BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 72(6), 2024, Article number: e151955
DOI: 10.24425/bpasts.2024.151955

ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE

Traffic accident prediction method based on multi-view spatial-temporal learning

Jian FENG[1]*, Tian LIU[1] and Yuqiang QIAO[2]

[1] College of Computer Science & Technology, Xi'an University of Science and Technology, Xi'an 710000, China
[2] Shaanxi Branch, China United Network Communications Group Co., Ltd., Xi'an 710000, China

Abstract. Traffic accident prediction is a crucial component of an intelligent traffic system, which is important to maintain citizen safety and decrease economic losses. Current methods for traffic accident prediction based on deep learning fail to consider the driving mechanisms of traffic accidents, so a novel traffic accident prediction method based on multi-view spatial-temporal learning is proposed, which represents the driving mechanism of traffic accidents from multiple views. Firstly, for the urban regions divided by grids, a new augmentation was designed to augment the spatial semantic information of regions through learnable semantic embedding, then deformable convolutional networks with non-fixed convolution kernels are used to learn dynamic spatial dependencies between regions and gated recurrent units are used to learn temporal dependencies, which can capture dynamic spatial-temporal evolution patterns of traffic accidents. Secondly, long short-term memory is employed to learn the traffic flow breakdown from the flow difference of adjacent time steps in each region to recognize the traffic accident precursor in the risk environment. Thirdly, accident patterns in different regions are learned from historical traffic flow to determine whether the flow is the dominant factor and capture the spatial heterogeneity of traffic accidents. Finally, the above features are fused for accident prediction at the regional level. Experiments are conducted on two real datasets, and the experimental results show that the proposed method outperforms eight benchmark methods.

Keywords: traffic accident prediction; spatial-temporal dependencies; accident precursor; spatial heterogeneity.

1. INTRODUCTION

The traffic safety situation is becoming increasingly serious due to the rapid development of motorized transportation, and traffic accidents are becoming an important factor in influencing the quality of life and safety level of people. Traffic accident prediction aims to predict future accidents by analyzing prior accidents and considering relevant factors comprehensively to reduce property damage and casualties. Meanwhile, other research fields in the traffic system [1–3] will also benefit from the development of traffic accident prediction, such as intelligent mobility [4], autonomous driving [5], and trajectory planning [6].

Traffic accident prediction is a typical spatial-temporal problem. The most recent paradigm employs deep learning techniques to learn the spatial-temporal dependencies from spatial-temporal grids or spatial-temporal graphs constructed by traffic accidents, traffic flow, road network topology, etc. Generally, convolutional neural networks (CNNs) [7] and graph convolutional networks (GCNs) [8] are used to learn spatial dependencies from spatial-temporal grids and spatial-temporal graphs respectively, and recurrent neural networks (RNNs) [9, 10] are used to learn temporal dependencies.

Due to the strong nonlinear learning capability, deep learning has demonstrated significant advancements in traffic accident prediction in recent years. However, traffic accidents result from multiple interacting factors, including individuals, vehicles, road conditions, environment, unforeseen events, etc., and their spatial-temporal dependencies are very complex. Current research does not sufficiently consider multiple aspects, including:

Spatial dependencies. On the one hand, the traffic conditions are constrained by the topology of the road network, so the interplay between adjacent regions is intricate. However, existing research often ignores the dynamic changes of mutual influence between adjacent regions after establishing the initial influence [11, 12], such as vehicle diversion resulting from road works. On the other hand, although there is physical location information of the accidents in the traffic accident data, the corresponding spatial semantic information of the accidents is not provided explicitly, which affects the in-depth analysis of the spatial pattern of traffic accidents. For example, shopping malls and residential regions have different accident patterns. Although Wang *et al.* [13] explored the semantic information behind points of interest (POIs) through the similarity between different regions, they ignored the dynamic changes of spatial dependence like other works. In contrast, Trilat *et al.* [14] considered dynamic views but failed to mine the semantic information behind specific regions.

Accident precursor. The initial phase of a traffic accident typically exhibits gradual or sudden changes in traffic conditions, which are manifested in varied degrees through traffic

*e-mail: fengjian@xust.edu.cn

Manuscript submitted 2024-01-09, revised 2024-08-06, initially accepted for publication 2024-09-01, published in November 2024.

flow parameters. Traditional traffic accident analysis is usually conducted from two scenarios: normal traffic environment and risky traffic environment, because the mechanisms of the two situations that trigger the change of traffic operation state are different. Nevertheless, the existing research for traffic accident prediction using deep learning focuses on capturing the nonlinear correlation between traffic flow and traffic accidents [11–14], disregards the underlying accident-caused mechanisms, and so neglects the accident precursor thoroughly.

Spatial heterogeneity. The primary factors for accidents in different regions are varied, leading to distinct accident patterns. For example, in rural regions, adverse weather conditions often cause the emergence of accidents, but in urban regions, accidents are common during rush hours. Existing research tends to handle diverse regions homogeneously, learning different accident patterns in various regions inadequately. Although a few studies [11, 15] address heterogeneity by treating different regions differently, these methods fail to explore the underlying influencing factors while increasing costs.

To this end, a traffic accident prediction method based on multi-view spatial-temporal learning (MVSTL) is proposed based on analyzing the potential mechanism of traffic accidents. MVSTL captures the complex spatial-temporal dependencies of traffic accidents by learning the influence from multiple perspectives among multi-source data such as traffic flow, POI, weather, date, and traffic accidents. The main contributions of the paper are as follows:

- MVSTL is proposed to investigate the occurrence rules of traffic accidents from three perspectives, namely spatial-temporal dependencies, accident precursor, and spatial heterogeneity.
- When learning spatial-temporal dependencies, on the one hand, the semantics of POI and weather are augmented by learnable semantic matrixes to strengthen spatial dependencies. On the other hand, deformable convolutional networks (DCNs) [16] are introduced to learn dynamic dependencies between regions adaptively.
- The accident precursor is learned from the flow difference of adjacent time steps. The accident pattern of different regions is represented by the ratio of the current flow to the historical flow of the regions to learn the spatial heterogeneity of accidents.
- Experimental results in real datasets demonstrate that MVSTL outperformed the benchmark methods.

The rest of the research paper is organized as follows. Section 2 is a review of related work. The proposed method is detailed in Section 3. In Section 4, the effectiveness of the proposed method is demonstrated through experiments. Finally, Section 5 summarizes the research and directs future research.

2. RELATED WORKS

Research on traffic accident prediction can be roughly divided into two categories, statistical methods and machine learning-based methods.

