# A Multi-regional Spatio-temporal Network for Traffic Accident Risk Prediction

Qingrong Wang, Kai Zhang, Changfeng Zhu, Yutong Zhou

*Abstract*—Traffic accident risk prediction is a cornerstone of intelligent mobility, and the main challenge is adequately capturing dynamic spatial and temporal characteristics. However, most existing approaches ignore two critical features of traffic accident risk prediction. First, they introduce traffic conditions only at strictly periodic times, thus weakening the role of temporal continuity. Second, most methods consider the spatio-temporal characteristics of local or global regions, ignoring the effects of multi-regional convergence. Considering the aforementioned issues, we put forward a model of fused multi-regional spatio-temporal characteristics under periodic translation (ATCGCN). The model fuses adjacent and multi-adjacent regions to accurately predict road accident risk by capturing dynamic spatio-temporal correlations. Inspired by the first feature, this paper combines traffic data from adjacent and cycle panning periods, captures their temporal characteristics through GRU, and uses the attention mechanism to characterize the space-time features of multi-adjacent regions dynamically. To achieve the second feature, ATCGCN introduces CNN and GCN to capture the spatial correlations of adjacent and multi-adjacent regions, respectively, and then fuses the spatial and temporal interdependence of the two different regions to enhance the accuracy of traffic accident risk prediction effectively. Through executing trials on a pair of actual datasets from the real world, we highlight the efficacy of each component in ATCGCN and demonstrate that its predictive performance surpasses that of several existing approaches.

*Index Terms*—Traffic Accident Risk, Multi-regional Convergence, Cycle Panning Periods, Spatio-temporal Characteristics

## I  INTRODUCTION

The increasing number of motor vehicles has led to a corresponding rise in the frequency of traffic accidents. This phenomenon has significantly impacted public life, as it represents a primary contributing factor to traffic congestion. Accordingly, accurate prediction of traffic accident risk has emerged as an essential factor in the development of modern cities. Such predictions enable individuals to make more informed route-planning decisions and schedule their trips m

Qingrong Wang is a professor at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: wangqr003@163. com).

Kai Zhang is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, e-mail: zk611@qq.com).

Changfeng Zhu is a professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: cfzhu003@163. com)

Yutong Zhou is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 937245272@qq.com).

-ore effectively. Additionally, they provide traffic managem-ent departments with the information necessary to implement timely measures to prevent traffic accidents and reduce accident risk, thereby enhancing public safety. Given these factors, improving the accuracy of traffic accident risk prediction has become an urgent issue in academic research. As such, this paper aims to explore an approach to enhance the precision of traffic accident risk prediction, ultimately providing insights and recommendations to optimize the effectiveness of such models further.

In the early stages of research, researchers employed statistical techniques in traffic prediction studies to anticipate upcoming traffic scenarios by examining the patterns in historical time-series data. For example, the Autoregressive Integrated Moving Average (ARIMA) model [1] has been implemented to predict short-term flows during peak hours by analyzing historical and real-time data. Meanwhile, the HA model [2] was used to study time series prediction performance. However, extracting non-linear features using statistical methods in complex, large-scale traffic data was difficult. As a result, scholars have begun to use Machine Learning (ML), which can capture non-linear relationships, for complex traffic data modeling. Olutayo et al. [3] analyzed various factors related to the severity of traffic accidents using decision trees. Tamerius et al. [4] studied non-linear features that affect traffic accidents by analyzing the interaction between rainfall and traffic flow. Yu et al. [5] combined unsupervised learning feature extraction methods with supervised learning feature classification methods and proposed the H-ELM algorithm, which utilizes deep features to enhance the precision of traffic accident forecasting. Although ML has improved predictive performance to some extent, capturing the time correlation of traffic accidents requires further improvement and is heavily dependent on feature engineering, which ultimately affects the reliability of traffic collision risk estimation.

To fully capture the nonlinear time correlation of traffic accident sequence data and further improve the performance of traffic accident risk prediction, Deep Learning has gradually been applied to this field, providing new ideas for traffic accident risk prediction. Various neural network methods, including Recurrent Neural Networks (RNN), have been used to predict road accidents [6], to extract time series correlation. While it excels in handling short-term time series, it encounters issues like vanishing gradients when dealing with longer time series. Therefore, some scholars have tried to use Long Short-term Memory Networks (LSTM) [7] to capture the long-term dependencies of traffic prediction. Ren et al. [8] constructed a traffic accident risk prediction model based on LSTM by quantitatively analyzing traffic accident data and capturing its temporal cyclicity. Although LSTM shows certain advantages in capturing time correlation

[9][10]. These approaches emphasize the importance of individual data points in traffic data but often neglect the impact of non-linear spatial relationships among nodes within the framework of the road network.

To more effectively capture the complex non-linear spatial correlations within road networks, some researchers have used Convolutional Neural Networks (CNNs) to address the challenges of traffic prediction. For example, some scholars can display traffic flows as images to predict flow and speed over large areas of the network [11]. Moreover, several researchers [12][13] have attempted to integrate CNNs with RNN and its enhanced models for road accident prediction, attaining favorable outcomes. Although CNN performs better in feature extraction in Euclidean space, traffic road networks have non-euclidean correlations, making it challenging to extract the correlations between road sections fully. Therefore, related scholars have introduced Graph Convolutional Neural Networks (GCN) to apply spatial correlations of topological structures, which are more suitable for traffic accident prediction. Chen et al. [14] proposed the MRA-BGCN model to extend convolution to more general graph structure data. At the same time, STGCN [15] represented the problem with a graph and captured spatial correlations through a complete convolution structure, achieving better performance in capturing spatial features than CNN. However, the above research has yet to capture spatio-temporal features fully, so fully utilizing spatiotemporal features is the key to enhancing the precision of traffic accident risk forecasting.

To capture spatiotemporal features more comprehensively, relevant scholars have conducted in-depth research. Bao et al. [16] proposed a spatiotemporal convolutional Long Short-Term Memory Network (STCL-Net) by dividing the city into grids of different sizes, which can predict traffic accidents on a weekly, daily, and hourly basis, and effectively capture the short-term spatiotemporal features of the city. Du et al. [17] introduced a Hybrid Multiple DL Framework (HMDLF) for short-term traffic flow prediction, effectively dealing with the intricate nonlinear space-time relationships in multimodal traffic data. However, the studies above only capture local spatiotemporal features, ignoring the mutual influence of non-adjacent regions in the urban road network.

Therefore, scholars extended the local area to the global area. Sun et al. [18] proposed an end-to-end Global Spatiotemporal Graph Attention Network (GST-GAT) that captures dynamic spatiotemporal correlations through "global interaction + node query," improving the accuracy and speed of traffic prediction. Chen et al. [19] studied the impact of unevenly distributed traffic conditions in different spaces (suburbs and urban areas) on traffic flow prediction by considering the complexity of different prediction tasks. Wang et al. [20] applied GCN to model global multi-scale spatiotemporal correlations and similarities. Pan et al. [21] proposed a ST-MetaNet model based on deep meta-learning, which predicts traffic flow for all locations at once. Zhang et al. [22] combined GCN with LSTM based on the road network structure, using actual road segments as spatial prediction units, effectively capturing spatiotemporal correlations of traffic accident risks. Compared with previous research, some scholars have extended the prediction unit to the global area, effectively capturing spatiotemporal features and improving the accuracy of traffic accident risk prediction.

However, the above methods capture only static spatiotemporal features and cannot fully reflect the actual traffic conditions in the road network, weakening the role of long-term temporal dynamics in traffic accident risk prediction.

Currently, the Attention Mechanism (AM) has been employed in traffic forecasting because of its remarkable capacity for capturing temporal dependencies. It can handle long sequential data better than RNN and avoid issues such as vanishing and exploding gradients. In traffic accident forecasting, the AM can bolster the model's capability to extract the continuance and periodicity present in time series data. For example, some researchers have combined the AM with LSTM to enhance the performance of series prediction [23]. Guo et al. [24] proposed the AM-based Spatio-Temporal Graph Convolutional Network (ASTGCN) model, which investigated the short-term, daily, and weekly time dependencies and captured the dynamic space-time correlations that affect traffic flow. Zhang et al. [25] introduced a Structure Learning Convolution (SLC) module based on Attention Networks, which fully captured the dynamic features that affect traffic flow prediction. Zheng et al. [26] employed an architecture combining an encoder-decoder structure with multiple spatio-temporal attention blocks to capture the influence of time and space dimensions on traffic conditions. They proposed a Graph Multi-Attention Network (GMAN) model, effectively forecasting traffic conditions for various nodes in future time steps. Yao et al. [27] combined GCN with AM, allocating weights to nodes of different impact levels and incorporating node-adaptive learning to extract spatio-temporal features effectively. Despite the existing methods combining GCN with Attention Mechanism, which can capture dynamic spatio-temporal features during strict periodic periods well, they neglect the combination of traffic sequences in neighboring and period-shifted periods, resulting in inadequate capture of dynamic spatio-temporal correlations across multiple regions.

