

Bowen CAI, Léah CAMARCAT, Wen-long SHANG, Mohammed QUDDUS

A new spatiotemporal convolutional neural network model for short-term crash prediction

© The Author(s) 2024. This article is published with open access at link.springer.com and journal.hep.com.cn

Abstract Predicting short-term traffic crashes is challenging due to an imbalanced data set characterized by excessive zeros in noncrash counts, random crash occurrences, spatiotemporal correlation in crash counts, and inherent heterogeneity. Existing models struggle to effectively address these distinct characteristics in crash data. This paper proposes a new joint model by combining the time-series generalized regression neural network (TGRNN) model and the binomially weighted convolutional neural network (BWCNN) model. The joint model aims to capture all these characteristics in short-term crash prediction. The model was trained and tested using real-world, highly disaggregated traffic data collected with inductive loop detectors on the M1 motorway in the UK in 2019, along with crash data extracted from the UK National Accident Database for the same year. The short-term is defined as a 30-min interval, providing sufficient time for a traffic control center to implement interventions and mitigate potential hazards. The year was segmented into 30-min intervals, resulting in a highly imbalanced data set with over 99.99% noncrash samples. The joint model was applied to predict the probability of a crash occurrence by updating both the crash and traffic data every 30 min. The findings revealed that 75.3% of crashes and 81.6% of noncrash events were correctly predicted in the southbound direction. In the northbound direction, 78.1% of crashes and 80.2% of noncrash events were accurately captured. Causal analysis and model-based interpretation were used to analyze the relative importance of explanatory variables regarding their contribution to crashes. The results reveal that speed variance and speed are the most influential factors contributing to crash occurrence.

Keywords safety management, crash prediction, generalized regression neural network, binomial weighted CNN, variable importance

1 Introduction

Traffic crashes have significant economic and societal implications, including infrastructure damage, delays, and loss of life. In 2022, there were approximately 1.35 million fatalities globally due to traffic crashes (WHO, 2022), with 27,795 serious injuries and fatalities occurring in the UK alone (UK gov, 2022). Proactive traffic safety management has emerged as a primary approach to reduce traffic crashes, driven by the availability of advanced data sources such as loop detectors, cameras, radars, and lidars on motorways (Abdel-Aty et al., 2010). Real-time crash prediction (RTCP) is a key component of proactive traffic management aimed at identifying crash risks and implementing measures to ensure safe traffic conditions. The RTCP involves predicting crash occurrences within a short time frame using instantaneous traffic dynamics (Hossain et al., 2019).

Short-term crash prediction plays a crucial role in enhancing overall traffic safety, enabling proactive roadway safety management, and preventing potential traffic incidents (Zheng and Sayed, 2020). Lee et al. (2003) proposed a probabilistic Bayesian model to estimate crash risk in real time, emphasizing the need to monitor short-term traffic conditions. Cai et al. (2022) utilized computer vision techniques to extract dynamic traffic features from roadside surveillance cameras for crash prediction and found a strong correlation between crash characteristics and dynamic traffic data.

Although proactive traffic safety management systems have been extensively studied over the past decade, several challenges persist in short-term crash prediction. Unlike when predicting traffic conflicts between ego and preceding vehicles, accurately predicting the occurrence of an actual traffic crash is highly unlikely. Obtaining

Received Mar. 1, 2024; revised Jun. 12, 2024; accepted Jun. 18, 2024

Bowen CAI (✉), Léah CAMARCAT, Wen-long SHANG, Mohammed QUDDUS
Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK
E-mail: b.cai22@imperial.ac.uk

sufficient historical crash data requires several years of collection. Moreover, historical crash data are severely imbalanced, especially when segmented into short time windows. While many studies have focused on real-time prediction of traffic conflicts using sensors on ego vehicles, few have considered short-term prediction using roadway-based analysis.

In a study by Huang et al. (2020), the sensitivity of the time window on crash prediction accuracy was investigated for intervals of 1, 5, and 10 min. The researchers concluded that accurately predicting crashes 10 min prior to their occurrence is particularly challenging. This challenge is further amplified when predicting crashes within 30-min intervals. However, this timeframe provides an opportunity for traffic control centers and highway operational teams to implement effective safety measures without significantly disrupting traffic conditions.