2.1. Statistical methods

The statistical methods explore the relationship between variables based on statistical theory. Typical methods include the vector autoregressive model (VAR), autoregressive integrated moving average model (ARIMA), seasonal autoregressive integrated moving average model (SARIMA), and exponential smoothing (ES). For example, Li *et al.* [17] analyzed the different influences of traffic, weather, and socioeconomic characteristics on traffic collisions by VAR and Bayesian inference and found that different types of collisions have different trends during the prediction period. Getahun [18] modelled the trend of traffic accidents by ARIMA and found that traffic accidents within a week have an uneven distribution. Rabbani *et al.* [19] predicted the number of accidents by SARIMA and ES and found traffic accidents have considerable seasonality and non-stationarity. These studies are suitable for analyzing the influencing factors of accidents since they can reveal the characteristics of accidents, such as causality and randomness. However, while statistical methods are good at analyzing low-dimensional data, they are challenging to handle high-dimensional traffic data.

2.2. Machine learning-based methods

2.2.1. Traditional machine learning-based methods

Combining domain expert experience, early research based on traditional machine learning predict traffic accidents through learning the relationship between accident-related factors and accidents in high-dimensional data. Typical methods include Bayesian networks (BNs), support vector machines (SVMs), artificial neural networks (ANNs), etc. For example, Castro and Kim [20] explored variables that affect the degree of accident risk by BNs and found that lighting conditions and road types are the decisive factors of traffic accidents. Xiong *et al.* [21] studied the impact of precipitation and weather conditions on accidents based on SVMs. Lee *et al.* [22] used ANNs and k-nearest neighbours to explore the factors that influence accident duration. Fallah Tafti and Roshani [23] identified the most effective factors influencing accidents that occurred on the final sections of main access roads to the cities through ANNs. These works improve prediction accuracy compared to the statistical methods but rely on artificial feature extraction.

2.2.2. Deep learning-based methods

Both the statistical methods and the traditional machine learning-based methods only consider the temporal dependencies and ignore the spatial dependencies, so traffic accident prediction in these methods will be restricted by the road topology and fail. To learn the potential temporal and spatial characteristics of traffic accidents automatically and further improve the prediction accuracy, deep learning is applied to traffic accident prediction.

The deep learning-based method fuses heterogeneous data from multiple sources to realize grid division or graph construction and then extracts spatial-temporal dependencies using the powerful nonlinear learning capability of deep learning. The accident-related multi-source heterogeneous data include road network structure, weather, traffic flow, and other information.

Among them, the road network structure is often used to build topology, while other data are used as features of grids or nodes. As an early representative method, a stacked denoising autoencoder was used by Chen *et al.* [24] to learn the relationship between traffic accidents and human activities. On this basis, Chen *et al.* [25] added a CNN to analyze the spatial dependence of adjacent regions. In contrast, Sameen and Pradhan [26] learned the temporal dependencies between accidents based on long-short term memory (LSTM) [10]. However, these studies only model single spatial or temporal dependencies.

With the deepening of research, for spatial-temporal grid data, the current typical method is to learn spatial dependencies by a CNN and learn temporal dependencies by an RNN, and for spatial-temporal graph data, the spatial dependencies are learned by a GCN to adapt its non-European structure, and the temporal dependencies are learned by RNN-like methods, and then spatial-temporal dependencies are combined. For example, a CNN and an LSTM were combined to learn spatial-temporal dependencies from multi-source data by Yuan *et al.* [11]. Yu *et al.* [12] used roads as nodes to construct road network graphs and then learned spatial-temporal dependencies by combining GCN and temporal convolutions. Nevertheless, these methods learn spatial-temporal dependencies from fixed structures. To overcome this obstacle, Wang *et al.* [13] combined a CNN and a gated recurrent unit (GRU) [9] to learn spatial-temporal dependencies and additionally construct graphs by the similarity of accident risks, roads, and POIs between grids to learn global semantic spatial dependencies through a GCN. Trilat *et al.* [14] considered the time factor in calculating similarity and learned various dependency relationships between regions besides the traditional adjacency matrix through a GCN after constructing the graph. Wang *et al.* [27] constructed a graph based on learned features, hoping to learn spatial-temporal dependencies adaptively.

In general, early statistical methods are good at discovering the relationship between the influencing factors of traffic accidents, and can effectively reveal the mechanism of traffic accidents, but they ignore the spatial characteristics and have low accuracy. On the contrary, the deep learning-based methods as current mainstream methods can improve prediction accuracy, because they can learn high-dimensional and complex characteristics from traffic data automatically. However, existing studies usually ignore the mechanism analysis of traffic accidents and cannot fully consider the characteristics of accidents. On instinct, if the mechanism of traffic accidents can be considered in deep learning methods, working together with their complex nonlinear learning ability, the corresponding deep learning methods could be proposed, and the accuracy of traffic accident prediction should be further improved.

To this end, for spatial-temporal grid data, MVSTL combines a DCN and a GRU to learn the dynamic spatial-temporal dependencies of traffic accidents after using learnable embedding matrixes to augment semantics. And then, the accident precursor is learned by the flow difference of adjacent time steps by analyzing the occurrence mechanism of traffic accidents. Finally, considering that accident patterns have spatial heterogeneity, accident patterns in different regions are learned from the ratio of current flows to historical flows.

3. METHOD

3.1. Problem formulation

The traffic data are first introduced and then the traffic accident prediction problem is formalized.

3.1.1. Multi-source traffic data

Traffic data are divided into three categories: spatial data, temporal data, and spatial-temporal data. Among them, the spatial data is only related to the location, including the region and POI; the temporal data change over time, such as calendar information; the spatial-temporal data are affected by both location and time, including weather, traffic flow per unit time, accident risk level, and so on. Due to the inconsistent numerical range of these data, normalization is required. For enumerated data, such as weather and POI, one-hot encoding is used; for numerical data, normalized numerical encoding is used. The specific definition of multi-source data is explained below.

Traffic grid. Research city is divided into $I \times J$ grids according to latitude and longitude, each grid represents a region.

Traffic accident. Firstly, the accidents are matched to different time steps of the traffic grids based on the location and time. Then the accidents are divided into three levels according to the number of casualties: mild, moderate, and severe, and the corresponding accident levels are assigned as 1, 2, and 3. Finally, the accident risk of each grid at each time step is calculated by the weighted sum based on weights assignment to different accident levels [28]. At time step t, the accident risk of region m is $Y_{m,t} \in \mathbb{R}$ and the accident risk distribution of all regions is $Y_t \in \mathbb{R}^{I \times J}$.

