 2021): crime data for Chicago city in 2015. Crime types can be categorized into eight, including Theft, Criminal damage, Narcotics, Robbery, Assault, Deceptive practices, Burglary, and Battery.

Based on the original crime dataset, we aggregated the counts of different types of crimes for each day. For experimental validation, we selected data for a single month between August and December from two datasets to serve as the test set. We took the data from 6.5 months preceding the test data as training data, and 0.5 months of data as validation data. The learning rate was set to 0.01. The number of epochs for both datasets was established at 100, with a batch size of 64. The diffusion convolutional layer consists of two stacked RNN layers, each with a hidden layer size of 64. The number of graph network layers in both the DSGCN and DTGCN modules is 3. The sparsity coefficient k_s for the spatial graph learning layer was set to 50, and

Table 2
Performance comparison with the state-of-the-art baselines on crime forecasting (Precision and Recall).

Data	Month	Metric	ARIMA	GRU	SVR	RF	MiST	GWNet	CF	HAGEN	MRAGNN
CHI	August	Precision	0.4877	0.5230	0.5793	0.6761	0.6789	0.6678	0.7063	0.7026	**0.7278**
		Recall	0.4808	0.5018	0.6011	0.6929	0.6679	0.6450	0.6959	0.7301	**0.7477**
	September	Precision	0.5018	0.4883	0.5967	0.7093	0.6745	0.6532	0.7003	**0.7404**	0.7383
		Recall	0.4761	0.4623	0.6150	0.6783	0.6936	0.6470	0.6911	0.7057	**0.7233**
	October	Precision	0.4856	0.4708	0.6304	0.6858	0.6535	0.6447	0.6961	**0.7186**	0.7118
		Recall	0.4671	0.5028	0.6431	0.6964	0.6809	0.6476	0.6788	0.7083	**0.7357**
	November	Precision	0.4663	0.4721	0.5883	0.6656	0.6884	0.6062	0.6751	0.6828	**0.6982**
		Recall	0.4481	0.4598	0.6055	0.6972	0.6680	0.6307	0.6944	0.7109	**0.7244**
	December	Precision	0.4591	0.4762	0.6215	0.6889	0.6930	0.6066	0.6964	0.6822	**0.7041**
		Recall	0.4340	0.4462	0.6345	0.6598	0.6606	0.5813	0.6621	0.7074	**0.7209**
	Average	Precision	0.4801	0.4861	0.6433	0.6851	0.6937	0.6557	0.7148	0.7253	**0.7380**
		Recall	0.4612	0.4746	0.6398	0.6849	0.6742	0.6503	0.6845	0.7145	**0.7304**
LA	August	Precision	0.3833	0.4129	0.4664	0.5931	0.6008	0.5802	0.5896	0.6279	**0.6458**
		Recall	0.3656	0.3906	0.455	0.5635	0.5979	0.5643	0.6042	0.6186	**0.6203**
	September	Precision	0.3858	0.4181	0.4853	0.5696	0.582	0.5679	0.6017	0.6307	**0.6438**
		Recall	0.3583	0.4401	0.4697	0.5999	0.6107	0.5726	0.6146	0.6016	**0.6351**
	October	Precision	0.3567	0.3944	0.5363	0.5627	0.6153	0.5798	0.5905	**0.6255**	0.6123
		Recall	0.3856	0.4041	0.5202	0.5811	0.5880	0.5555	0.5864	0.6140	**0.6304**
	November	Precision	0.3873	0.4622	0.4784	0.5681	0.5475	0.5696	0.5796	0.6005	**0.6460**
		Recall	0.3651	0.4644	0.4965	0.5848	0.5265	0.5764	0.6047	0.6250	**0.6308**
	December	Precision	0.3609	0.4535	0.5235	0.5594	0.5156	0.5413	0.5663	0.6256	**0.6385**
		Recall	0.3827	0.4628	0.5361	0.5831	0.548	0.5611	0.5943	0.5956	**0.6179**
	Average	Precision	0.3748	0.4282	0.4980	0.5706	0.5722	0.5678	0.5856	0.6220	**0.6373**
		Recall	0.3715	0.4324	0.4955	0.5825	0.5742	0.5660	0.6008	0.6110	**0.6269**

for the type graph learning layer, k_s was set at 4. The parameter λ for the multi-label focal loss was configured to 2.