This paper proposes a new approach called the time series generalized regression binomial weighted convolutional neural network (TGRCN) model for short-term traffic crash prediction. The model utilizes dynamic weights sampled from binomial regression. The TGRCN is a joint model consisting of a time series generalized regression neural network (TSGRNN) and a weighted CNN model (WCNN). By considering instantaneous spatial temporal structures in traffic dynamics and their relationships with traffic crashes, this technique effectively addresses the issue of data imbalance in the model. To the best of our knowledge, this is the first implementation of a time-series GRNN for crash prediction. The uniqueness of this method lies in its integration of a binomial weighted CNN.

The time-series GRNN is employed to predict the traffic state at time t based on data from time $t - 1$, allowing for the prediction of crashes within the same time dimension using the WCNN. A forensic analysis of the model's performance is conducted through causal effect analysis, enabling targeted actions to reduce crash risk. The model's output not only identifies the higher risk of a crash but also provides insights into the causes behind it.

To train and test the model, real-world data are utilized. Specifically, traffic data collected from loop detectors on the M1 motorway in the UK in 2019 are employed, and crash data are extracted from the STATS19 database for the same year. The data set consists of 244 crashes and more than 99.999% noncrash events. To address the data imbalance, the synthetic minority oversampling technique (SMOTE) is used to oversample the crash data. These oversampled data are then used to train the weighted CNN and obtain a classifier that captures the relationship between crash data and dynamic traffic variables.

The joint TGRCN model conducts crash prediction by updating crash and traffic data every 30 min. The results surpass those mentioned in many other papers on real-time crash prediction. Our method successfully detects crash events and classifies noncrash events. To validate

the superiority of our new method, we compare the results from the TGRCN joint model with those of several other types of neural networks, including multilayer perceptron (MLP), a time series artificial neural network, a long short-term memory network (LSTM), a spatiotemporal graph convolutional neural network (STGCN), a generalized autoregressive conditional heteroskedasticity (GARCH), and an autoregressive integrated moving average (ARIMA). We evaluate their prediction accuracy in short-term crash occurrence and find that the STGCN performs the best among the others. Although the spatiotemporal graph convolutional neural network is widely applied in predicting urban city traffic data, as initiated by Yu et al. (2018), it has not yet been applied to short-term crash prediction. Our proposed model in this paper shows significant improvements in the results, outperforming the STGCN graph neural network in short-term crash prediction.

Additionally, we employ the Intermediate ConvNet Activation Visualization and Local Interpret Model-agnostic Explanation (ICAV-LIME) combined method for model interpretability. This method is used to visualize the feature maps in different convolution layers in the CNN and calculate the importance of the input variables.

This paper will first discuss related works in Section 2, followed by data preparation and presentation in Section 3. Section 4 proposes the methodology, and Section 5 presents and discusses the results. Finally, the work is concluded in Section 6.

2 Related works

The analysis of short-term crash prediction has attracted considerable interest and focus in modern traffic analysis. The primary goal of short-term crash prediction is to proactively manage road safety by exploring the hypothesis that the likelihood of a crash occurring is connected to instantaneous dynamic traffic data during that period (Zheng and Sayed, 2020). However, due to the scarcity of actual crashes, crash risk and traffic conflicts are often utilized as proxies for real crash occurrences. Hossain and Muromachi (2012) constructed a Bayesian-based network to predict real-time crash risk instead of the actual number of crashes. Zheng and Sayed (2020) employed traffic conflicts extracted from informative vehicle trajectories as an intermediary for crash prediction. Machiani and Abbas (2016) proposed a real-time surrogate safety measure to assess safety at intersections based on traffic conflicts. Using actual crash data to develop real-time and short-term crash prediction models is inherently challenging due to the highly imbalanced data set. Additionally, while ARIMA has been extensively employed in constructing prediction models, it is unsuitable for short-term crash prediction because traffic-crash data consist of nonnegative, skewed distributions, thereby

violating the underlying normal distribution assumption of errors (Quddus, 2008; Cai et al., 2024).

Hossain et al. (2019) classified real-time crash prediction methods into two categories: statistical methods and data-driven methods. They demonstrated that commonly used statistical methods include logit and probit models, as well as mixed generalized linear models. Machine learning classification models and Bayesian networks are also popular data-driven methods. Basso et al. (2018) applied their model to raw imbalanced data without generating synthetic data. They initially employed a random forest method to determine the most significant variables and then developed an RTCP model using logistic regression and SVM, achieving an accuracy of 67.89% with a false positive rate of 20.94%.