Traffic flow. Traffic flow consists of the inflow and outflow of vehicles at each time step in each region. For example, the inflow and outflow of region m at time step t are $X_{m,t}^{F_I}, X_{m,t}^{F_O} \in \mathbb{R}$, respectively, and the traffic flow of all regions is $[X_t^{F_I}, X_t^{F_O}] \in \mathbb{R}^{I \times J \times d_F}$, where the feature dimension $d_F = 2$.

POI. POIs generally refer to all geographical objects that can be abstracted into points, especially some geographical entities that are closely related to people's lives, such as shopping malls, hospitals, gas stations, and so on. According to the relationship with traffic accidents, seven types of POIs are used in this paper: residence, school, culture facility, recreation, social service, transportation, and commercial premises. Since the distribution of POIs does not change with time in the short term, the number of POIs in each traffic grid reflects the geographic characteristics of different regions. $X_m^P \in \mathbb{R}^{d_P}$ is used to represent the distribution of POIs in region m, where $d_P = 7$. All regional POI distribution is denoted as $X^P \in \mathbb{R}^{I \times J \times d_P}$.

Weather. The weather includes temperature and weather conditions, which are collected consistent with the time interval of traffic data. While the temperature is represented by a normalized numerical value, the weather conditions are enumerated and represented by one-hot encoding, including five categories: sunny, rainy, snowy, cloudy, and foggy. Therefore, the weather of region m at time step t is denoted as $X_{m,t}^W \in \mathbb{R}^{d_W}$, where $d_W = 6$, and the weather of all regions is denoted as $X_t^W \in \mathbb{R}^{I \times J \times d_W}$.

Calendar. Calendar information is represented by one-hot encoding. A specific time in a day is represented by 24-bit one-hot encoding, a week is represented by 7-bit one-hot encoding, and whether it is a holiday is represented by 1-bit one-hot encoding. The calendar information of region m at time step t is $X^C_{m,t} \in \mathbb{R}^{d_C}$, where $d_C = 32$. The calendar information for all regions is $X^C_t \in \mathbb{R}^{I\times J\times d_C}$.

3.1.2. Traffic accident prediction

Based on the above multi-source historical data, traffic accident prediction aims to find a function $f = (\cdot)$ to predict the accident risk at the next time step. Let $X_t = [Y_t, X^F_t, X^P, X^W_t, X^C_t]$ be the multi-source data collected before time step $T-1$, then the prediction of the traffic accident risk at the target time T is shown in equation (1)

$$\hat{Y}_T = f\left(X_{T-1-I_w\times n_w},\ldots,X_{T-n_h},\ldots,X_{T-1}\right), \tag{1}$$

where n_h, n_w denotes the number of hours and weeks before the target time step, and I_w is the number of time steps in one week, as shown in Fig. 1. Such inputs are designed to learn proximity and periodicity simultaneously.

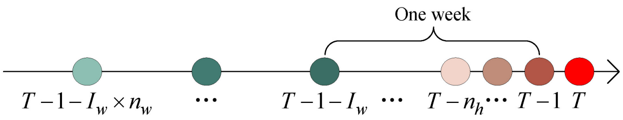

Fig. 1. Time steps of historical data

3.2. Architecture of MVSTL

MVSTL consists of three modules, namely the data preprocessing module, spatial-temporal feature learning module, and prediction module, as shown in Fig. 2. The multi-source data is processed by the data preprocessing module to usable input. The spatial-temporal feature learning module is divided into three submodules to learn the spatial-temporal dependencies, accident precursor, and spatial heterogeneity of accidents from the perspective of accident driving mechanism. Among them, the spatial-temporal dependencies learning submodule uses a DCN to learn dynamic spatial dependencies based on augmenting the spatial semantic information of accidents, then uses a GRU to learn temporal dependencies; the accident precursor learning submodule takes the flow difference and accident risk as input, and learns the accident precursor through LSTM; the spatial heterogeneity learning submodule learns accident patterns in different regions from the ratio of current flow to historical flow through a fully connected (FC) layer. The prediction module fuses the outputs of the three submodules and then predicts the accident risk in the next time step through FC.

3.3. Spatial-temporal dependencies learning submodule

The most crucial problem in traffic accident prediction is capturing complex spatial-temporal dependencies. When learning spatial dependencies, on the one hand, although existing research has constructed the road network structure and learned its spatial dependencies in different ways, they often ignore the semantic information in the structure. There are various POIs in cities, and the distribution of POIs determines whether the current region belongs to a commercial region, office region, or education region, which can help reveal the spatial semantic pattern of the accident and explore the occurrence rules of the accident. On the other hand, the existing research often uses CNNs to learn the spatial dependencies of traffic grids. However, a CNN has a limited capacity to learn long-distance dependen-

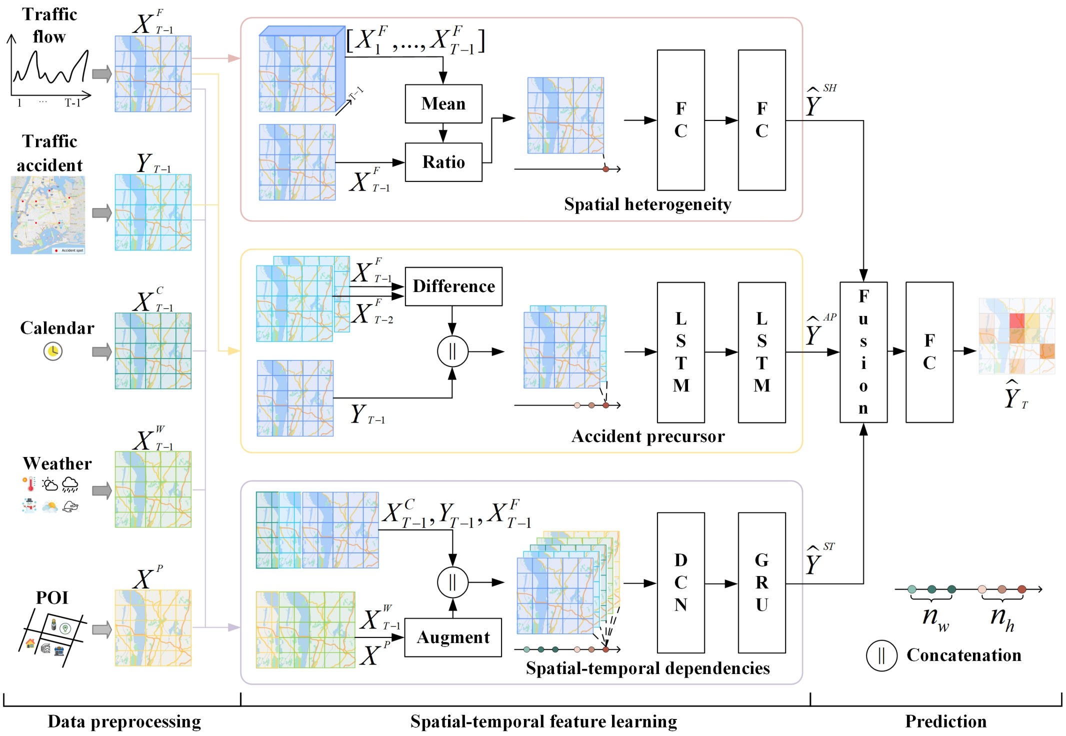

Fig. 2. Architecture of MVSTL

cies and also lacks the ability to learn the dynamic changes of spatial dependencies.