Evaluation Metric: In our model, the goal of crime occurrence prediction is to predict whether a crime of any type in the target region will occur or not. Therefore, we generalize it as a multi-label classification problem, aiming to classify whether multiple types of crimes occur within the target region. Similar to CrimeForecaster (Sun et al., 2021) and HAGEN (Wang et al., 2022), two metrics (Micro-F1 and Macro-F1) are used to evaluate the crime occurrence prediction model. In addition, we have incorporated commonly used classification metrics such as Precision, Recall, and ROC-AUC into our experimental results.

5.2. Baselines

We utilized the following methods as baselines for comparison:

ARIMA (Chen et al., 2008): As a classical time forecasting model, ARIMA combines both autoregressive and moving average methods.

GRU (Fu et al., 2016): GRU is a variant of recurrent neural networks, which is widely employed in time prediction and commonly used in crime occurrence prediction research.

SVR (Wu et al., 2004): SVR (Support Vector Regression) is a regression prediction model that uses a machine learning method.

Random Forest (Speiser et al., 2019): Random Forest(RF) integrates multiple decision trees to achieve crime occurrence prediction.

Graph WaveNet (GWNet) (Wu et al., 2019): Graph WaveNet is a predictive model used for spatio-temporal graph modeling. It accurately captures spatial correlations in spatio-temporal data by learning an adaptive dependency matrix that describes the relationships between nodes.

MiST (Huang et al., 2019): MiST is a deep learning-based crime occurrence prediction model that learns spatio-temporal correlations between regions.

CF (Sun et al., 2021): CF is an end-to-end crime occurrence prediction model. This model leverages a gated recurrent network with diffusive convolution modules to learn temporal recursiveness and spatial dependencies.

HAGEN (Wang et al., 2022): HAGEN introduces a homophily perception constraint to optimize the generation of region-relevant graphs. This approach proposes a constraint that neighboring nodes have similar crime occurrence patterns, thus aligning with the diffusion convolution mechanism.

5.3. Forecast results and analysis

In Table 2, 3 and 4, comprehensive presentations of the crime occurrence prediction results on the CHI and LA datasets have been provided. We employed five evaluation metrics to thoroughly assess the prediction performance of all methods. The following are the key findings distilled from Tables 2–4:

As can be seen in Table 2, our model (MRAGNN) shows better performance in most cases for the Chicago (CHI) and Los Angeles (LA) city datasets. On the LA dataset, the MRAGNN model exhibits high mean values of 0.6373 and 0.6269 for both Precision and Recall, respectively, which indicates that the model is not only able to accurately identify crime events when predicting crime events (high Precision), but also able to cover a larger number of crime events (high Recall). On the CHI dataset, while the MRAGNN model still performs better, other models such as on HAGEN also show high values. This observation may be attributed to variations in crime patterns between the two cities, as well as differences in dataset size and quality, which can influence the adaptability of various models. Our MRAGNN model performs particularly well in specific months compared to the other models. Similarly, the MRAGNN model has better ROC_AUC results for most months across both datasets, as shown in Table 4. For the multilabel classification task, a higher ROC_AUC value (i.e., close to 1.0) means that the model performs well in distinguishing between positive and negative classes with multiple labels. This result demonstrates that MRAGNN performs well in predicting crime data with unbalanced distribution.

Effectiveness of graph network-based models: As shown in Table 3, models like CF, HAGEN, and MRAGNN, which are based on graph networks, demonstrate notable performance advantages in crime occurrence prediction tasks, surpassing other baseline models. These models excel in capturing crime correlations between different regions, leading to more accurate predictions. The MiST model achieves significant

Table 3
Performance comparison with the state-of-the-art baselines on crime forecasting(Micro-F1 and Macro-F1).