Theofilatos et al. (2019) compared various machine learning and deep learning methods for crash prediction and demonstrated that deep neural network models exhibited the highest accuracy compared to random forests, decision trees, logistic regression, SVMs, and shallow neural networks. Yang et al. (2022) proposed the reinforcement learning tree method, which outperformed the deep neural network tree model by detecting 96% of crashes with a false alarm rate of 10%. Convolutional neural networks (CNNs) were initially developed by Yann LeCun, a computer scientist, during the 1980s (Lecun and Bengio, 1995). This type of artificial neural network utilizes a mathematical operation called convolution. CNNs were originally designed for image processing and focused on pixel manipulation; however, CNNs have since been applied to various other domains, including crash prediction. Basso et al. (2021) conducted research using a multiple-input CNN combined with a deep convolutional generative adversarial network, which yielded the most favorable results among oversampling methods. Rahim and Hassan (2021) also employed a CNN model alongside a customized fl-loss function. They transformed their crash data into images using a nonlinear dimensionality reduction technique called t-SNE and a convex hull algorithm.

The utilization of historical crash data entails handling highly imbalanced data sets due to the rarity of crash events. Consequently, oversampling methods have been extensively employed in the literature. Common techniques for generating synthetic data include the SMOTE, adaptive boosting, and generative adversarial networks (GANs). Man et al. (2019) not only demonstrated the effective use of GANs for synthetic data generation but also showed the model's success when applying transfer learning to other motorways. To address the issue of zero-inflation, some researchers suggest using models that inherently account for this aspect. Pew et al. (2020) argued that zero-inflated models are appropriate for specific scenarios and outperform other state-of-the-art approaches. Their work concluded that zero-inflated binomial regression is the most effective method. Song

et al. (2021) provided an example of the use of a zero-inflated binomial regression model in a case study involving freeway bridges. Additionally, CNNs are highly adept at handling imbalanced data due to their ability to recognize specific features. Yu et al. (2020) proposed a focal loss function that enhances the ability of CNNs to predict crashes in imbalanced data sets.

The significance of understanding the events leading up to a crash to predict the crash itself cannot be overstated. Therefore, incorporating time dependencies is crucial for real-time crash prediction. Numerous studies have employed time-series methods in an effort to address this challenge. Sun and Sun (2015) developed a dynamic Bayesian network (DBN) model using time sequence traffic data and compared it with a static Bayesian network. Their findings demonstrated that considering time improves both accuracy and transferability. Hassouna and Al-Sahili (2020) utilized an ARIMA method to examine the impact of the crash data set's size, measured in years, on the accuracy of predictions. Yuan et al. (2019) proposed a long short-term memory recurrent neural network (LSTM-RNN). While many studies have focused on real-time traffic conflict prediction methods, few have addressed real-time short-term crash prediction. The latter is critical because it allows proactive safety management to be implemented.

Some studies have combined spatial and temporal features. For instance, Li et al. (2020) developed an LSTM-CNN crash prediction model that captures temporal dependencies using LSTM and detects time-invariant attributes using a CNN. An improved approach incorporating an attention layer into the LSTM model has been proposed (Hema and Kumar, 2022; Li and Abdel-Aty, 2020). The inclusion of the attention layer has been shown to significantly enhance the model's performance in capturing time-series data, further highlighting the importance of considering time series in crash prediction.

LSTM models are commonly used to analyze temporal data, but in regard to studying traffic and crash data, they have some limitations. One major drawback is their inability to capture seasonality or calendar effects. Furthermore, the gated system of LSTMs does not consider time in the same way that traditional time-series methods do to perform partial autocorrelation. To address these limitations, this study proposes the use of a time series generalized regression neural network (Martinez et al., 2022). This method not only has an improved ability to capture time series patterns but is also more interpretable.

The concept of lag is employed to define the number of time steps the model should consider in the past. This can be either predetermined or fine-tuned during the training process by the model itself. Additionally, the model incorporates a weighted sum approach at the end of each layer, giving more importance to the most significant training patterns. The use of the radial basis function

(RBF) adds efficiency and speed to the training process, as RBFs have single-pass learning, only one parameter, and produce deterministic results (Martínez et al., 2022). Therefore, this method is well suited for real-time prediction of short-term traffic conditions and will perform effectively on the real-time data collected for this study.