The studies mentioned above have employed various approaches to capture the spatiotemporal characteristics that impact traffic accident risk by partitioning the road network into regions or using individual road segments as prediction units, thus enhancing the performance of road accident risk prediction to some extent. However, they must still integrate the space and time characteristics of adjacent regions and the changeable non-linear spatio-temporal correlations of multiple neighboring regions into prediction. Instead, most scholars have only considered local or global regions, limiting the capacity to capture non-linear spatiotemporal features fully. Furthermore, concerning the temporal correlation that affects traffic accident risk, most scholars have yet to consider the shift of long-period segments, only focusing on strict periodicity, which weakens the continuance and periodicity of the temporal sequence and reduces the accuracy of traffic collision hazard.

To overcome these difficulties, this paper presents a model, namely the Attention-based Temporal and Cross-regional Graph Convolutional Network (ATCGCN), which integrates spatiotemporal features from multiple regions to predict future traffic accident risk. The model takes adjacent time periods and long-period shifted traffic sequences as input. The spatiotemporal features of adjacent regions are captured

using a combination of CNN and GRU. In contrast, multiple neighboring regions' dynamic non-linear spatiotemporal correlations are captured using an Attention-GRU-GCN model. Finally, the spatiotemporal features of the two different regions are weighted and fused to reduce the impact of unnecessary features, such as little fog and low-traffic volume roads, on the model prediction, thereby improving the accuracy of traffic collision hazard prediction.

The primary contributions of this study are as follows:

(1) The fusion of spatiotemporal features from adjacent regions and the dynamic nonlinear spatiotemporal correlation from multiple neighboring regions enhance the prediction of traffic collision hazards. The model effectively captures both short-range and extended-range temporal dependencies, along with spatial characteristics like road hierarchy and connectivity, across multiple scales.

(2) The Dynamic Time Warping (DTW) algorithm [28] and the road network spatial graph are utilized to build a spatiotemporal relation graph. The adjacency matrix is updated into the Laplacian matrix used in GCN to more effectively represent the space relationships within the road network.

(3) By integrating the periodic shift of peak periods and the neighboring traffic sequence and adopting the Attention mechanism, this study dynamically expresses various spatio-temporal features.

(4) Experimental results on NYC and Chicago datasets demonstrate that the ATCGCN model outperforms other existing methods regarding prediction performance.

## II  THEORETICAL BASIS

### A. Problem Description

The road network graph $G$ is constructed by partitioning the similar road segments in the network into $a \times b$ regions. An undirected graph $G = (V, E, A)$ is then formed by using these regions as nodes to describe the network topology of the road network. $G$ consists of a node set $V = \{v_1, v_2, \cdots, v_n\}$ with n nodes, an edge set $E = \{e_{12}, e_{13}, \cdots, e_{ij}\}$ representing the connectivity between regions $i$ and $j$, and an adjacency matrix $A \in \mathbb{R}^{n \times n}$ for the graph, which is constructed using $DTW(X_{v_i, a}, X_{v_j, a})$ to establish the similarity between two road segments.

When predicting traffic accident risk, we often define the severity of a traffic accident in a particular area during time interval t by the number of casualties in that area [29], which can classify into three categories: minor injuries, moderate injuries, and severe injuries.

This study aims to predict traffic collision hazards in all related neighboring areas during time interval $T+1$ by utilizing the historical data of time interval $T$. The model takes the features $X_T \in \mathbb{R}^{a \times b \times d}$ of all neighboring areas during time interval $T$ as input, where $d$ represents the feature dimension of each area, including traffic accident risk, weather conditions, traffic volume, POI, and other relevant features.
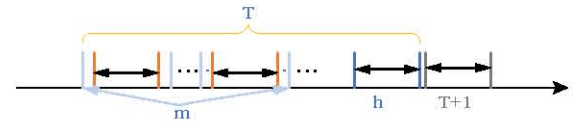


Fig. 1.  Time series

In addition, it is necessary to incorporate both adjacent time intervals and cyclically shifted time intervals. As shown in Fig. 1, this includes the $h$ adjacent time intervals, as indicated by the blue section, and the $m$ cyclically shifted time intervals, as indicated by the light blue section.
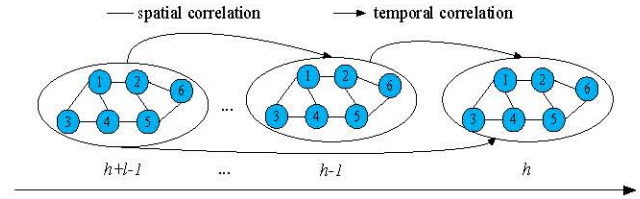


Fig.2 Temporal correlation

Fig. 2 illustrates that the traffic conditions in the previous time interval strongly correlate with those in the next time interval. The urban road network is complex, and road segments in a region have a high degree of connectivity, especially for short segments, which are prone to widespread congestion during peak hours. Moreover, the congestion dissipation rate is slow, significantly increasing the risk of traffic accidents. Additionally, POIs such as schools and shopping malls around roads can induce significant traffic flow, and there are mutual influences among them. For example, Node 1's traffic conditions are affected by Node 2, Node 5, and Node 6. In densely populated urban areas, even sub-adjacent nodes' traffic conditions can still affect each other, such as Node 5 having a particular impact on Node 1.

Hence, researchers partition the entire urban traffic network into various areas according to the city's latitude and longitude. The author uses traffic historical sequence data to analyze the spatiotemporal characteristics of nearby and multiple neighboring regions and predict the likelihood of traffic accidents in future time intervals. Finally, the research problem is to forecast the region feature graph $G_{T+1}$ at $T+1$ time based on the given region feature graph $(G_1, G_2, \cdots, G_T)$ for the previous $T$ time intervals, as shown in Eq. (1).

$$G_{T+1} = f[(G_1, G_2, \cdots, G_T), X_T] \qquad (1)$$

Where: $G_T \in \mathbb{R}^{a \times b \times d}$ is the regional feature map at the time $T$, $X_T \in \mathbb{R}^{a \times b \times d}$ is the region containing d-dimensional features, and $f$ is the road accident risk prediction model.

### B. Temporal Characteristics

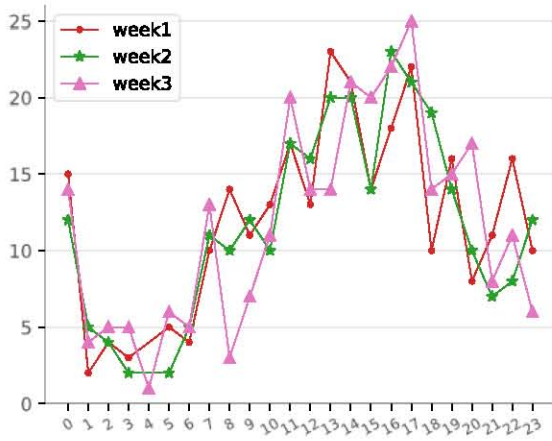Traffic accidents within a region are not only influenced by adjacent time intervals but also exhibit periodic trends.

Fig. 3.  Accident periodicity



Fig. 4.  Path matrix

Fig. 3 depicts the count of traffic accidents that took place on the same day for the previous three weeks, revealing a higher occurrence of accidents during daily peak hours, thus indicating the cyclic nature of traffic accident risk prediction.

However, capturing only long-period features of the same time interval may only partially reflect the long-term time dependence. Typically, the period between 11:00 and 14:00 is considered a peak period, with a higher risk of traffic accidents. However, as shown in Fig. 3, the number of accidents between 10:00 and 11:00 and 14:00 and 15:00 is also relatively high. Therefore, capturing only strict period features of the same time interval will not fully reflect the cyclic continuity of time features. Consequently, in addition to capturing strict period time correlation, it is necessary to research periodic shifts by adding Z time intervals to the same time interval of the previous m weeks. For example, for the period between 11:00 and 14:00, two-time intervals are added before and after two-time points, i.e., from 10:00 to 15:00, to capture long-term time dependence.

*C  Map of Space-time Relationships*

The DTW algorithm calculates the likeness between two-length time series data by employing dynamic programming. It constructs both one-to-many and many-to-one matches to minimize the cumulative distance between them, deriving the time series' similarity. This paper uses the idea to construct a time correlation matrix of traffic accident volume between any two nodes in the road network, where each node represents a segment of temporal sequence data. The author compute the distance between any two nodes $i$ and $j$ by denoting the traffic accident volumes of node $i$ and $j$ as $X_{v_i,a}$ and $X_{v_j,a}$, respectively, and using Eq. (2).

$$d(t,l) = |X_{v_i,T,a} - X_{v_j,T,a}| \tag{2}$$

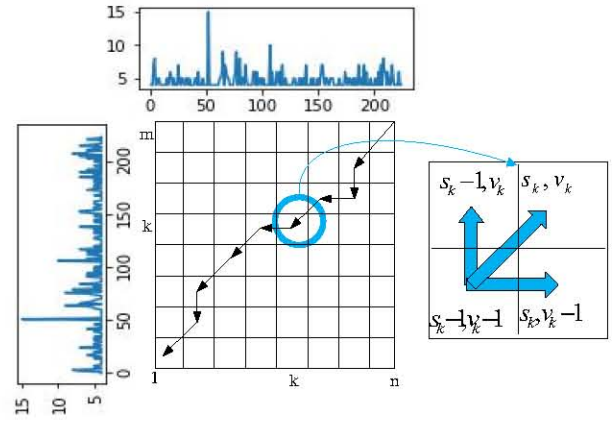Where: $X_{v_i,T,a}$ is the traffic incident volume of node $i$ at the period $T$.