Data	Month	Metric	ARIMA	GRU	SVR	RF	MiST	GWNet	CF	HAGEN	MRAGNN
CHI	August	Micro-F1	0.4867	0.5047	0.5822	0.6860	0.6719	0.6593	0.7052	0.7209	**0.7315**
		Macro-F1	0.4205	0.4339	0.3424	0.3423	0.6176	0.6201	0.6636	0.6791	**0.6831**
	September	Micro-F1	0.4836	0.4818	0.6064	0.6967	0.6892	0.6512	0.6929	0.7211	**0.7332**
		Macro-F1	0.3979	0.4206	0.3288	0.3587	0.6351	0.6111	0.6491	0.6779	**0.6850**
	October	Micro-F1	0.4757	0.4886	0.6395	0.6933	0.6692	0.6456	0.6931	0.7129	**0.7297**
		Macro-F1	0.3881	0.4125	0.3352	0.3624	0.6211	0.6068	0.6506	0.6738	**0.6813**
	November	Micro-F1	0.4520	0.4657	0.5935	0.6833	0.6766	0.6238	0.6774	0.6946	**0.7172**
		Macro-F1	0.3667	0.4059	0.3296	0.3523	0.6262	0.5849	0.6356	0.6528	**0.6692**
	December	Micro-F1	0.4528	0.4655	0.6261	0.6795	0.6753	0.5933	0.6773	0.6907	**0.7131**
		Macro-F1	0.3697	0.4034	0.3325	0.3540	0.6138	0.5581	0.6379	0.6467	**0.6627**
	Average	Micro-F1	0.4702	0.4813	0.6095	0.6878	0.6764	0.6346	0.6892	0.7080	**0.7249**
		Macro-F1	0.3886	0.4153	0.3337	0.3539	0.6228	0.5962	0.6474	0.6661	**0.6763**
LA	August	Micro-F1	0.3711	0.3931	0.4569	0.5798	0.5991	0.5699	0.6038	0.6216	**0.6356**
		Macro-F1	0.3036	0.3285	0.2119	0.2054	0.4920	0.5079	0.5359	0.5509	**0.5683**
	September	Micro-F1	0.3668	0.4357	0.4748	0.5849	0.5998	0.5719	0.6035	0.6210	**0.6371**
		Macro-F1	0.2959	0.3160	0.2005	0.2097	0.4877	0.4973	0.5180	0.5443	**0.5562**
	October	Micro-F1	0.3722	0.3994	0.5302	0.5742	0.5956	0.5611	0.5886	0.6165	**0.6283**
		Macro-F1	0.3010	0.3197	0.1992	0.4492	0.4754	0.4891	0.5028	0.5455	**0.5590**
	November	Micro-F1	0.3800	0.4643	0.4899	0.5790	0.5393	0.5711	0.5922	0.6190	**0.6309**
		Macro-F1	0.3090	0.3890	0.2048	0.4519	0.4999	0.5002	0.5083	0.5373	**0.5507**
	December	Micro-F1	0.3730	0.4543	0.5294	0.5685	0.5343	0.5570	0.5841	0.6113	**0.6242**
		Macro-F1	0.3010	0.3662	0.2084	0.4401	0.4944	0.4843	0.4921	0.5250	**0.5426**
	Average	Micro-F1	0.3726	0.4294	0.4962	0.5773	0.5736	0.5662	0.5944	0.6179	**0.6312**
		Macro-F1	0.3021	0.3439	0.2049	0.3513	0.4899	0.4958	0.5114	0.5406	**0.5554**

Table 4
Performance comparison with baselines on crime forecasting (ROC_AUC).

Dataset	Month	ARIMA	GRU	SVR	RF	MiST	GWNet	CF	HAGEN	MRAGNN
CHI	August	0.4860	0.5138	0.5998	0.7094	0.6927	0.6667	0.6931	0.7350	**0.7412**
	September	0.5012	0.5045	0.6382	0.7041	0.6814	0.6318	0.7205	0.7377	**0.7395**
	October	0.5347	0.5447	0.6744	0.6914	0.6900	0.6432	0.7168	0.7160	**0.7190**
	November	0.5288	0.5569	0.6302	0.6870	0.6752	0.6590	0.7121	**0.7294**	0.7218
	December	0.5198	0.5610	0.6696	0.6858	0.6341	0.6055	0.7137	0.7159	**0.7253**
	Average	0.5141	0.5362	0.6424	0.6955	0.6747	0.6412	0.7112	0.7268	**0.7294**
LA	August	0.4236	0.5024	0.5863	0.6583	0.6771	0.6713	0.7008	**0.7203**	0.7184
	September	0.3943	0.5330	0.6122	0.6615	0.7024	0.6910	0.7131	**0.7221**	0.7205
	October	0.4186	0.4640	0.6547	0.6757	0.6748	0.6638	0.7066	0.7190	**0.7269**
	November	0.4663	0.5796	0.6208	0.6487	0.6910	0.6917	0.7062	0.7227	**0.7280**
	December	0.4361	0.5174	0.6492	0.6525	0.6817	0.6935	0.7024	0.7182	**0.7193**
	Average	0.4278	0.5193	0.6242	0.6593	0.6854	0.6827	0.7058	0.7205	**0.7226**