To solve the above problems, the learning of spatial-temporal dependencies is divided into three steps:

(1) Semantic augmentation

Firstly, the semantic information of each region is augmented by converting the original POI number into the semantic accumulation of different POI through the embedding matrix. The calculation process is shown in equation (2)

$$E^P = X^P \times W_P \in \mathbb{R}^{I\times J\times d'_P}, \tag{2}$$

where $W_P \in \mathbb{R}^{d_P\times d'_P}$ is a trainable embedding matrix, d'_P is the embedding dimension of the enhanced POI, and E^P is the augmented POI.

Secondly, different weather conditions have varied impacts on traffic patterns. For example, bad weather can lead to adverse traffic conditions, such as traffic jams and mixed traffic of pedestrians and vehicles, which can easily cause accidents. Therefore, the weather information X_t^W needs to be augmented as well. By the augmented method same as the POI, the augmentation result is $E_t^W \in \mathbb{R}^{I\times J\times d'_W}$, where d'_W is the embedding dimension of the enhanced weather information.

(2) Spatial dependencies learning

After augmenting the POI and weather information, a DCN is employed to learn dynamic spatial dependencies for multi-source data. Unlike a CNN, a DCN can learn the relationship between grids adaptively by changing the shape mapping in the target receptive field of the convolution kernel through an offset adding, which can better capture the dynamic spatial dependencies between traffic grids.

The augmented multi-source data is $[X^{ST}_{T-1-I_w\times n_w}, \cdots, X_t^{ST}, \ldots, X^{ST}_{T-1}]$, where $X_t^{ST} = [Y_t, X_t^F, E^P, E_t^W, X_t^C]$. For each time step, the specific convolution operation is

$$H_t^l = \sigma\left(H_t^{l-1} * W_c^l + b_c^l\right), \tag{3}$$

where $H_t^0 = X_t^{ST}$, $*$ indicates the convolution operation, W_c^l and b_c^l indicate the parameters of the DCN in the l-th layer, σ is the activation function Relu, and H_t^l is the representation of the l-th layer at the time step t. Finally, the output of the L layer DCN is $[H^L_{T-1-I_w\times n_w}, \ldots, H_t^L, \ldots, H^L_{T-1}]$.

(3) Temporal dependencies learning

The temporal dependencies between different time steps are learned by a GRU after the spatial dependencies learning. Taking H_t^L as an example, the calculation process of the GRU is as follows:

$$r_t = \sigma\left(W_r H_t^L + U_r h_{t-1}^{ST}\right), \tag{4}$$

$$z_t = \sigma\left(W_z H_t^L + U_z h_{t-1}^{ST}\right), \tag{5}$$

$$\tilde{h}_t^{ST} = \tanh\left(W_h H_t^L + r_t \odot U_h h_{t-1}^{ST}\right), \tag{6}$$

$$h_t^{ST} = z_t \odot h_{t-1}^{ST} + (1-z_t) \odot \tilde{h}_t^{ST}, \tag{7}$$

where W_r and U_r, W_z and U_z, W_h and U_h are the weights of the reset gate, update gate, and $tanh$ function, respectively, h_{t-1}^{ST} is the hidden state of the previous moment, and $\odot$ denotes the Hadamard product. The output of the last time step is the output of the spatial-temporal dependencies learning submodule

$$\hat{Y}^{ST} = h_{T-1}^{ST}. \tag{8}$$

3.4. Accident precursor learning submodule

The occurrence of traffic accidents is accidental. However, existing research has proved that the occurrence of accidents is also inevitable in some cases from the perspective of traffic flow changes, since the traffic flow parameters will change in some pattern before and after the accident [29]. This change is called accident precursor. From the perspective of accident causes, accident precursor is mainly divided into two situations. One is a sudden change in traffic conditions, and the other is the cascading effect brought by an existing accident, namely a secondary accident.

In the first case, traffic accidents are triggered by changes in traffic flow due to the increase in traffic demand. But in the second case, a sharp drop in traffic capacity caused by the occurrence of existing traffic accidents leads to traffic flow breakdown and is often accompanied by the risk of secondary accidents. So, the change in traffic flow can reflect the accident precursor to some extent.

To learn the accident precursor, the flow difference of adjacent time steps is calculated first, and then it is combined with the accident risk to form sequence data, recorded as $[X^{AP}_{T-n_h}, \ldots, X_t^{AP}, \ldots, X^{AP}_{T-1}]$, where $X_t^{AP} = \left[Y_t, X_t^{D_I}, X_t^{D_o}\right]$, $X_t^{D_I}$ and $X_t^{D_o}$ represent the difference between the inflow and outflow at time step t and time step $t-1$, respectively, and the calculations are shown in equations (9) and (10)

$$X_t^{D_I} = X_t^{F_I} - X_{t-1}^{F_I}, \tag{9}$$

$$X_t^{D_o} = X_t^{F_o} - X_{t-1}^{F_o}. \tag{10}$$

Then, LSTM is used to learn the accident precursor. Take X_t^{AP} as an example, the calculation is as follows:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}^{AP}, X_t^{AP}\right] + b_f\right), \tag{11}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}^{AP}, X_t^{AP}\right] + b_i\right), \tag{12}$$

$$\tilde{c}_t = \tanh\left(W_c \cdot \left[h_{t-1}^{AP}, X_t^{AP}\right] + b_c\right), \tag{13}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{14}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}^{AP}, X_t^{AP}\right] + b_o\right), \tag{15}$$

$$h_t^{AP} = o_t \odot \tanh\left(c_t\right), \tag{16}$$

where h_t^{AP} is the hidden state at time step t, c_t denotes the state of the memory unit, W_f and b_f, W_i and b_i, W_o and b_o, W_c and b_c denote the weight and bias of the forgetting gate, input gate, output gate, and tanh function, respectively. The output of the last time step is used as the output of the accident precursor learning submodule

$$\hat{Y}^{AP} = h_{T-1}^{AP}. \tag{17}$$

3.5. Spatial heterogeneity learning submodule

There are different causes of accidents in different geographic regions. For example, some regions may be prone to traffic accidents due to road design defects, while others may be due to heavy traffic. Therefore, different regions may have different accident patterns, namely spatial heterogeneity. There are many factors that affect spatial heterogeneity, but some are not easy to learn, so we only learn spatial heterogeneity from the traffic flow.