As shown in Fig. 4, the distance obtained by Eq. (2) calculates the minimum path using dynamic programming, representing the similarity between two-time series.

$$L(s_k, v_k) = d(s_k, v_k) + \min \begin{bmatrix} L(s_k, v_k - 1) + \\ L(s_k - 1, v_k) + \\ L(s_k - 1, v_k - 1) \end{bmatrix} \tag{3}$$

$$DTW(X_{v_i,a}, X_{v_j,a}) = min[L(X_{v_i,a}, X_{v_j,a})] \tag{4}$$

Where: $(s_k, v_k)$ represents the coordinates of the points passed in the path matrix shown in Fig. 4, and $L(s_k, v_k)$ denotes the minimum distance at the $(s_k, v_k)$ coordinates, thus obtaining the time correlation matrix, as shown in Eq. (5).

$$\begin{bmatrix} DTW(X_{v_1,a}, X_{v_1,a}) & \cdots & DTW(X_{v_1,a}, X_{v_n,a}) \\ \vdots & \ddots & \vdots \\ DTW(X_{v_n,a}, X_{v_1,a}) & \cdots & DTW(X_{v_n,a}, X_{v_n,a}) \end{bmatrix} \tag{5}$$
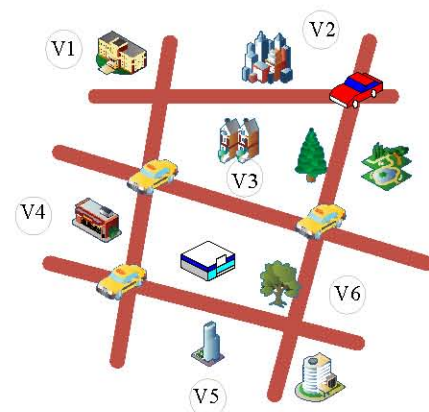


Fig. 5   Traffic road network map

To more accurately represent how road networks change over time, this research uses the time correlation matrix obtained from Eq. (5) in conjunction with the spatial road network (as depicted in Fig. 5) to create a temporal graph denoted as S. V represents the set of nodes in the spatial road network graph. D is the set of temporal similarity degrees between nodes. The smaller the temporal similarity degree

value, the closer the traffic accident volumes between the two nodes. By setting a given similarity degree threshold, any two nodes can be judged whether they are similar, thus establishing a temporal correlation matrix as expressed in Eq. (6).

$$A_T(v_i, v_j) = \begin{cases} 1, & DTW(X_{v_i,a}, X_{v_j,a}) < S_{threshold} \\ 0, & others \end{cases} \quad (6)$$

Where: $A_T(v_i, v_j)$ is the temporal correlation matrix, and $DTW$ is the temporal similarity between nodes.
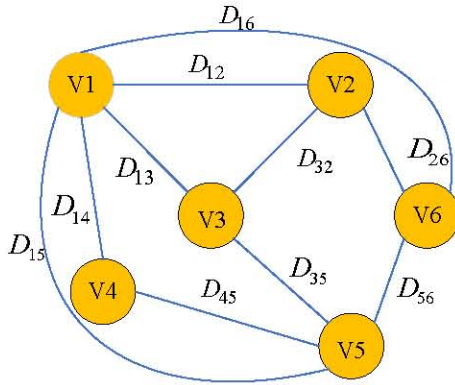


Fig. 6.   Spatio-temporal relationship diagram

To capture the spatial relationship of nodes within the context of spatiotemporal relationships, we propose constructing a spatiotemporal relationship graph $S_g(V, A_g)$ by combining a temporal and spatial graph. Specifically, $V$ represents the node set in the spatial graph, while $A_g$ signifies the combined adjacency matrix of the spatial graph integrated with the temporal correlation matrix. Using $S_g$, we acquire the spatial similarity of the transportation grid and subsequently select the top K nodes with the highest similarity scores to construct an undirected road network graph that belongs to a certain region, as depicted in Fig. 6.

$$A_g(v_i, v_j) = \begin{cases} 1, & A(v_i, v_j) = 1 \, or \, A_T(v_i, v_j) = 1 \\ 0, & others \end{cases} \quad (7)$$

Where: $A(v_i, v_j)$ is the adjacency matrix of the transportation grid diagram and $A_g(v_i, v_j)$ is used to update the Laplacian matrix of the GCN.

### III   MODEL BUILDING

#### A. Model Framework

Fig. 7 depicts the proposed ATCGCN model framework in this study. The ATCGCN model comprises four inputs: traffic network, temporal features, vehicle collision volume, and outside influences. The road network has spatial correlations, and this study extracts spatial features of adjacent regions and multiple neighboring regions separately. The spatial features of adjacent regions use the grid area as the input of CNN. In contrast, multiple neighboring regions employ GCN to capture spatial features, and the proximity

matrix of the space-time relationship graph replaces the Laplacian matrix of GCN. In the temporal feature extraction, neighboring periods are considered, and long-period shifted periods are incorporated, providing near-term and long-range temporal relationships for forecasting traffic accident susceptibility. Moreover, the AM is used in multi-neighborhood areas to capture the dynamic spatial and temporal interrelations of neighboring periods and periodic shift periods. Traffic accident volumes can serve as the basis for determining accident risk levels and calculating road section similarities. This study also incorporates outside influences such as weather, temperature, and POI to boost the model's capacity for generalization. Finally, the spatio-temporal correlations of two different spatial neighborhoods are weighted and fused, and the forecast is generated via a densely connected layer.
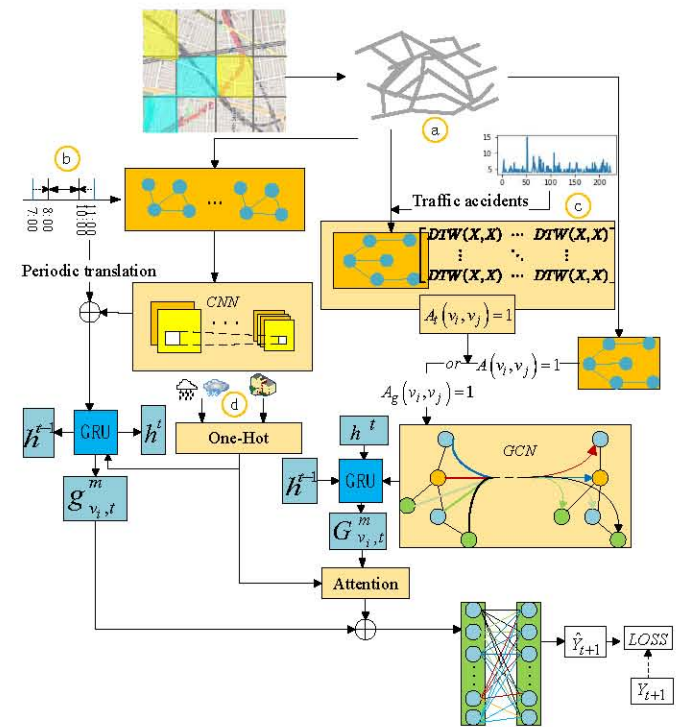


Fig. 7.   ATCGCN-Risk model framework

#### B. Capture of Adjacent Areas Features

##### B.1 Spatial Feature

Since regions tend to have similar POI and road features [30] and adjacent regions are directly connected and features relatively concentrated, we use regions as nodes and utilize local CNNs to capture adjacent region features, achieving extraction of spatial correlations. We divide the target region $i$ and its adjacent regions at period $T$ into $a \times b$ size and take the features $X_{v_i, T}$ of the region as inputs. Eq. (8) defines the convolutional layer of the local CNN.

$$X_{v_i, T}^k = ReLu\left(W^k * X_{v_i, T}^{(k-1)} + b^k\right) \quad (8)$$

Where: $W^k$ and $b^k$ are learnable parameters, and the spatial correlation of the adjacent region $X_{v_i, T}^k$ are output after the

convolution of the stacked $k$ layers.