Interpretation of the causes of traffic crashes is crucial for proactively preventing motor vehicle collisions. In a study conducted by Roilson et al. (2018), an assessment of the perspectives of law enforcement, the opinions of ordinary drivers, and road accident records was performed to investigate the main factors contributing to traffic crashes. Athiappan et al. (2022) utilized the Spearman ranking frequency index to rank influential factors that lead to traffic accidents and found that human factors were the primary cause of accidents, followed by vehicle factors and geometric factors. Elyassami et al. (2020) employed gradient boosting trees to predict traffic crashes and discovered that disregarding traffic signals, road design issues, poor visibility, and adverse weather conditions were influential factors in road accidents. The use of deep learning methods is often referred to as a “black box” due to challenges in interpreting the relationship between classification results and classifiers. Traditional machine learning model interpretation methods, such as random forests and gradient boosting, cannot be employed to conduct feature selection and analysis in deep neural networks. Therefore, a method known as local interpret model-agnostic explanation (LIME) is proposed in this study to analyze the contributing factors and variable importance in short-term crash incidents using a CNN model. LIME explains the predictions of any classifier by approximating it locally with an interpretable model (Ribeiro et al., 2016). A common approach to model-agnostic explanation is to learn a potentially interpretable model by capturing the explanation as a gradient vector that reflects a similar locality intuition on the predictions of the original model (Ribeiro et al., 2016). LIME is commonly used to explain the importance of classifiers in image classification and ensemble learning tree-based methods. In our study, it is applied to analyze crash-indicating factors in CNN classification.

3 Data presentation and processing

The MIDAS data set, which contains microscopic traffic data collected by loop detectors every minute, is stored in a database operated by Mott McDonald. This data set consists of five prediction variables: speed variance, speed, flow, occupancy, and headway. These variables effectively capture the dynamic traffic state. By accessing the location coordinates of the junctions and loop detectors, loop detectors can be classified into different road segments spanning junctions to junctions. They are also

separated in the northbound and southbound directions.

Averaging the loop detector data over road segments allows for compensation because certain loop detectors do not collect data at certain times. The traffic variables were averaged over all the lanes and a period of 30 min, except for flow, which was summed over that time period and then averaged over all the lanes. Additionally, the speed variance was calculated and added as a new variable. Our data cover junctions 1 to 12 and junctions 15 to 28 in both directions.

Crashes that occurred on M1 in 2019 were extracted from the database. A map matching algorithm was applied to compute the exact coordinates of the crash locations (Quddus et al., 2003). Based on the precise location, the crashes were classified into the different road segments used to group loop detector traffic data. The direction of the vehicle just before the crash was used to determine the vehicle’s direction, as the traffic data were separated for northbound and southbound travel. When the direction was not available, it was randomly assigned. As the traffic data are formatted in 30-min intervals spanning the entire year, the crashes need to be assigned to a 30-min interval. This was achieved using the date and time of the crashes and converting it to the corresponding traffic data index. Once the time index was computed, a new crash variable was added to the traffic data, indicating whether a crash occurred (1) or not (0).

The M1 motorway is one of the most important motorways in England, connecting London to Leeds and passing through several important cities such as Milton Keynes, Northampton, Leicester, Nottingham, and Sheffield. It was the first interurban motorway built in the UK. The data set built for this study covers a significant portion of the motorway, starting at junction 1 near London and ending at junction 25 near Nottingham. However, due to construction work for the smart motorway project, junctions 13 to 15 were not available in 2019.

The traffic variables are presented in Table 1, with descriptive statistics of the data for the whole year of 2019 for junctions 1 to 2 as an example.

4 Methodology

Short-term crash data exhibit high dispersion and sparsity. The presence of extremely imbalanced data sets poses challenges in capturing pro-crash traffic dynamics and accurately predicting crashes. The temporal and spatial characteristics of dynamic traffic variables play significant roles in crash occurrence. This paper proposes a model composed of two components: (i) a time series generalized regression and (ii) a binomial weighted convolutional neural network. The model structure is comprehensive because it addresses temporal characteristics such as autocorrelation and seasonality, along with spatial correlation and imbalanced sample issues, by