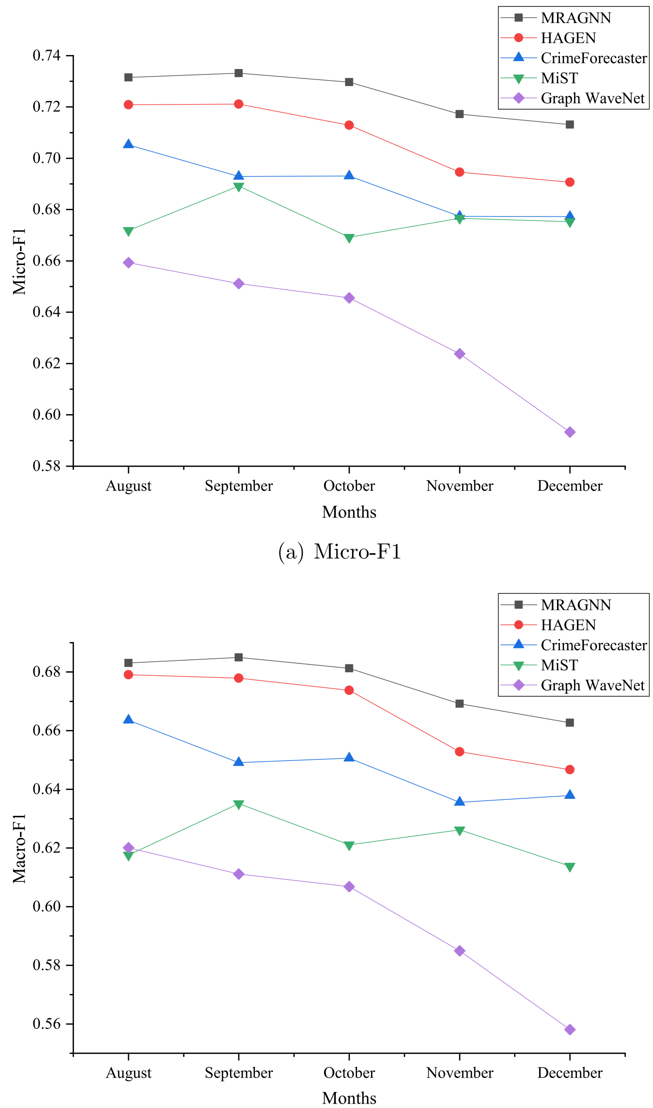
predictive performance by using a grid structure to represent crime correlations between different regions. Compared to MiST, MRAGNN shows a noteworthy improvement: 7.17 % in average micro-F1 and 8.59% in average macro-F1 on the CHI dataset. This substantial enhancement confirms the outstanding performance of MRAGNN. It also emphasizes the significant potential of graph structures in capturing spatial correlations in crime occurrences across different regions.

Effectiveness of graph learning with the Type Correlations: In both the CHI and LA datasets, MRAGNN delivered an outstanding performance in the testing data spanning five months, clearly surpassing other baseline models. Specifically, the performance of the MRAGNN model consistently outperformed that of the best-performing HAGEN model. Concerning the Micro-F1 metric on the CHI dataset, it achieved an average score of 0.7249, representing an overall improvement of 2.39%.

Regarding the five-month Macro-F1 metric on the CHI dataset, our model also exhibited significant improvements. As depicted in Fig. 5, compared to several baseline models that solely employed graph structures to model spatial correlations within crime data, our proposed MRAGNN model performed exceptionally well. Notably, even when considering data from November and December, our model's predictive performance remained relatively stable, further affirming the usefulness of dedicating a branch to modeling type correlations in crime data. These types correlations include valuable crime pattern information, aiding the predictive model in capturing crime occurrence patterns with greater precision.

On the LA dataset, our model exhibited a similar overall performance improvement trend as observed in the CHI dataset. The average Micro-F1 and average Macro-F1 metrics improved by 2.16% and 2.73%, respectively, relative to the HAGEN model. The consistent improvement in performance highlights the versatility and effectiveness of the MRAGNN model in the field of crime occurrence prediction.

Effectiveness of type information: To verify the significance of category information in crime prediction, we conducted comparative experiments using two datasets, CHI and LA, both sampled from the month of August. The experiments included three models: CF, HAGEN, and MRAGNN, along with two model variants, CF+type and HAGEN+type, which incorporate category information. To fully exploit the category-temporal information, we added category graph structure modeling to both CF and HAGEN. By employing a specialized category graph network, we derived the CF+type and HAGEN+type variants.



(a) Micro-F1

(b) Macro-F1

Fig. 5. The results of two metrics for different models on the CHI dataset.