Firstly, the average value of historical traffic flow is calculated, and the ratio of the traffic flow at the current moment to the historical average traffic flow is calculated to learn accident patterns in different regions. Specifically, for the target time step T, the input is recorded as $X^{SH} = [X^{A_I}_{T-1}, X^{A_O}_{T-1}]$, where $X^{A_I}_{T-1}$, $X^{A_O}_{T-1} \in \mathbb{R}$ are the ratios of the current inflow and outflow to the average value of historical flow, respectively:

$$\bar{X}^{F_I}_{T-1} = \frac{1}{N_w} \sum_{i=0}^{N_w-1} X^{F_I}_{T-1-i\times I_w}, \tag{18}$$

$$\bar{X}^{F_O}_{T-1} = \frac{1}{N_w} \sum_{i=0}^{N_w-1} X^{F_O}_{T-1-i\times I_w}, \tag{19}$$

$$X^{A_I}_{T-1} = \frac{X^{F_I}_{T-1}}{\bar{X}^{F_I}_{T-1}+1}, \tag{20}$$

$$X^{A_O}_{T-1} = \frac{X^{F_O}_{T-1}}{\bar{X}^{F_O}_{T-1}+1}, \tag{21}$$

where $\bar{X}^{F_I}_{T-1}$ and $\bar{X}^{F_O}_{T-1}$ denote the average inflow and outflow of each region, and N_w represents the number of historical weeks in the dataset. In the equation (20) and (21), $\bar{X}^{F_I}_{T-1}$ and $\bar{X}^{F_O}_{T-1}$ are added 1 to prevent calculation errors when the traffic is 0.

Accident patterns $\hat{Y}^{SH}$ is learned by FC for X^{SH}

$$\hat{Y}^{SH} = FC\left(X^{SH}\right). \tag{22}$$

3.6. Prediction module

The prediction module fuses the outputs of the above three submodules for the final prediction. Considering the impact of the fusion method on the final prediction result, the $1*1$ convolution is used for vector reduction and redundant information elimination, and then the Hadamard product is used for fusion. The accident risk prediction is performed through FC after fusion.

$$\hat{Y}_T = FC\left(W_{ST} * \hat{Y}^{ST} \odot W_{AP} * \hat{Y}^{AP} \odot W_{SH} * \hat{Y}^{SH}\right), \tag{23}$$

where $*$ represents the convolution operation, $\odot$ represents the Hadamard product. W_{ST}, W_{AP} and W_{SH} are parameters of the convolution kernel, and $\hat{Y}_T$ is the accident risk of all regions at the target time step.

3.7. Model training

Since traffic accidents are sparse compared to normal travel data, the model will be more inclined to predict non-accident for the target moment, which makes model training difficult.

To address the sparsity of accidents, a joint loss function is used in training [30]. Among them, the mean square error (MSE) loss function focuses on reflecting the distribution of low-risk accidents, and the mean absolute error (MAE) loss function focuses on reflecting the distribution of high-risk accidents. The sparsity problem can be alleviated by combining these two loss functions. The calculation process of MSE and MAE is as follows:

$$\mathscr{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i \left(Y(i) - \hat{Y}(i)\right)^2, \tag{24}$$

$$\mathscr{L}_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i \left|Y(i) - \hat{Y}(i)\right|, \tag{25}$$

where N represents the number of samples, each sample is composed of accident risk values of all regions at a specific time. $Y(i)$ and $\hat{Y}(i)$ represent the true value and predicted value of the i-th sample, respectively, and λ_i represents the weight of the i-th sample, and different weights are given according to the risk of the corresponding sample [13].

The final joint loss function is

$$\mathscr{L} = \mathscr{L}_{\text{MSE}} + \mathscr{L}_{\text{MAE}}. \tag{26}$$

4. EXPERIMENTS AND ANALYSIS

In this section, MVSTL is evaluated on two datasets, and the experiments are designed to answer the following questions.

1. Question 1: Does MVSTL outperform competing baselines?
2. Question 2: Are the key components of MVSTL helpful for prediction?
3. Question 3: How does the fusion method affect the method's performance?
4. Question 4: How efficient is MVSTL?

4.1. Experiment preparation

4.1.1. Datasets

The experimental data comes from the public data of the governments of New York (NYC) and Chicago (Chicago). It contains five types of data on the two cities, including traffic accidents, taxi orders, POI, weather, and time data. The specific information is shown in Table 1.

Table 1
Datasets

Dataset	NYC	Chicago
Time range	01/01/2013– 31/12/2013	01/02/2016– 30/09/2016
Number of traffic accidents	147 K	44 K
Number of taxi orders	173 179 K	1744 K
Number of POIs	15 625	–
Number of weather	8760	5823

4.1.2. Baseline

The baseline includes two classic machine learning-based methods and six latest deep learning-based methods, in order:

1. XGBoost: The ensemble learning model uses a regression tree as the base learner.
2. MLP: Multilayer Perceptron.
3. GRU: Gated RNN approach, good at learning temporal dependencies in data.
4. SDCAE [25]: Introducing CNNs to learn spatial dependencies in stacked denoising autoencoders.
5. ConvLSTM [11]: Combining CNNs with LSTMs to learn spatial-temporal dependencies.
6. ST-RistNet [28]: Combining GCNs and GRUs to learn spatial-temporal dependencies in traffic flow.
7. GSNet [13]: GCNs and GRUs are used to learn spatial-temporal dependencies; the semantic spatial-temporal dependencies are learned through GCNs based on three semantic graphs constructed according to road features, POI, and risks.
8. MG-TAR [14]: GCNs and temporal attention are used to learn various dependencies from graphs constructed from environmental data besides the traditional adjacency matrix.

Among them, XGBoost and MLP represent efficient machine learning methods. In deep learning methods, a GRU and an SDCAE, respectively, represent methods that only learn temporal or spatial dependencies. The remaining methods are designed based on spatial-temporal dependencies.

4.1.3. Experimental environment and settings

The operating system environment is Ubuntu 18.04, and the development framework is Pytorch 1.8.1. The hardware equipment uses NVIDIA RTX3080Ti GPU for training, and its CPU is Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40 GHz, and the memory is 32 GB.