Table 1 Descriptive statistics for explanatory variables

Variable	Label	Flow	Speed	Occupancy	Headway	Speed variance
Unit	Crash:1 Noncrash:0	Vehicles/30 min per lane	km/h	Vehicles/m	m	km/h
Mean	0	131.47	81.78	5.06	84.22	62.00
	1	133.75	87.82	2.97	96.43	60.67
Standard deviation	0	4.56	14.20	9.78	20.91	29.62
	1	19.04	1.96	0.58	6.71	21.88
Min	0	0.00	6.70	0.00	0.00	0.00
	1	112.25	85.56	2.34	92.08	35.84
Max	0	441.0	99.00	73.38	139.63	481.00
	1	148.5	88.98	3.48	104.16	77.10

extracting crash features from the projected chromagram image. The functional architecture of the joint model is illustrated in Fig. 1. Initially, a TSGRNN was employed to predict future traffic states X_{t+1} based on a sequence of past and current traffic states $X_{t-k, \dots, t}$. The traffic state is characterized by five variables: speed, speed variance, flow, occupancy, and headway. The output of this component is then fed into a Binomially WCNN to predict the crash probability \hat{Y}_{t+1} in the next 30-min interval. However, merely predicting the crash probability is insufficient for a traffic control center to take any intervention measures without understanding the contributing factors to collision risk. Hence, a forensic analysis is conducted on the predictive model's results to identify such factors and enable targeted actions for crash prevention. The following section offers a brief discussion of the two components of the model.

4.1 Time series generalized regression neural network (TSGRNN)

The generalized regression neural network (GRNN) is a generalized version of the ANN that has shown superior performance in various domains compared to ANNs (Martinez, et al., 2022; Martínez-Blanco, et al.2016; Martinez, et al., 2018.). In contrast to the commonly used back propagation in ANNs, the GRNN utilizes a memory-based RBF network as a hidden layer, allowing for the estimation of continuous variables and convergence to linear or nonlinear regression surfaces (Specht, 1991). This makes the GRNN suitable for handling sparse data in multidimensional measurement spaces, enabling both classification and regression tasks in cases where linearity cannot be justified (Specht, 1991).

However, the conventional GRNN fails to incorporate the time series structure, including autocorrelation, seasonality, and moving average, which are crucial in temporal-related use cases. To address this limitation, an RBF neural network is introduced. It projects nonlinear data onto a hyperplane and includes temporal parameters

to capture time features in the following ways: 1) it uses single-pass learning; 2) it sets smoothing parameters to tune the weighted average in the output; and 3) it produces deterministic results after choosing the smoothing parameter (Martinez, et al., 2022). This modified model is referred to as the TSGRNN.

The model and its application can be represented as follows:

$$x_{t+1} = f(X_{t-k, \dots, t}), \quad (1)$$

where X is a vector of the traffic characteristics, f is the time-series GRNN model and k is the lag value. The GRNN is composed of three layers: an input layer, a hidden layer of RBF neurons, and an output layer. The radial basis function is the multivariate Gaussian function:

$$G(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right), \quad (2)$$

where x_i represents the center parameter, σ^2 represents the smoothing parameter, and x is the input variable vector. The structure of the GRNN can be displayed as follows in Fig. 2.

Each data point in the time series inputs is assigned a weight proportional to its proximity to the prediction time, thus emphasizing the importance of closer values over those further away. The values are then aggregated in a weighted sum, as presented in Eq. (3):

$$\hat{y} = \sum_{i=1}^n w_i y_i, \quad w_i = \frac{\exp\left(-\frac{\|y - y_i\|^2}{2\sigma^2}\right)}{\sum_{j=1}^n \exp\left(-\frac{\|y - y_j\|^2}{2\sigma^2}\right)}. \quad (3)$$

The weights sum to one, and the smoothing parameter influences the number of variables that play a significant role in the outcome (Martinez, et al., 2022). When the smoothing parameter σ is large, the weights are small and similar, and the output is close to the target average.

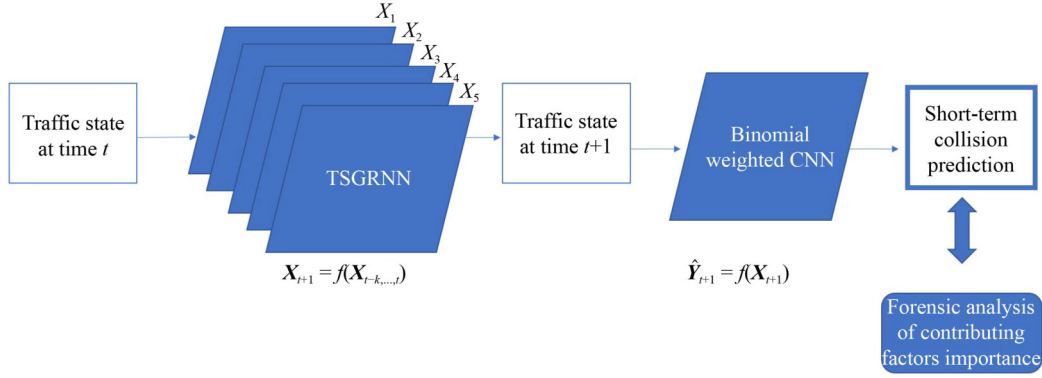


Fig. 1 Functional architecture of the joint model.