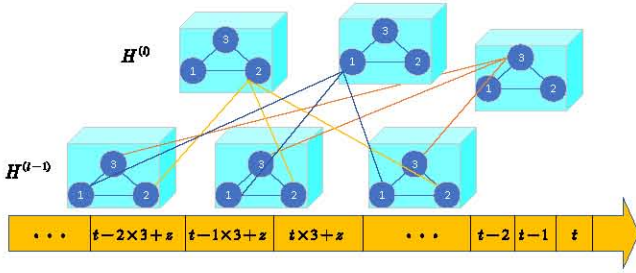### B.2. Spatio-temporal Correlation



Fig. 8.    Cycle translation status

To apprehend the time-related associations of neighboring and periodically shifted periods, the authors employed a GRU to grasp the time-based dependencies of road accident risks over time. As shown in Fig. 8, we utilize the features of adjacent periods and z-period shifted cycles to predict the traffic accident risk of period $T+1$. Specifically, they extracted the features of $h$ periods, such as $t-0$, $t-1$, and $t-2$, as the features of neighboring periods and the features of $m$ periods, such as $t \times 3+z$, $t-1 \times 3+z$, and $t-2 \times 3+z$, as the features of periodically shifted periods. Finally, $T(h+m)$ period information was input into the GRU module to acquire near-term and far-term time-related dependencies. Eq. (9) describes the computational process of the GRU.

$$g_{v_i,T}^{m,z} = GRU\left([X_{v_i,T}^{m,z}; P_{v_i,T}^{m,z}], g_{v_i,T}^{m,z-1}\right) \qquad (9)$$

Where: $X_{v_i,T}^{m,z}$ represents the spatial feature output of region $i$ at time $T$, while $P_{v_i,T}^{m,z}$ represents outside influences such as weather and holidays. $g_{v_i,T}^{m,z}$ denotes the spatiotemporal feature of the adjacent region $i$ after a periodic shift of $z$ periods.

### C. Capture of Multi-neighbor Regions Features
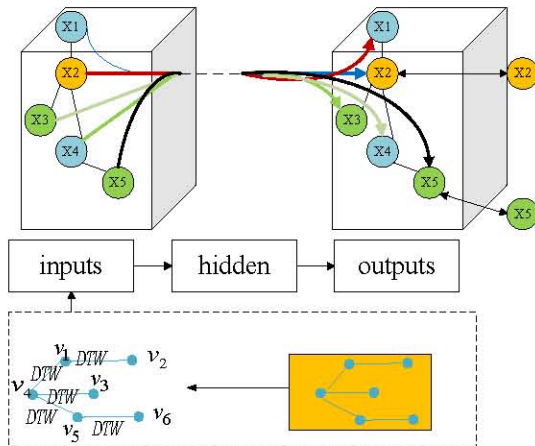
### C.1. Spatial Feature



Fig. 9.    Spatial features of multi-neighborhood areas

The urban road networks' structural characteristics are similar to the graphs' topological structure. However, due to the geographical conditions of roads and the impact of surrounding infrastructure, the spatial influence between some adjacent regions may be weaker than between non-adjacent regions. Meanwhile, traffic congestion or temporary traffic control in non-adjacent areas may significantly impact adjacent downstream areas. Therefore, we employ the idea of DTW to divide the road network into multiple regions and combine it with GCN to capture spatial correlations. As shown in Fig. 9, this approach captures the spatial features of multiple neighboring regions.

In Part C of Section II, the authors constructed a spatiotemporal relationship graph based on the DTW algorithm. They updated the GCN Laplacian matrix with $A_g(v_i, v_j)$ to extract spatial dependencies among multiple neighboring regions based on GCN features. Eq. (10) provides the formula for GCN computation.

$$f = RELU\left(A_g RELU\left(A_g X_T W^0 + b^0\right)W^1 + b^1\right) \quad (10)$$

Where: $A_g \in R^{n \times n}$ is the adjacency matrix constructed by $A(v_i, v_j)$ and $A_t(v_i, v_j)$, $W^0 \in \mathbb{R}^{d \times d_e}$ is the learnable parameters of layer 0, $b^0 \in \mathbb{R}^{d_e}$ is the adjustable parameters of the layer 0, $d_e$ is the convolution kernel size, $f \in \mathbb{R}^{a \times b \times d_e}$ is the spatial feature output of the multi-neighborhood region, and the activation function is $RELU$.

### C.2. Dynamic Spatio-temporal Correlation

Similarly, after obtaining the spatial correlations among multiple neighboring regions, the authors utilized GRU to capture the time characteristics of adjacent periods and periodically shifted periods. They also use Attention to compute node weights, enabling the dynamic expression of spatiotemporal features and ultimately improving the accuracy of vehicle collision hazard forecasting, as shown in Eq. (11).

$$G_{v_i,T}^{m,z} = GRU\left([f_{v_i,T}^{m,z}; P_{v_i,T}^{m,z}], G_{v_i,T-1}^{m,z}\right) \qquad (11)$$

Where: $f_{v_i,T}^{m,z} \in \mathbb{R}^{d_i}$ is the output of GCN, $d_i$ indicates the quantity of kernels in the $l$ layer convolution kernel, while $G_{v_i,T-1}^{m,z-1}$ corresponds to the output from the previous moment, $G_{v_i,T}^{m,z}$ is the region's output in the period $T$, and $P_{v_i,T}^{m,z}$ indicates the external factors.
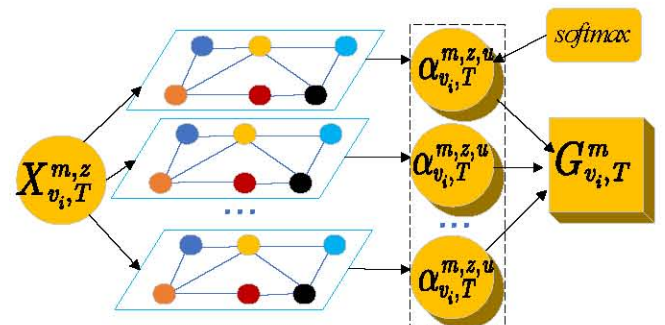


Fig. 10.    Attentional mechanism

In actual road networks, spatiotemporal features are dynamically changing, and neighboring nodes have different degrees of influence on the target node, which can be adaptively captured by attention to capture the correlations between regions. As shown in Fig. 10, different weights are dynamically assigned to diverse nodes (regions) at varying time steps. Additionally, we calculate the weighted sum of node $v_i$ and provide its new feature vector in Eq. (12).

$$G_{v_i,T}^m = \sigma\left(\sum_{v_i \in V, z \in Z} \alpha_{v_{ij},T}^{m,z} * W * G_{v_j,T}^{m,z}\right) \quad (12)$$

Where: $\alpha_{v_{ij},T}^{m,z}$ denotes the attention score of node $v_i$ at the period with periodic translation $z$, $W$ is a learnable parameter, $\sigma$ is the activation function, and $G_{v_i,T}^m$ is the output of dynamic spatiotemporal correlations among multiple neighboring regions.

Eq. (13) is the weight factor of the node.

$$\varepsilon_{v_{ij},T}^{m,z} = a(WG_{v_i,T}^{m,z}, WG_{v_j,T}^{m,z}) \quad (13)$$

Where: $a(\cdot)$ denotes the function used to compute the relationship between a pair of nodes and $\varepsilon_{v_{ij},T}^{m,z}$ is the weight coefficient of adjacent nodes $v_i$ to $v_j$.

Eq. (14) is normalized by softmax.

$$\alpha_{v_{ij},T}^{m,z} = \frac{exp\left(\varepsilon_{v_{ij},T}^{m,z}\right)}{\sum_{v_i \in V, z \in Z} exp\left(\varepsilon_{v_{ij},T}^{m,z}\right)} \quad (14)$$

At this point, $\alpha_{v_{ij},T}^{m,z}$ is brought into Eq. (12) to update the new feature vector.

To fully capture the dynamic spatio-temporal correlation, we employ a multi-headed attention mechanism by combining vectors of different single attentions, as shown in Eq. (15). This mechanism allocates attention to the target node. Also, it considers neighboring nodes, thus avoiding excessive attention concentration on the self-position.

$$G_{v_i,T}^m = \|_{u-1}^U \sigma\left(\sum_{v_i \in V} \alpha_{v_{ij},T}^{m,z,u} * W^u * G_{v_j,T}^{m,z}\right) \quad (15)$$

Where: $\alpha_{v_{ij},T}^{m,z,u}$ is the importance of the $u$ head attention mechanism, and $W^u$ is the learnable parameter.

### D. Fusion Prediction

We perform a weighted combination to fuse the spatiotemporal correlations of two different spatial neighborhoods. The prediction unit preserves the spatial and temporal characteristics of adjacent regions and the dynamic spatio-temporal features of multiple neighboring regions. As a result, a fully connected layer is utilized to capture the complex nonlinear relationships better, as shown in Eq. (16).

$$\hat{Y} = FC(g_{v_i,T}^m * W_1 + G_{v_i,T}^m * W_2) \quad (16)$$

Where: $W_1$ and $W_2$ are the adjustable parameters, $FC$ is the fully connected layer, and $\hat{Y}$ represents the traffic accident risk of all relevant neighborhoods in the period

$T+1$.

### D.1. Loss Function

Throughout the learning phase, to make the predicted results as close to the ground truth as possible, we adopt a loss function to reduce the error. Also, assign weights to samples with different traffic accident risk values during calculation. Eq. (17) is the loss function.

$$LOSS(Y, \hat{Y}) = \frac{1}{2} \sum_{j \in J} \mu_j \left(Y(j) - \hat{Y}(j)\right)^2 \quad (17)$$

Where: $Y$ denotes the observed value, $\hat{Y}$ is the model output value, and $\mu_j$ means the traffic accident risk-weighting factor.