The results, as shown in Table 5, demonstrate that category information significantly enhances the performance of crime prediction models. The models that integrate category information (CF+type and HAGEN+type) achieved higher Micro-F1 and Macro-F1 scores compared to their original counterparts (CF and HAGEN) on both the CHI and LA datasets. Specifically, the Micro-F1 scores of CF+type and HAGEN+type improved by approximately 1.50% and 0.78%, respectively, on the CHI dataset, and the improvements on the LA dataset were similarly notable.

Additionally, the MRAGNN model, which explicitly incorporates category information, achieves the best overall prediction performance. This suggests that MRAGNN effectively captures the fusion of category features and spatio-temporal information, further proving the value of integrating category information in crime prediction models.

5.4. Analysis of the model's time complexity

In this section, we analyze the time complexity of the model. We calculated the average training, validation, and testing times per epoch for the proposed model and several baseline models across two types of datasets. The baseline models for comparison include HAGEN and CrimeForecaster. All experiments were conducted on a server equipped with an RTX 3090 GPU, and the batch size was kept consistent across all models.

Table 5
Comparison of model performance with type information on CHI and LA datasets.

Model	CHI dataset		LA dataset	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1
CF	0.7052	0.6636	0.6038	0.5359
CF+type	0.7158	0.6724	0.6092	0.5403
HAGEN	0.7209	0.6791	0.6216	0.5509
HAGEN+type	0.7265	0.6817	0.6238	0.5542
MRAGNN	**0.7315**	**0.6831**	**0.6356**	**0.5683**

Table 6
Training, Validation, and Testing Time per epoch for Models on CHI and LA Datasets.

Dataset	Model	time/epoch(s)		
		training time	validation time	testing time
CHI	CrimeForecaster	13.3	3.5	3.6
	HAGEN	30.7	3.1	3.2
	MRAGNN	34.3	3.5	3.7
LA	CrimeForecaster	7.5	1.7	1.8
	HAGEN	29.4	1.1	2.3
	MRAGNN	44.0	1.2	2.4

The results in the Table 6 indicate that the model's computational efficiency has decreased after incorporating category information into the crime prediction. However, the differences in computational time among all models are minimal, and our model shows significant improvements when combined with the previous prediction results. The experimental results demonstrate that our model achieves a win–win effect in both predictive performance and computational efficiency.

5.5. Ablation study

We have completed ablation experiments for the proposed model, which aimed at assessing the contribution of each key module to our prediction framework. We considered three different degraded variants: **MRAGNN-TT**, **MRAGNN-GL**, and **MRAGNN-FL**.

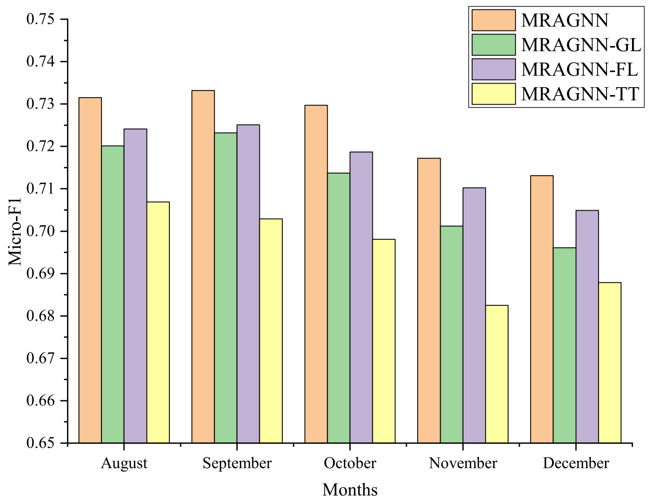
(1) **MRAGNN-TT** removes the type-temporal modeling branch, resulting in a spatio-temporal prediction model similar to Graph WaveNet and CF.

(2) **MRAGNN-GL** removes the graph learning layers, avoiding the use of spatial and type graph learning to construct spatial graph structures representing crime-relatedness between regions and type graph structures representing relationships between different crime types. In this variant, predefined spatial graphs and type relationship graphs based on the similarity of different type features are used instead.