The number of training batches is 32 and the learning rate is 1e–6. The early stopping mechanism is enabled during the training process, and the patience is 5.

4.1.4. Evaluating metrics

Traffic accident risk prediction is a regression task, for which root mean square error (RMSE) is used as an evaluation index. At the same time, considering that the prediction result is the accident risk distribution and contains multiple prediction regions, the recall rate and mean average precision (MAP) are introduced to evaluate the hit rate of each time step for the risk regions:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y(i)-\hat{Y}(i)\right)^2} \tag{27}$$

$$Recall = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|R_i \cap \hat{R}_i\right|}{|R_i|}, \tag{28}$$

$$MAP = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{|R_i|} p(j)\times r(j)}{|R_i|}, \tag{29}$$

where R_i and $\hat{R}_i$ are the sets of actual and predicted top $|R_i|$ highest risk regions for sample i, respectively. $p(j)$ represents the precision ranking list from 1 to j. $r(j) = 1$ indicates that accidents have occurred in region j, otherwise $r(j) = 0$.

4.2. Experiment 1: Comparative experiment

For question 1, Experiment 1 compared the performance of MVSTL with each baseline. Table 2 shows the comparison results in the two data sets, where $*$ represents the Chicago dataset.

Table 2
Performance comparison

Methods	RMSE	Recall	MAP	RMSE∗	Recall∗	MAP∗
XGBoost	10.513	21.56%	0.103	15.104	12.01%	0.048
MLP	8.488	27.41%	0.119	11.948	16.71%	0.056
GRU	7.852	30.43%	0.151	11.617	17.63%	0.069
SDCAE	8.020	31.08%	0.152	11.613	17.59%	0.066
ConvLSTM	7.674	31.27%	0.173	11.717	19.38%	0.079
ST-RiskNet	7.660	32.46%	0.181	11.487	20.04%	0.083
GSNet	7.671	33.42%	0.185	11.373	21.11%	0.090
MG-TAR	7.810	30.19%	0.184	**10.607**	18.43%	0.091
MVSTL	**7.622**	**34.55%**	**0.195**	11.868	**21.98%**	**0.105**

The bold ones in the table are the optimal results, and the underlined ones are the suboptimal results. It can be seen from Table 2 that MVSTL generally has lower RMSE and higher recall and MAP. Lower RMSE indicates that MVSTL is more accurate in predicting the risk of all regions. Higher recall and MAP indicate that MVSTL has a higher hit rate for the prediction of high-risk regions, and the prediction results are more correlated with the real risk distribution. Therefore, considering all metrics, MVSTL outperforms all baselines.

Among baselines, the machine learning-based methods XGBoost and MLP perform poorly because they process each piece of data individually and ignore the spatial dependencies between data. The performance of deep learning-based methods has improved. For example, a GRU performs better in modelling the time series of accident data because it can capture short-term proximity and long-term periodicity, further confirming the importance of modelling temporal dependencies in traffic accident prediction. SDCAE models the spatial dependencies of adjacent regions by stacking multi-layer convolutional neural networks but ignores the temporal dependencies and spatial dependencies of the global region. ConvLSTM can capture the temporal and spatial dependencies of traffic accidents simultaneously by combining convolutional neural networks and long short-term memory networks. ST-RiskNet and GSNet achieve good results by modelling local spatial-temporal dependencies and global spatial-temporal similarities based on static convolution kernels and fixed graph structures. The above two methods can capture the global spatial dependencies to a certain extent by constructing the global similarity graph but cannot capture its dynamic changes. MG-TAR considers dynamic changes by considering time factors but ignores accident precursors. In contrast, MVSTL considers the dynamic spatial-temporal dependencies between regions and deeply explores the problem of accident precursor and spatial heterogeneity. At the same time, the hidden semantic information behind POI and weather is also extracted to learn the traffic accident mode better, so the best results are obtained. The experimental results illustrate that it is feasible and effective to capture multiple accident characteristics from the perspective of accident driving mechanisms.

4.3. Experiment 2: Ablation experiment

For question 2, Experiment 2 verified the impact of different submodules in MVSTL on the prediction results. To this end, three model variants, MVSTL-ST, MVSTL-AP, and MVSTL-SH, were designed, representing the removal of the spatial-temporal dependencies learning submodule, the accident precursor learning submodule, and the spatial heterogeneity learning submodule, respectively. Figure 3 shows the results in the NYC dataset. The results obtained on the Chicago dataset are similar and will not be repeated here.

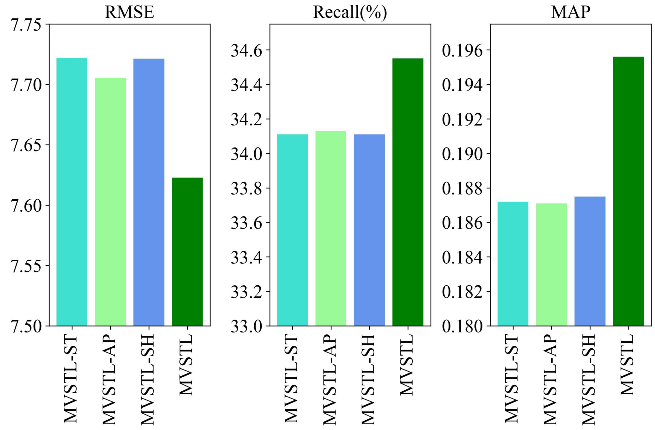

Fig. 3. Ablation experiments in NYC

It can be seen that MVSTL has the best overall performance because it considers the occurrence mode of accidents from multiple perspectives. Removing any one of the modules ignores a certain characteristic of the accident, resulting in a decline in the effect. This phenomenon reflects that each accident characteristic is helpful to the prediction, and the combination of different accident characteristics also has a positive impact.

4.4. Experiment 3: Parameter experiment

For question 3, Experiment 3 discusses the impact of different fusion methods on performance. There are six common fusion methods, including concatenate (C), LSTM (L), point-wise addition (P), CNN (N), max-pooling (M), and Hadamard product (H). To find a suitable fusion method for the prediction module, six variants were designed, denoted by MVSTL-C, MVSTL-L, MVSTL-P, MVSTL-N, MVSTL-M, and MVSTL-H. The experimental results are shown in Fig. 4.