## IV EXPERIMENTAL ANALYSIS

### A. Experimental Data

This paper utilizes two traffic accident datasets, the NYC and Chicago datasets, which contain information on the latitude and longitude of the accidents, time, and number of casualties. Since taxis often leave one area and enter another after some time, the number of times they leave or arrive at a certain area is used as the traffic flow information. The risk level of vehicle collisions is divided according to the number of casualties. Other discrete data, such as POI data, weather conditions, and holidays, are normalized through One-Hot encoding. Table 1 provides an overview of the datasets.

TABLE I DATASET OVERVIEW TABLE

| Data Type | Features |
|---|---|
| Accident Data | Longitude and latitude, number of casualties, etc. |
| Traffic flow | Number of cab inflows/outflows |
| POI Data | Schools, shopping malls, neighborhoods, etc. |
| Weather | Temperature, humidity, etc. |
| Holiday | Yes, No |
| Risk Level | Mild, Moderate, and Severe |
| Peak time | 8:00-13:00, 16:00-21:00 |

### B. Evaluation Indicators

We employ six evaluation metrics in this study: Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Recall, Precision, and F1-Score. The specific definitions are as follows:

$$MSE = \sum_{i=1}^n \left(y_i - \hat{y}_i\right)^2 / n \quad (18)$$

$$MAE = \sum_{i=1}^n |y_i - \hat{y}_i| / n \quad (19)$$

$$RMSE = \sqrt{\sum_{i=1}^n \left(y_i - \hat{y}_i\right)^2 / n} \quad (20)$$

$$Recall = TP/TP + FN \quad (21)$$

$$Precision = TP/TP + FP \quad (22)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (23)$$

Where: $y_i$ is the observed value, $\hat{y}_i$ is the output value, TP is the traffic accident that occurred, and the prediction is correct, FN is the traffic accident that occurred, but the forecast is wrong, FP is the traffic accident that did not occur, but the forecast occurred.

### C. Parameter Setting

This study used the PyTorch framework to implement the proposed model. We split the dataset into three sets: 60% for training, 20% for validation, and 20% for testing. We set the size of each region to $1km \times 1km$, the adjacent period $T = 3$, and the long period $m$ set to 3. We designed the GRU with two layers and set the number of road similarities for building an undirected road network graph to $K = 10$, with two convolutional layers. We set the periodic translation periods to $z = 0, z = 1$, and $z = 2$. The learning rate was set to 0.0001, batch size to 16, epoch to 20, and step to 350. Additionally, we set the weights for different risk levels to {0.05, 0.25, 0.50}.

### C.1. Parameter Analysis

The selection of hyperparameters is crucial for network training. As a result, this study conducted further optimization analysis on the GRU hidden layer unit number, GCN filter number, batch size, and epoch on the Chicago dataset.

(1) The count of hidden units within the GRU layer has an impact on the efficiency of the model.
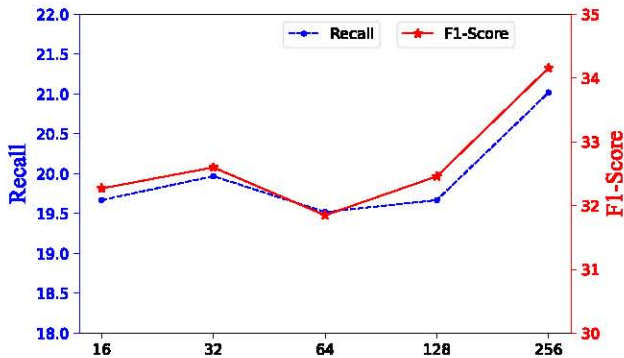
Fig. 11.    Number of hidden layer cells

We configured the GRU with 16, 32, 64, 128, and 256 hidden units, respectively. Fig. 11 illustrates that with 256 hidden units, both Recall and F1-Score evaluation metrics outperformed those with other quantities of units. Consequently, we opted for 256 as the count of hidden units in the GRU.

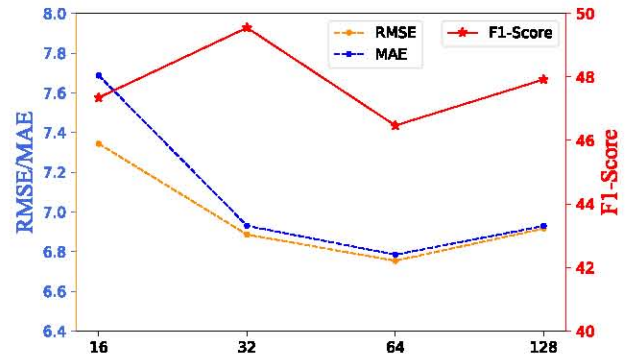(2) The effect of GCN filter count on the model.

Fig. 12.    GCN-filters

As shown in Fig. 12, when we set GCN-filters to 64, the RMSE and MAE were at their minimum, and the F1-Score was also the minimum, with a value of 46.47%. However, when GCN-filters were set to 32, its RMSE and MAE were the second smallest, and its F1-Score was the highest, reaching 49.55%. Therefore, setting GCN-filters to 32 made the ATCGCN model optimal.

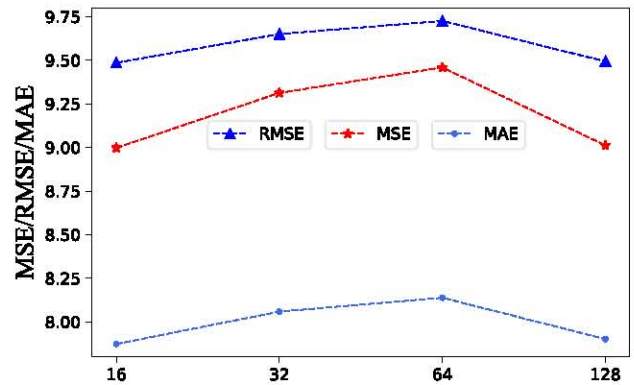(3) The effect of Batch-size on the model.

Fig. 13.    Batch-size

Based on the results in Fig. 13, increasing batch size leads to higher computational cost and lower training performance due to a larger input size per batch. On the contrary, using a batch size of 16 allows for a smaller input size per batch, which improves gradient descent accuracy and training efficiency. Therefore, a batch size of 16 is the optimal choice to achieve the best performance for the ATCGCN model.

(4) The effect of epoch value on model training.

It is crucial to halt the training throughout the model training phase when the loss function no longer demonstrates substantial alterations. Doing so can improve the training efficiency while preventing the propagation of erroneous features to subsequent predictions. In this study, we trained the model on the Chicago dataset for 17, 18, 19, and 20 epochs and analyzed the variations in the loss function to determine the optimal training steps.

Based on the results presented in Fig. 14a, 14b, and 14c, it can be inferred that the loss value exhibits a relatively stable trend when the step size is approximately 200 for each epoch, whereas increasing the step size results in fluctuations in the loss value.