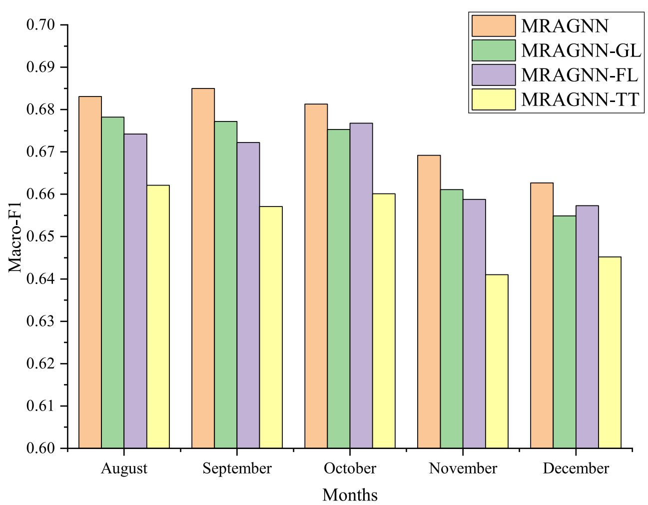
(3) **MRAGNN-FL** does not use the proposed multi-label classification focal loss. In this variant, Binary Cross-Entropy (Chen & Liao, 2023) is used to replace the proposed multi-label classification focal loss.

The experimental results are shown in Fig. 6. We compared the complete MRAGNN model with these variants and found that removing any component from MRAGNN leads to a decrease in performance. This emphasizes the importance of each component and validates the rationality of our model design. Specifically, the MRAGNN-TT exhibited the poorest performance, highlighting the crucial significance of modeling crime-type correlations for the predictive model. The presence of the type-temporal graph neural network branch allows for capturing complex interactions and dependencies between different crime types, which is essential for accurately predicting crime incidents.

The MRAGNN-GL variant shows a decrease in performance compared to the complete MRAGNN model, further demonstrating the importance of the graph learning layers in MRAGNN. The introduction



(a) Micro-F1



(b) Macro-F1

Fig. 6. Evaluation on the ablated variants of MRAGNN.

of spatial graph learning and type graph learning layers enables more precise modeling of spatial correlations and type correlations within crime data compared to predefined graph structures. The MRAGNN-FL variant exhibits lower prediction performance than the MRAGNN model. This is because crime data contains a substantial number of zero values, leading to bias in the prediction model's learning of crime representations. The ablation results confirm the effectiveness of our proposed multi-label classification focal loss in addressing prediction bias caused by imbalanced crime distributions.

Through these ablation experiments, we have validated the importance of each component in the MRAGNN model, emphasizing the crucial role of crime-type correlation information in improving the performance of crime occurrence prediction models.

5.6. Discussion of social issues

Crime prediction models often raise complex social issues related to bias and fairness. In this section, we thoroughly examine the social implications of our proposed model and the strategies employed to mitigate potential challenges.

Our model addresses a multi-label classification task, focusing on predicting whether different types of crimes will occur in a specific region, rather than modeling individual criminal acts or estimating the total number of crimes across areas. The model is trained solely on historical crime occurrence data and does not incorporate demographic variables, such as population distribution, racial composition, or socioeconomic factors. This design choice aims to minimize the risk of social bias, such as racial discrimination, ensuring that predictions rely exclusively on correlations within the data, without introducing sensitive or subjective social assumptions. However, crime data itself is not entirely neutral. In areas with higher crime rates, stronger law enforcement efforts often lead to more frequent crime reporting. This can create a "feedback loop", where higher crime predictions reinforce increased surveillance and skew resource allocation toward these areas. Even if crime levels improve, the model may continue to classify such regions as high-risk due to the persistent pattern of high-frequency reporting. This unintended bias risks exacerbating social inequalities by justifying disproportionate policy measures that negatively impact vulnerable communities.

To address the issue of data imbalance, we incorporate a multi-label focal loss function, which reduces the model's tendency to overemphasize high-crime regions and frequent crime types. This approach promotes more balanced predictions across different areas and ensures that less common crime types or low-frequency events are not overlooked. While this adjustment improves fairness to some extent, it does not fully eliminate the structural biases embedded in the crime data.

6. Conclusion

To address the challenge of crime-type correlations and mitigate prediction bias caused by the imbalanced distribution of crime data, we propose a novel model called Multi-type Relations Aware Graph Neural Networks (MRAGNN). This model effectively captures both spatial–temporal and type-temporal dependencies within crime data, offering a more holistic approach to crime prediction. In addition, we designed an improved multi-label classification focal loss to address the issue of imbalanced crime data, ensuring that the model does not overemphasize regions or crime types with higher occurrence frequencies. Extensive experiments conducted on crime datasets from two cities validate the effectiveness of our approach, demonstrating significant improvements in predicting crime occurrences. We believe this research will serve as an important contribution to the advancement of crime prediction methodologies.