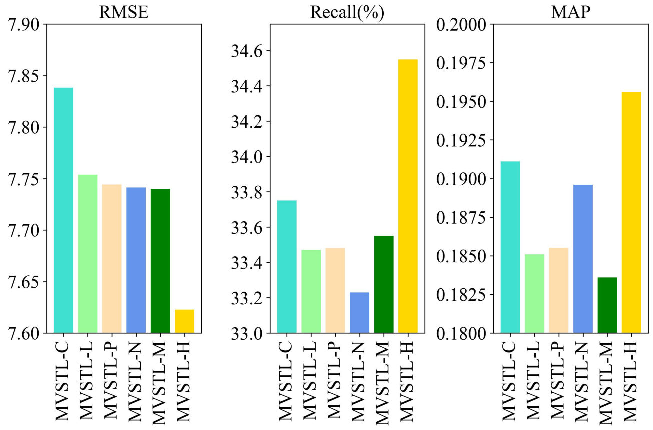

Fig. 4. Fusion methods in NYC

As shown in Fig. 4, the Hadamard product has the best effect when used to fuse the output results of the three sub-modules. This is because it can expand the importance of variables to a certain extent and increase attention to important information. LSTM may filter out some important information through the gating mechanism, so the result is lower than the Hadamard product fusion method. Concatenate and point-wise addition average the importance of information without considering the interaction between different space-time vectors, and certain semantic information will be lost. Max pooling will cause some hidden information to be lost. Therefore, Hadamard is used for the final feature fusion.

4.5. Experiment 4: Efficiency experiment

For question 4, Experiment 4 compared the model efficiency by recording the single-step prediction time of each model on the two data sets, and the results are shown in Fig. 5.

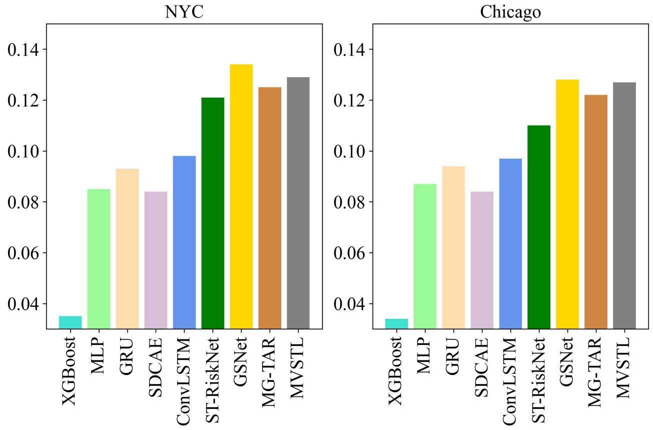

Fig. 5. Prediction time (s) in NYC and Chicago

As shown in Fig. 5, the prediction time of each model is within 0.5 s. Machine learning-based methods have short prediction times, while deep learning-based models have high time complexity, resulting in long prediction times. Compared to the improvement in performance, the increased time cost of deep learning methods is acceptable, and the efficiency of MVSTL can support its practical application.

Overall, the results of four experiments show that MVSTL can effectively predict traffic accidents.

5. CONCLUSIONS

This paper proposes a new deep-learning method for traffic accident prediction, namely MVSTL. MVSTL is designed from the perspective of accident driving mechanism, which makes up for the shortcomings of existing research. The key design of MVSTL is to take different modules to extract relevant features from the data by analyzing three accident characteristics, including dynamic spatial-temporal dependencies, accident precursor, and spatial heterogeneity. After fusing the learned features, MVSTL can achieve regional-level accident risk prediction.

Three metrics are used to evaluate the proposed method, and the experimental results show that MVSTL outperformed all baseline, especially in terms of the accuracy of predictions for regions with high accident risk. This proves that introducing traffic accident driving mechanisms can improve the accuracy of prediction results, which can bring new thinking to related research and further introduce theoretical knowledge to guide model design. It is worth mentioning that MVSTL designed based on accident characteristics is more interpretable compared to other methods and may be helpful for some applicable occasions. Moreover, experimental results prove that MVSTL prediction speed is fast enough to meet the needs of an intelligent traffic system for real-time accident prediction.

However, MVSTL has limitations. On the one hand, it is designed for grid data to predict the risk of accidents in the region. We may hope to be able to predict the risk of accidents at the road level, which also places higher demands on the granularity of data. On the other hand, it fails to consider the cascading effect of traffic accidents in depth. The occurrence of an accident may lead to a series of chain reactions, resulting in the probability of subsequent accidents. Therefore, we will collect fine-grained data in the future to study road-level accident prediction, and further study the cascading effects of traffic accidents through the theory of complex network dynamic evolution.

ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China under Project 2022YFB3304401.

REFERENCES

[1] R. Tomasz, E. Szczepański, and M. Jacyna, “Safety factor in the sustainable fleet management model,” *Arch. Transp.*, vol. 49, no. 1, pp. 103–114, 2019, doi: 10.5604/01.3001.0013.2780.

[2] E. Macioszek, A. Granà, and S. Krawiec, “Identification of factors increasing the risk of pedestrian death in road accidents involving a pedestrian with a motor vehicle,” *Arch. Transp.*, vol. 65, no. 1, pp. 7–25, 2023, doi: 10.5604/01.3001.0016.2474.

[3] M. Izdebski, I. Jacyna-Gołda, and P. Gołda, “Minimisation of the probability of serious road accidents in the transport of dangerous goods,” *Reliab. Eng. Syst. Safe.*, vol. 217, p. 108093, 2022, doi: 10.1016/j.ress.2021.108093.

[4] J. Murawski, E. Szczepański, I. Jacyna-Gołda, M. Izdebski and D. Jankowska-Karpa, “Intelligent mobility: A model for assessing the safety of children traveling to school on a school bus with the use of intelligent bus stops,” *Eksploat. Niezawodn.*, vol. 24, no. 4, pp. 695–706, 2022, doi: 10.17531/ein.2022.4.10.

[5] Y. Qian, H. Sun, and S. Feng, “Obstacle avoidance method of autonomous vehicle based on fusion improved A*APF algorithm,” *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 71, no. 2, p. e144624, 2023, doi: 10.24425/bpasts.2023.144624.

[6] Y. Qian, C. Deng, J. Xu, X. Qu, and Z. Song, “Occlusion-aware collision avoidance trajectory planning with potential collision risk assessment for autonomous vehicle,” *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 72, no. 4, p. e149819, 2024, doi: 10.24425/bpasts.2024.149819.

[7] M. Méndez, M.G. Merayo, and M. Núñez, “Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model,” *Eng. Appl. Artif. Intel.*, vol. 121, p. 106041, 2023, doi: 10.1016/j.engappai.2023.106041.