(a) Epoch value of 17



(b) Epoch value of 18



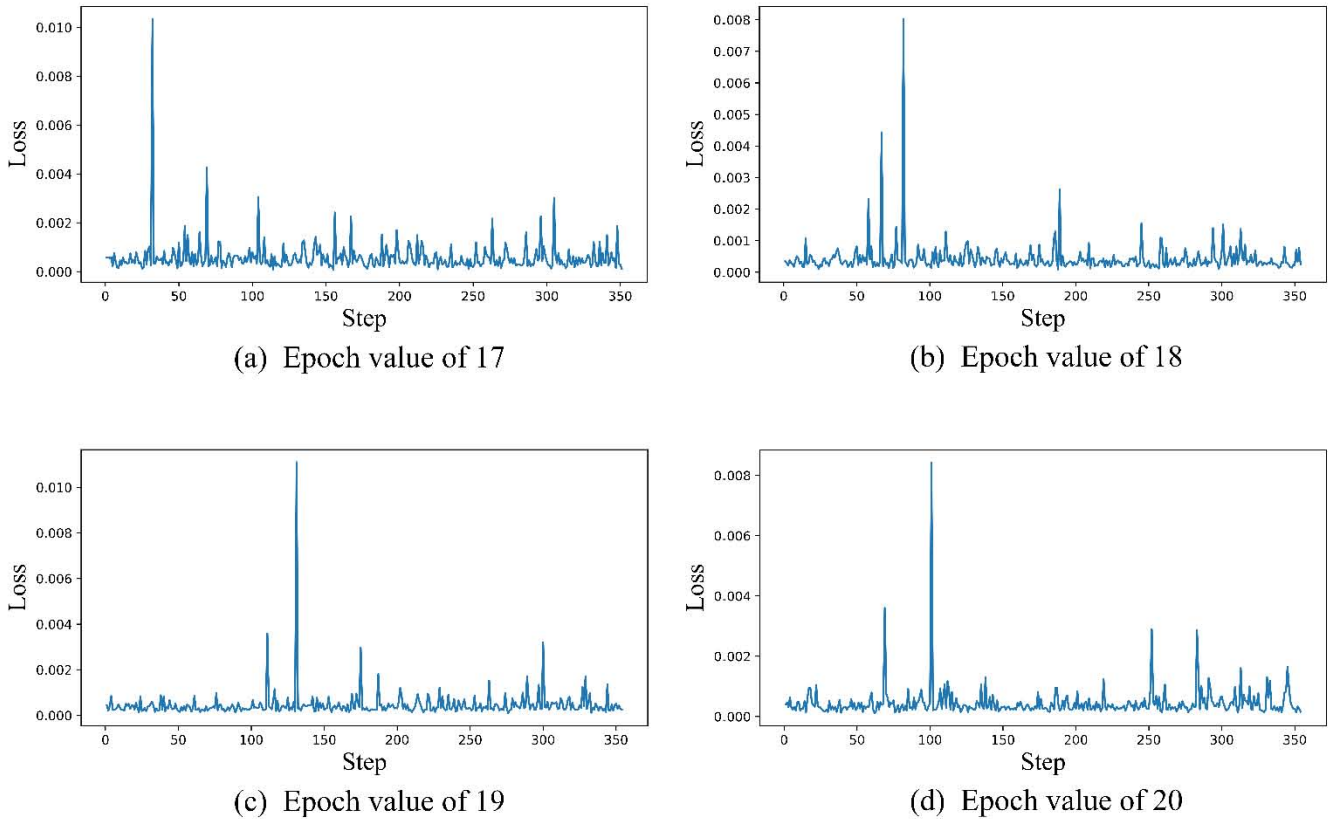(c) Epoch value of 19



(d) Epoch value of 20

Fig. 14. Variation of loss for different values of epoch

Hence, optimal training performance is attained with a step size of approximately 200, independent of the epoch. Specifically, when the epoch is 19, the training loss value increases smoothly with a step size of around 200. Moreover, Fig. 14d indicates that when the epoch is 20, and the step size is around 200, the trend of the loss value remains stable for a longer period, suggesting that the model has reached its optimal performance on the training set.

(5) Effect of epoch value on model validation.

After analyzing the influence of epoch value on model training, we determined that the model reaches its best performance with an epoch value of 20. We carried loss function experiments on 20% of the data randomly selected from the evaluation set to verify whether the training set has genuinely reached the optimal state.
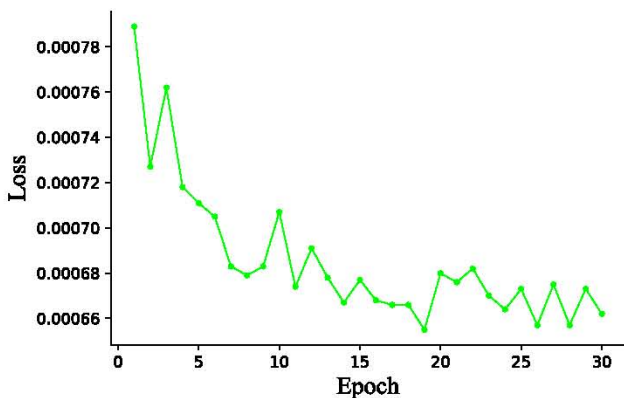
As shown in Fig. 15, the loss value decreases in the epoch range of 15 to 20, reaching its minimum when the epoch value approaches 20. Therefore, our model achieves optimal performance when the epoch value is set to 20.

### D. Results Analysis

#### D.1. Long-period Advection Correlation Analysis

(1) NYC data set validation

Experiments were executed using the NYC dataset to assess the effectiveness of our model in forecasting traffic accident risks following a temporal shift. Specifically, we considered three scenarios: $z = 0$ for strict periodicity, $z = 1$ for periodicity shifted by one-time intervals, and $z = 2$ for periodicity shifted by two-time intervals.
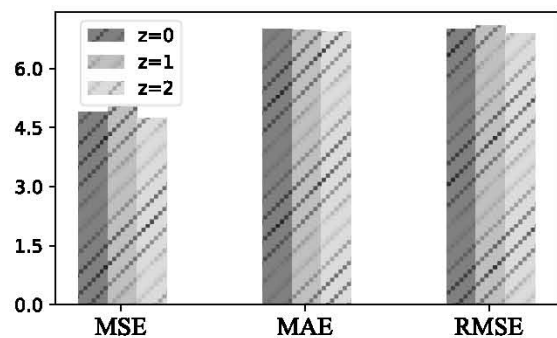


Fig. 15. Loss variation of the validation set



Fig. 16. Comparison of translation periods

Fig. 16 shows that when $z = 2$, the error metric is relatively lower than the other two shift periods, indicating superior performance. Although the MAE value for $z = 2$ is slightly lower than the other two shift periods, it has little impact on the results, as traffic accidents do not follow a smooth increasing pattern. And extreme weather conditions will influence accidents, such as heavy rain and fog, lead to outliers in the accident dataset. Moreover, the MAE value shows a consistent downward trend. Therefore, the proposed model achieves optimal performance when $z = 2$.

(2) Chicago dataset validation

To further confirm the efficacy of accounting for periodic shifts in our suggested model (ATCGCN), we analyzed F1-Score and Precision on the Chicago dataset. Together with the analysis of the validation set's loss value shown in Fig. 15, we determined that the best efficiency of the model is obtained when the epoch is set to 24.
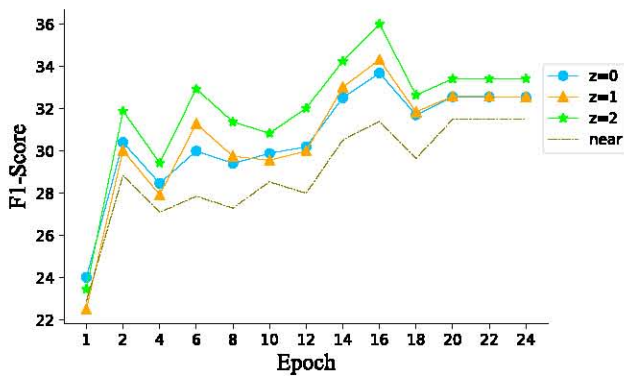


Fig. 17.    F1-Score

According to Fig. 17, the F1-Score gradually increases after each epoch of training, which serves as a comprehensive evaluation metric combining precision and recall. A higher F1-Score value indicates better performance of the proposed model. When $z = 2$, the F1-Score demonstrates a more favorable trend than other time periods. If only considering the neighboring traffic conditions without taking into account the periodicity, the F1-Score value remains lower than that of longer cyclic patterns. Thus, the proposed model achieves the optimal performance when $z = 2$.
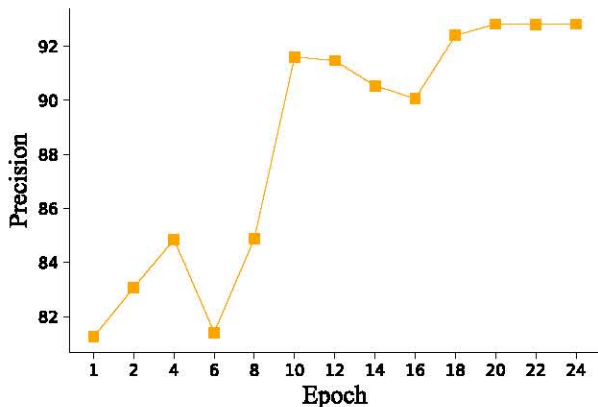


Fig.18.    Precision

As per the outcomes depicted in Fig. 18, it is noticeable that the model reaches the greatest precision at epoch 20 when $z = 2$, and it remains stable at epochs 22 and 24. It confirms the proposed model performs optimally when $z = 2$ and the epoch is 20. The precision performance also indicates that the model accurately predicts traffic accident risks.

*D.2. Comparative Analysis of Benchmark Models*

This study evaluates the proposed ATCGCN model by contrasting it with four established baseline models:
(1) Support Vector Machines (SVM) [31].
(2) Gated Recurrent Unit Network (GRU).
(3) Long Short-Term Memory Network (LSTM).
(4) T-GCN model [32], which models both spatial and temporal correlations.
(5) ConvLSTM model [33], which captures spatiotemporal features of local regions.

Table II displays the comparative outcomes between the ATCGCN and other models using the NYC dataset.

According to the experimental results in Table II, the proposed ATCGCN model outperforms other models in terms of MSE, MAE, RMSE, and Recall. The results indicate the relatively poor predictive performance of SVM, GRU, and LSTM models. when comparing the ATCGCN model and the LSTM model, the LSTM model's MSE is 0.675 higher than the ATCGCN model's, suggesting that neglecting spatial features while considering only temporal dependence leads to a decline in prediction accuracy. Although the T-GCN model improves the model's predictive accuracy by capturing spatial and temporal correlations, it ignores the translational of time series for the periodic window, resulting in an MAE 0.5660 higher than that of the ATCGCN model. The ConvLSTM model only considers local spatiotemporal features and fails to capture the dynamic spatiotemporal correlations of multiple neighboring regions, resulting in an RMSE 0.2296 higher than that of the ATCGCN model.