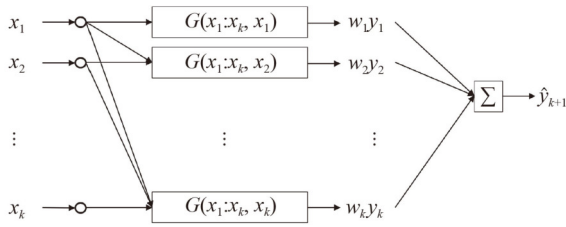


Fig. 2 TSGRNN structure (adapted from Martinez et al., 2022).

If the smoothing parameter σ is small, only the closest targets have significant weights that influence the prediction (Martinez, et al., 2022).

In the interim, the selection of appropriate lag values is crucial when considering seasonality. The autocorrelation (ACF), partial autocorrelation (PACF), and time differencing methods proposed in the ARIMA model can be utilized to determine lag values. Generally, if there is evidence of a seasonal period, lags should be chosen based on the seasonal patterns observed, such as weekly, monthly, or quarterly. The significant lag values are then selected to establish training patterns for testing the linear relationship (Martinez, et al., 2022).

4.2 Binomial weighted sequential convolutional neural network (WCNN)

Convolutional neural networks are commonly employed in tasks such as image recognition, pattern recognition, and signal processing due to their ability to learn translation invariant patterns with spatial hierarchies (Chollet, 2018). CNNs, as powerful classifiers for feature extraction and spatial correlation analysis, have been applied to various anomaly detection tasks involving highly imbalanced data sets. In this context, CNNs project sequential data onto chromagram images and employ convolution neurons to extract spatial features for classification purposes. To handle extremely imbalanced data, CNNs assign different weights to different classes. The less representative class typically receives a higher weight, compelling CNNs to adjust the weights in the direction of

the gradient descent during the training process (Han and Jeong, 2020). This paper proposes a binomial sampling weighted CNN to capture the overdispersion characteristics in traffic crash data. Dynamic weights following a binomial distribution are assigned to each training step when fitting the CNN model by tuning the parameters to increase the probability of correctly classifying traffic crashes from noncrash events. The CNN model proposed in this study has a sequentially based structure consisting of three convolutional layers and two densely connected layers, forming a formidable learner. Dropout is also utilized to mitigate overfitting. The rectified linear unit function is used as the activation function. In total, 20,898 parameters are tuned through a backpropagation process. The architecture of this deep learning neural network consists of 30 epochs, with each epoch trained to tune the parameters over 100 steps. The structure is illustrated in Fig. 3.

SMOTE was used to generate crash data for training the CNN model, enabling the learner to stabilize its gradients and tune the parameters with an adequate amount of crash data. However, the model was tested using the original data set to assess its accuracy in real-world scenarios, thereby validating its suitability for industry applications.

4.3 Joint model of the generalized regression neural network and weighted CNN (TGRCN)

The conventional spatiotemporal model predicts the future state at time $t+1$ by considering data from time $t-k$ to t . Although this model captures the autocorrelated temporal structure and incorporates spatial characteristics with a one-step lag, it fails to account for the spatial correlation between traffic crashes and dynamic traffic data occurring in the same time dimension. As a result, a novel approach is needed to accurately predict traffic crashes by jointly modeling dynamic traffic data in the same time dimension. In this study, we propose a joint model called the TGRCN. The TGRCN model combines the TSGRNN model, which predicts the values of covariates at time $t+1$, with the weighted CNN (WCNN)