In future research, we will delve deeper into leveraging crime-type correlations to further enhance prediction accuracy. We also aim to incorporate additional factors, such as socio-economic indicators, demographic data, and external environmental variables, to enrich the model's understanding of crime patterns and improve its predictive power. Additionally, generalizing the model across datasets with varying distributions remains a key area for exploration. This will involve designing adaptable modules that can efficiently model spatial and crime-type graph structures, allowing for more accurate predictions in diverse urban settings. In addition to enhancing prediction performance, improving the interpretability of crime prediction models will be a key focus in future research. While graph-based models like MRAGNN can effectively capture complex spatio-temporal and crime-type relationships, they are often seen as "black-box" models, which limits their transparency in practical applications such as public safety decision-making. Therefore, a critical direction will be to explore methods that offer more interpretable insights into the crime prediction process.

CRediT authorship contribution statement

Shun Wang: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Yong Zhang:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing. **Xinglin Piao:** Resources, Data curation. **Xuanqi Lin:** Validation, Writing – original draft, Visualization. **Yongli Hu:** Supervision, Funding acquisition. **Baocai Yin:** Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgments

This work has been supported by National Key R&D Program of China (No. 2021ZD0111902), National Natural Science Foundation of China (No. 62072015, U21B2038, U19B2039), Beijing Natural Science Foundation (No. 4222021), R&D Program of Beijing Municipal Education Commission (No. KZ202210005008).

Data availability

Data will be made available on request.

References

Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in Neural Information Processing Systems*, *33*, 17804–17815.

Boukabous, M., & Azizi, M. (2022). Multimodal sentiment analysis using audio and text for crime detection. In *International conference on innovative research in applied science, engineering and technology* (pp. 1–5). IEEE.

Chen, B., & Liao, Y. (2023). Spatio-temporal deep fusion graph convolutional networks for crime prediction. In *Proceedings of the 2023 7th international conference on machine learning and soft computing* (pp. 75–81).

Chen, P., Yuan, H., & Shu, X. (2008). Forecasting crime using the arima model. In *International conference on fuzzy systems and knowledge discovery (Vol. 5)* (pp. 627–630). IEEE.

Fu, R., Zhang, Z., & Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. In *Youth academic annual conference of Chinese association of automation* (pp. 324–328). IEEE.

Hu, K., Li, L., Liu, J., & Sun, D. (2021). Duronet: A dual-robust enhanced spatial-temporal learning network for urban crime prediction. *ACM Transactions on Internet Technology*, *21*(1), 1–24.

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

Huang, C., Zhang, J., Zheng, Y., & Chawla, N. V. (2018). DeepCrime: Attentive hierarchical recurrent networks for crime prediction. In *Proceedings of ACM international conference on information and knowledge management* (pp. 1423–1432).

Jin, G., Li, F., Zhang, J., Wang, M., & Huang, J. (2022). Automated dilated spatio-temporal synchronous graph modeling for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 1–11. http://dx.doi.org/10.1109/TITS.2022.3195232.

Li, F., Feng, J., Yan, H., Jin, G., Yang, F., Sun, F., Jin, D., & Li, Y. (2023). Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution. *ACM Transactions on Knowledge Discovery from Data*, *17*(1), 1–21. http://dx.doi.org/10.1145/3532611, URL https://dl.acm.org/doi/10.1145/3532611.

Li, Z., Huang, C., Xia, L., Xu, Y., & Pei, J. (2022). Spatial-temporal hypergraph self-supervised learning for crime prediction. In *IEEE international conference on data engineering* (pp. 2984–2996). IEEE.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*.

Liang, W., Wang, Y., Tao, H., & Cao, J. (2022). Towards hour-level crime prediction: A neural attentive framework with spatial–temporal-categorical fusion. *Neurocomputing*, *486*, 286–297.

Liao, R., Wang, X., Li, L., & Qin, Z. (2010). A novel serial crime prediction model based on Bayesian learning theory. In *International conference on machine learning and cybernetics (Vol. 4)* (pp. 1757–1762). IEEE.

Lin, H., Gao, Z., Xu, Y., Wu, L., Li, L., & Li, S. Z. (2022). Conditional local convolution for spatio-temporal meteorological forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*, 7470–7478. http://dx.doi.org/10.1609/aaai.v36i7.20711, URL https://ojs.aaai.org/index.php/AAAI/article/view/20711.

Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, *30*(3), 491–497.

Morenoff, J. D., & Sampson, R. J. (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. *Social Forces*, *76*(1), 31–64.