Given this, in this paper, we investigate the periodic shift of time series and integrate the spatio-temporal features of neighboring and multi-neighborhood regions. The ATCGCN model achieved superior predictive performance.

We used the same parameters and conducted experiments on the Chicago dataset. Table III demonstrates that the ATCGCN model has improved the prediction accuracy to a certain extent and outperformed other methods.
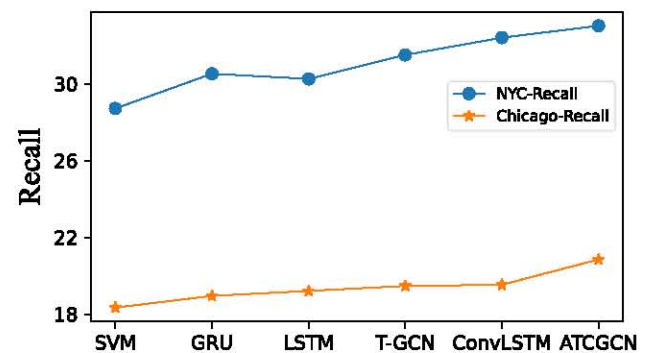


Fig. 19.    Recall of different models in NYC/Chicago

TABLE. II COMPARATIVE ANALYSIS OF VARIOUS MODELS ON THE NYC DATASET

| Evaluation index / Models | MSE | MAE | RMSE | Recall% |
|---|---|---|---|---|
| SVM | 5.5543 | 7.6142 | 7.6413 | 28.72 |
| GRU | 5.2596 | 7.4767 | 7.4273 | 30.53 |
| LSTM | 5.4164 | 7.5236 | 7.5448 | 30.26 |
| T-GCN | 5.3642 | 7.4961 | 7.4726 | 31.51 |
| ConvLSTM | 4.8793 | 7.2483 | 7.1154 | 32.41 |
| ATCGCN | **4.7414** | **6.9301** | **6.8858** | **33.03** |

Table. III COMPARISON OF MODEL PERFORMANCE ON THE CHICAGO DATASET

| Evaluation index / Models | MSE | MAE | RMSE | Recall% |
|---|---|---|---|---|
| SVM | 7.8261 | 8.3175 | 8.7823 | 18.34 |
| GRU | 7.6109 | 81673 | 8.6764 | 18.95 |
| LSTM | 7.7052 | 8.1726 | 8.7462 | 19.21 |
| T-GCN | 7.6275 | 8.0014 | 8.6244 | 19.47 |
| ConvLSTM | 7.3173 | 7.8531 | 8.5846 | 19.53 |
| ATCGCN | **7.1572** | **7.7136** | **8.4600** | **21.02** |

As shown in Fig. 19, the recall of our proposed method on both datasets exhibits an upward trend, indicating that the correctness of predicting positive traffic accident samples gradually improves. This experiment verifies the generalization ability of the ATCGCN. Therefore, the model can be utilized to predict road accident risk in different cities.

### D.3. Analysis of Ablation Experiments

(1) Different component validation analysis

To demonstrate the effectiveness of fully considering the spatiotemporal features of adjacent and multiple neighboring regions, three experiments were conducted on the NYC dataset in this study:

(a) TCGCN, a model without an attention mechanism.

(b) TACCN, a model that only considered the dynamic spatiotemporal features of adjacent regions.

(c) AGCN, a model that used strictly periodic data without using period-shifted data.
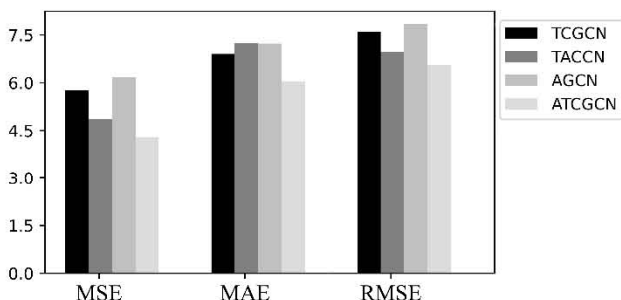


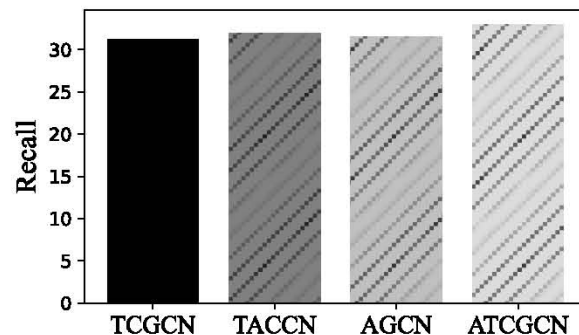Fig. 20. Comparison results of variant models



Fig. 21. Comparison results of variant models

Fig. 20 and Fig. 21 show certain changes in the evaluation metrics, including MAE, RMSE, and Recall, as different model components are progressively reduced. The results indicate that incorporating the attention mechanism and dynamic spatiotemporal features of multiple adjacent regions into the model can substantially enhance predicting traffic accidents.Moreover, the experiment results are unsatisfactory when removing the cyclically shifted data, as it weakens the long-period traffic accident volume trend. Therefore, all the above components can effectively enhance the precision of traffic accident risk forecasting.

(2) Different neighborhood validation analysis

To better demonstrate the effectiveness of fusing spatiotemporal correlations of adjacent and multiple adjacent regions, this paper analyzes the impact of captured spatiotemporal features in different regions on model accuracy through changes in error metrics on the NYC dataset.

a.    Adjacent area

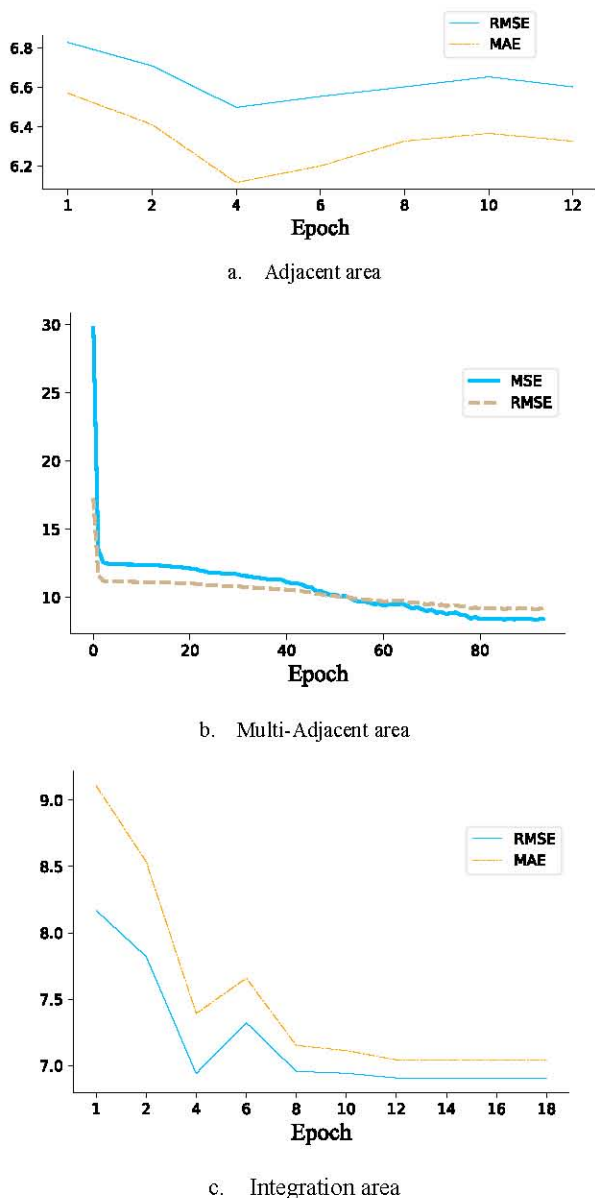

b.    Multi-Adjacent area



c.    Integration area

Fig. 22.    Variation of error in different neighborhoods

As shown in Fig. 22a, both RMSE and MAE show an upward trend when the epoch is 12. The experiment has already ended, indicating that considering only the spatiotemporal features captured by neighboring areas is insufficient. If we only consider the spatiotemporal features of multiple neighboring areas, the model will extract unnecessary features, leading to a longer training time, as shown in Fig. 22b. Only when the epoch value reached 80 did RMSE and other indicators drop below 10, which may lead to overfitting. As shown in Fig. 22c, by fusing the spatiotemporal features of these two types of areas for prediction, MAE and RMSE can converge quickly, thereby capturing more valuable features and improving the accuracy of the model's predictions.

## V    CONCLUSION

This study focuses on investigating the problem of vehicular collision hazard forecasting by capturing the spatiotemporal correlations across multiple regions. To this end, we propose the ATCGCN model, which combines the feature learning of adjacent and multiple neighboring regions to obtain the spatiotemporal correlations that impact traffic accident risk. Experimental outcomes show that the ATCGCN effectively overcomes the limitations of only considering adjacent or multiple neighboring regions and exhibits good convergence performance. Moreover, introducing the attention mechanism in multiple neighboring regions to capture the dynamic features of long-period shifts leads to more accurate traffic accident risk prediction, resulting in more consistent decision-making with real-world scenarios.