[8] Y. Shin and Y. Yoon, “PGCN: Progressive graph convolutional networks for spatial–temporal traffic forecasting,” *IEEE Trans. Intell. Transp.*, vol. 25, no. 7, pp. 7633–7644, 2024, doi: 10.1109/TITS.2024.3349565.

[9] N.S. Chauhan, N. Kumar, and A. Eskandarian, “A Novel Confined Attention Mechanism Driven Bi-GRU Model for Traffic Flow Prediction,” *IEEE Trans. Intell. Transp.*, vol. 25, no. 8, pp. 9181–9191, 2024, doi: 10.1109/TITS.2024.3375890.

[10] F. Kavehmadavani, V.D. Nguyen, T.X. Vu, and S. Chatzinotas, “Intelligent traffic steering in beyond 5G open RAN based on LSTM traffic prediction,” *IEEE Trans. Wirel. Commun.*, vol. 22, no. 11, pp. 7727–7742, 2023, doi: 10.1109/TWC.2023.3254903.

[11] Z. Yuan, X. Zhou, and T. Yang, “Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data,” in *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 984–992, doi: 10.1145/3219819.3219922.

[12] L. Yu, B. Du, X. Hu, L. Sun, L. Han, and W. Lv, “Deep Spatio-Temporal Graph Convolutional Network for Traffic Accident Prediction,” *Neurocomputing*, vol. 423, pp. 135–147, 2021, doi: 10.1016/j.neucom.2020.09.043.

[13] B. Wang, Y. Lin, S. Guo, and H. Wan, “GSNet: Learning Spatial-Temporal Correlations from Geographical and Semantic Aspects for Traffic Accident Risk Forecasting,” in *35th AAAI Conference on Artificial Intelligence*, 2021, pp. 4402–4409, doi: 10.1609/aaai.v35i5.16566.

[14] P. Trirat, S. Yoon and J.G. Lee, “MG-TAR: Multi-view graph convolutional networks for traffic accident risk prediction,” *IEEE Trans. Intell. Transp.*, vol. 24, no. 4, pp. 3779–3794, 2023, doi: 10.1109/TITS.2023.3237072.

[15] B. An, A. Vahedian, X. Zhou, W.N. Street, and Y. Li, “Hintnet: Hierarchical knowledge transfer networks for traffic accident forecasting on heterogeneous spatio-temporal data,” in *2022 SIAM International Conference on Data Mining (SDM)*, 2022, pp. 334–342, doi: 10.1137/1.9781611977172.38.

[16] J. Dai *et al.*, “Deformable Convolutional Networks,” in *IEEE international conference on computer vision*, 2017, pp. 764–773, doi: 10.1109/ICCV.2017.89.

[17] Z. Li, H. Yu, G. Zhang, and J. Wang, "A Bayesian Vector Autoregression-based Data Analytics Approach to Enable Irregularly-spaced Mixed-frequency Traffic Collision Data Imputation with Missing Values," *Transp. Res. Part. C Emerg. Technol.*, vol. 108, pp. 302–319, 2019, doi: 10.1016/j.trc.2019.09.013.

[18] K.A. Getahun, "Time Series Modeling of Road Traffic Accidents in Amhara Region," *J. Big. Data*, vol. 8, no. 1, pp. 1–15, 2021, doi: 10.1186/s40537-021-00493-z.

[19] M.B.A. Rabbani *et al.*, "A Comparison Between Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ES) Based on Time Series Model for Forecasting Road Accidents," *Arab. J. Sci. Eng.*, vol. 46, no. 11, pp. 11113–11138, 2021, doi: 10.1007/s13369-021-05650-3.

[20] Y. Castro and Y.J. Kim, "Data Mining on Road Safety: Factor Assessment on Vehicle Accidents Using Classification Models," *Int. J. Crashworthiness*, vol. 21, no. 2, pp. 104–111, 2016, doi: 10.1080/13588265.2015.1122278.

[21] X. Xiong, L. Chen, and J. Liang, "A New Framework of Vehicle Collision Prediction by Combining SVM and HMM," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 699–710, 2017, doi: 10.1109/TITS.2017.2699191.

[22] Y. Lee, C.H. Wei and K.C. Chao, "Non-parametric machine learning methods for evaluating the effects of traffic accident duration on freeways," *Arch. Transp.*, vol. 43, no. 3, pp. 91–104, 2017, doi: 10.5604/01.3001.0010.4228.

[23] M. Fallah Tafti, and R. Roshani, "Development of models to study traffic accidents on the final sections of access roads to the cities: a case study of three major Iranian cities," *Arch. Transp.*, vol. 59, no. 3, pp. 129–148, 2021, doi: 10.5604/01.3001.0015.2646.

[24] Q. Chen, X. Song, H. Yamada, and R. Shibasaki, "Learning Deep Representation from Big and Heterogeneous Data for Traffic Accident Inference," in *30th AAAI Conference on Artificial Intelligence*, 2016, pp. 338–344, doi: 10.1609/aaai.v30i1.10011.

[25] C. Chen, X. Fan, C. Zheng, L. Xiao, M. Cheng, and C. Wang, "SDCAE: Stack Denoising Convolutional Autoencoder Model for Accident Risk Prediction Via Traffic Big Data," in *16th International Conference on Advanced Cloud and Big Data (CBD)*, 2018, pp. 328–333, doi: 10.1109/CBD.2018.00065.

[26] M.I. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Appl. Sci.*, vol. 7, no. 6, p. 476, 2017, doi: 10.3390/app7060476.

[27] S. Wang, Y. Zhang, X. Piao, X. Lin, Y. Hu, and B. Yin, "Data-unbalanced traffic accident prediction via adaptive graph and self-supervised learning," *Appl. Soft Comput.*, vol. 157, p. 111512, 2024, doi: 10.1016/j.asoc.2024.111512.

[28] B. Wang, H. Wan, S. Guo, and Y. Lin, "Local and Global Spatial-Temporal Networks for Traffic Accident Risk Forecasting," *Front. Comput. Sci.*, vol. 15, no. 9, pp. 1694–1702, 2021, doi: 10.3778/j.issn.1673-9418.2008093.

[29] M. Abdel-Aty, A. Pande, C. Lee, V. Gayah, and C.D. Santos, "Crash Risk Assessment Using Intelligent Transportation Systems Data and Real-time Intervention Strategies to Improve Safety on Freeways," *J. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 107–120, 2007, doi: 10.1080/15472450701410395.

[30] P. Trirat and J.G. Lee, "DF-TAR: A Deep Fusion Network for Citywide Traffic Accident Risk Prediction with Dangerous Driving Behavior," in *Web Conference 2021*, 2021, pp. 1146–1156, doi: 10.1145/3442381.3450003.