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*(12), 1565–1567.

Parameswaran, M., & Whinston, A. B. (2007). Social computing: An overview. *Communications of the Association for Information Systems*, *19*(1), 37.

Quick, M., Li, G., & Brunton-Smith, I. (2018). Crime-general and crime-specific spatial patterns: A multivariate spatial analysis of four crime types at the small-area scale. *Journal of Criminal Justice*, *58*, 22–32.

Rajapakshe, C., Balasooriya, S., Dayarathna, H., Ranaweera, N., Walgampaya, N., & Pemadasa, N. (2019). Using cnns rnns and machine learning algorithms for real-time crime prediction. In *International conference on advancements in computing* (pp. 310–316). IEEE.

Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M., & Zelnik-Manor, L. (2021). Asymmetric loss for multi-label classification. In *Proceedings of the 2021 IEEE/CVF international conference on computer vision* (pp. 82–91).

Sampson, R. J., & Groves, W. B. (2017). Community structure and crime: Testing social-disorganization theory. In *Social, ecological and environmental theories of crime* (pp. 93–122). Routledge.

Sharma, M. (2014). Z-CRIME: A data mining tool for the detection of suspicious criminal activities based on decision tree. In *International conference on data mining and intelligent computing* (pp. 1–6). IEEE.

Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. *Journal of Quantitative Criminology*, *35*, 831–854.

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, *134*, 93–101.

Steffensmeier, D. J., Allan, E. A., Harer, M. D., & Streifel, C. (1989). Age and the distribution of crime. *American Journal of Sociology*, *94*(4), 803–831.

Sun, J., Yue, M., Lin, Z., Yang, X., Nocera, L., Kahn, G., & Shahabi, C. (2021). CrimeForeCaster: crime prediction by exploiting the geographical neighborhoods' spatiotemporal dependencies. In *Machine learning and knowledge discovery in databases. applied data science and demo track: European conference, ECML pKDD 2020, ghent, Belgium, September 14–18, 2020, proceedings, part v* (pp. 52–67). Springer.

Sutskever, I., Martens, J., & Hinton, G. E. (2011). Generating text with recurrent neural networks. In *Proceedings of international conference on machine learning* (pp. 1017–1024).

Wang, C., Lin, Z., Yang, X., Sun, J., Yue, M., & Shahabi, C. (2022). Hagen: Homophily-aware graph convolutional recurrent network for crime forecasting. In *Proceedings of the AAAI conference on artificial intelligence (Vol. 36)* (pp. 4193–4200).

Wang, S., Zhang, Y., Hu, Y., & Yin, B. (2023). Knowledge fusion enhanced graph neural network for traffic flow prediction. *Physica A. Statistical Mechanics and its Applications*, *623*, Article 128842.

Wang, S., Zhang, M., Miao, H., Peng, Z., & Yu, P. S. (2022). Multivariate correlation-aware spatio-temporal graph convolutional networks for multi-scale traffic prediction. *ACM Transactions on Intelligent Systems and Technology*, *13*(3), 38:1–38:22. http://dx.doi.org/10.1145/3469087, URL https://dl.acm.org/doi/10.1145/3469087.

Wu, C.-H., Ho, J.-M., & Lee, D.-T. (2004). Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*, *5*(4), 276–281.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24.

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for deep spatial-temporal graph modeling. In *Proceedings of the twenty-eighth international joint conference on artificial intelligence*. International Joint Conferences on Artificial Intelligence Organization.

Yi, F., Yu, Z., Zhuang, F., & Guo, B. (2019). Neural network based continuous conditional random field for fine-grained crime prediction. In *International joint conference on artificial intelligence* (pp. 4157–4163).

Zhao, X., Fan, W., Liu, H., & Tang, J. (2022). Multi-type urban crime prediction. In *Proceedings of the AAAI conference on artificial intelligence (Vol. 36)* (pp. 4388–4396).

Zhao, X., & Tang, J. (2018). Crime in urban areas: A data mining perspective. *ACM Sigkdd Explorations Newsletter*, *20*(1), 1–12.

Zhuang, D., Wang, S., Koutsopoulos, H., & Zhao, J. (2022). Uncertainty quantification of sparse travel demand prediction with spatial-temporal graph neural networks. In *Proceedings of ACM SIGKDD conference on knowledge discovery and data mining* (pp. 4639–4647). Washington DC USA: ACM, http://dx.doi.org/10.1145/3534678.3539093, URL https://dl.acm.org/doi/10.1145/3534678.3539093.