The sensitivity analysis of the model components shows different effects of incorporating different components into the ATCGCN model, which improves the prediction accuracy to different degrees. The primary characteristics include the following.

(1) To address the issue of time cycle continuity. We flatten the long period of the time series based on the adjacent times and dynamically represent the spatio-temporal characteristics by the AM. In this paper, we verify the effectiveness of period shifting on the NYC/Chicago dataset, respectively, and the experiments show that the model performance is optimal when $z = 2$.

(2) We obtain the TCGCN, TACCN, and AGCN models by removing different components to analyze the influence of various components on the model. The evaluation metrics indicate that considering the dynamic spatiotemporal features of multiple neighborhoods and the AM can enhance the prediction performance of the ATCGCN model. Additionally, experimental results were more favorable when incorporating a periodic shift.

(3) To enhance the precision of forecasting traffic accident risk. This study employed two model structures: CNN+GRU and GCN+GRU+Attention. The former captures the spatiotemporal features of adjacent areas, while the latter captures the dynamic nonlinear spatiotemporal correlations among multiple adjacent areas. Finally, the spatiotemporal features of the two models are weighted and fused. Compared to the baseline model, our model performs better in evaluation metrics such as MSE, MAE, and Recall.

Despite the ATCGCN model demonstrating an enhancement in prediction outcomes compared to the baseline model, its predictive scope is limited to the regional level. It has yet to be narrowed down to the level of individual road segments, which is a limitation of this study. Furthermore, factors such as driver inattention, pedestrian non-compliance with traffic rules, and other random events may contribute to traffic accidents in traffic networks. Therefore, future research must consider these stochastic factors to enhance the model's generalizability.

## REFERENCES

[1]    Kumar, S. Vasantha, and Lelitha Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, no. 3, pp. 1-9, 2015.

[2]    J. Y. Campbell and S. B. Thompson, "Predicting excess stock returns out of sample: can anything beat the historical average," *The Review of Financial Studies*, vol. 21, no. 4, pp. 1509–1531, 2008.

[3]    Olutayo, V A, Eludire A A, "Traffic accident analysis using decision trees and neural networks," *International Journal of Information Technology & Computer Science*, vol. 6, no. 2, pp. 22-28, 2014.

[4] Tamerius, J. D., Zhou, X., Mantilla, R., & Greenfield-Huitt, T. "Precipitation effects on motor vehicle crashes vary by space, time, and environmental conditions," *Weather, Climate, and Society*, vol. 8, no. 4, pp. 399-407, 2016.

[5] Yu J, Li B, Chen QM, "Research on traffic accident prediction method based on H-ELM," *Journal of Nanjing University (Natural Sciences)*, vol. 54, no. 5, pp. 896-903, 2018.

[6] Li, Yaguang and Yu, Rose and Shahabi, Cyrus and Liu, Yan. "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," *6th International Conference on Learning Representations*, arXiv.1707.01926, pp. 1-16, 2018.

[7] Zhang, Chaoyun; and Patras, Paul, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," *Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 231-240, 2018.

[8] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu, and Jinzhi Lei. "A deep learning approach to the citywide traffic accident risk prediction," *IEEE Conference on Intelligent Transportation Systems (ITSC)*, vol. 2018-November, pp. 3346-3351, 2018.

[9] Yu, Rose and Li, Yaguang and Shahabi, Cyrus and Demiryurek, Ugur and Liu, Yan, "Deep learning: a generic approach for extreme condition traffic forecasting," *Proceedings of the 17th SIAM International Conference on Data Mining*, pp. 777-785, 2017.

[10] Kong, Fanhui, and Li, Jian and Jiang, et al. "Big data-driven machine learning enabled traffic flow prediction," *Transactions on Emerging Telecommunications Technologies*, vol. 30.no. 9, issn. 21615748, 2019.

[11] Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4: issn. 23318422, 2017.

[12] Chen K, Chen F, Lai B, et al, "Dynamic spatio-temporal graph-based CNNs for traffic flow prediction," *IEEE Access*, vol. 8, pp. 185136 – 185145, 2020.

[13] Yang, Gang and Wang, Yunpeng and Yu, Haiyang and Ren, Yilong and Xie, Jindong, "Short-term traffic state prediction based on the spatiotemporal features of critical road sections," *Sensors*, vol. 18, no. 7, https://doi.org/10.3390/s18072287, 2018.

[14] Chen, W., Chen, L., Xie, Y., Cao, W., Gao, Y., & Feng, "Multi-range attentive bicomponent graph convolutional network for traffic forecasting," *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, pp. 3529-3536, 2020.

[15] Yu, Bing, Haoteng Yin, and Zhanxing Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2018-July, pp. 3634-3640, 2018.

[16] Bao, Jie, Pan Liu, and Satish V. Ukkusuri, "A spatiotemporal deep learning approach for city-wide short-term crash risk prediction with multi-source data," *Accident analysis & prevention*, vol. 122, pp. 239-254, 2019.

[17] Du, Shengdong and Li, Tianrui and Gong, Xun and Horng, Shi-Jinn, "A hybrid method for traffic flow forecasting using deep multimodal learning," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 85-97, 2018.

[18] Sun, Bin and Zhao, Duan and Shi, Xinguo and He, Yongxin, "Modeling global spatial-temporal graph attention network for traffic prediction," *IEEE Access*, vol. 9, pp. 8581-8594, 2021.

[19] Chen, Changlu and Liu, Yanbin and Chen, Ling and Zhang, Chengqi, "Bidirectional spatial-temporal adaptive transformer for urban traffic flow forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-13, 2022.

[20] Wang Beibei, Wan Huaiyu, Guo Shengnan, Lin Youfang, "Traffic accident risk prediction by integrating local and global spatio-temporal features," *Computer Science and Exploration*, vol. 15, no. 9, pp. 1694-1702, 2021.

[21] Pan, Zheyi and Liang, Yuxuan and Wang, Weifeng and Yu, Yong and Zheng, Yu and Zhang, Junbo, "Urban traffic prediction from spatio-temporal data using deep meta-learning," *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1720-1730, 2019.

[22] Zhang Yankong, Lu Jiapin, Zhang Shuaichao, Ji Xiaopeng, Luo Yuetong, Chen Wei, "A short-term risk prediction method for urban traffic accidents based on road network structure," *Journal of Intelligent Systems*, vol. 15, no. 04, pp. 663-671, 2020.

[23] Huang Sen, "Time series prediction based on improved deep learning," *IAENG International Journal of Computer Science*, vol. 49, no. 4, pp. 1133-1138, 2022.

[24] Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H, "Attention-based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 1, pp. 922-929, 2019.

[25] Zhang, Q., Chang, J., Meng, G., Xiang, S., & Pan, C, "Spatio-temporal graph structure learning for traffic forecasting," *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 1, pp. 1177-1185, 2020.

[26] Zheng, C., Fan, X., Wang, C., & Qi, J, "Gman: A graph multi-attention network for traffic prediction," *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 1, pp. 1234-1241, 2020.

[27] Xiaomin, Yao; Xinlan, Zhang; Zhengguo, Zhang, "Traffic flow prediction model based on time graph attention," *Research on Computer Application*, vol. 39, no. 3, pp. 770-773+ 779, 2022.

[28] Liu, Yi-Cheng and Li, Zhi-Peng and Lv, Chun-Pu and Zhang, Tao and Liu, Yan, "Network-wide Traffic Flow Prediction Research Based on DTW Algorithm Spatial-temporal Graph Convolution," *Journal of Transportation Systems Engineering and Information Technology*, vol. 22, no. 3, pp. 147-157+178, 2022

[29] Chen, Quanjun and Song, Xuan and Yamada, Harutoshi and Shibasaki, Ryosuke, "Learning deep representation from big and heterogeneous data for traffic accident inference," *30th AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, pp. 338-344, 2016.

[30] Lin, Z., Feng, J., Lu, Z., Li, Y., & Jin, D, "Deepstn+: "Context-aware spatial-temporal neural network for crowd flow prediction in the metropolis," *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 1, pp. 1020-1027, 2019.

[31] Do, Hung Ngoc and Vo, Minh-Thanh and Luong et al, "Speed limit traffic sign detection and recognition based on support vector machines," *International Conference on Advanced Technologies for Communications*, pp. 274-278, 2017.

[32] Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., & Li, H, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848-3858, 2019.

[33] Shi, Xingjian and Chen, Zhourong and Wang, et al, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, pp. 802-810, 2015.

**Qingrong Wang**, born in 1977, is a professor and master's supervisor, a teacher in the Department of Computer Science and Technology, whose main research interests are the application of big data and data mining in intelligent transportation.

**Kai Zhang**, the correspqonding author, was born in 1997 in Guizhou, China. He is currently a master student at Lanzhou Jiaotong University, majoring in computer technology. His main research interests are machine learning and intelligent transportation