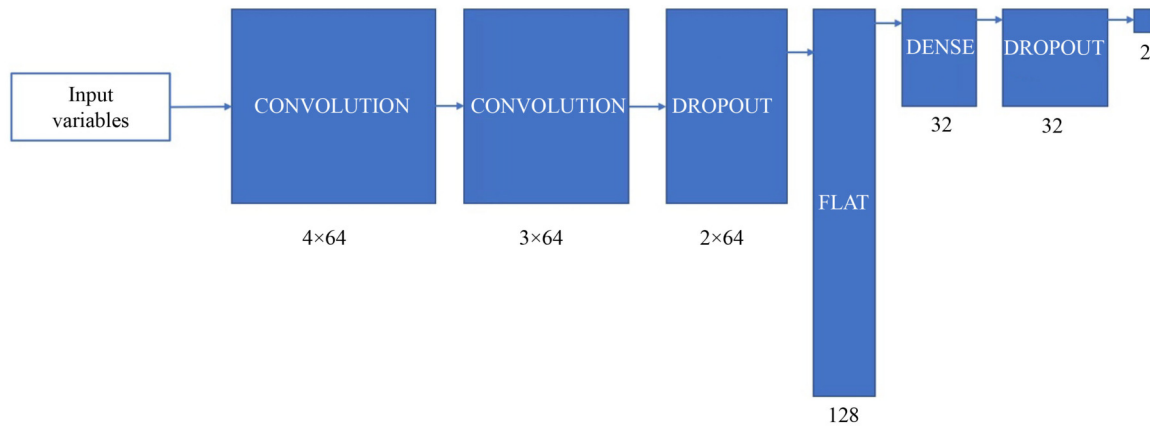


Fig. 3 Sequential convolutional neural network structure.

model, which forecasts the probability of a crash event in the same time dimension. This unique framework allows us to capture spatial hierarchies and analyze spatial features rather than solely relying on temporal autocorrelation. Utilizing binomial weights empowers CNNs with strong classification abilities, specifically in detecting and studying underrepresented classes. Consequently, the TSGRNN model is utilized to obtain the predicted values of dynamic traffic variables at time $t + 1$, serving as input for the CNN to construct the chromagram image at the same time stamp. The TSGRNN learns temporal features, while the weighted CNN extracts spatial correlation features within the same time dimension. By leveraging both methods, the joint model effectively decomposes crash characteristics and maximizes the prediction ability of the TSGRNN and CNN, compensating for their respective limitations.

To evaluate the performance of the TGRCN model, we compare it with state-of-the-art models that capture both spatial and temporal components in crash data. One such model is the STGCN. The STGCN model consists of multiple spatiotemporal convolutional blocks, each composed of two gated sequential convolution layers and one spatial graph convolution layer (Yu et al., 2020). To address the gradient overdispersion issue and reduce network load, residual connections and a bottleneck strategy are implemented within each block (Yu et al., 2020). The input, v_{t-M+1}, \dots, v_t , is uniformly processed by these spatiotemporal blocks to effectively handle spatial and temporal dependencies. The output layer then integrates these comprehensive features to generate the final prediction (Yu et al., 2020).

In this paper, we propose the Transformer-based Spatial-Temporal Transformer Network (STTNs), along with the Graph Convolutional Network Accelerator on Versal ACAP Architecture (H-GCN) and the Temporal Graph Convolutional Network (T-GCN), to predict near-future crashes (Zhang et al., 2022; Zhao et al., 2018). STTNs incorporates a graph neural network (GNN) variant called spatial transformer, which dynamically models

directed spatial dependencies using a self-attention mechanism to capture real-time conditions and traffic directions (Xu et al., 2020). Xu et al. (2020) defined a short-term temporal range as less than 30 min and a long-term temporal range as greater than 30 min. STTNs leverages both dynamical directed spatial dependencies and long-range temporal dependencies to improve the accuracy of long-term traffic condition prediction (Xu et al., 2020).

H-GCN, proposed by Zhang et al. (2022), is a programmable logic and AI engine-based hybrid accelerator that utilizes the emerging heterogeneity of Xilinx Versal Adaptive Compute Acceleration Platforms to achieve high-performance GNNs. Each graph is partitioned into three subgraphs based on inherent heterogeneity (Zhang et al., 2022). H-GCN is also employed in traffic flow prediction, offering fast speed and moderate accuracy. The T-GCN combines a graph convolutional network (GCN) with a gated recurrent unit (GRU) (Zhao et al., 2018). The GCN captures spatial dependence, while the GRU learns the dynamic changes in temporal traffic data structures (Zhao et al., 2018). The T-GCN is widely used for predicting traffic dynamic patterns at urban junctions.

4.4 Causal interpretable implications of variable contributions from TGRCN predictions

Causal analysis and model-based interpretation go beyond association analysis and correlation. They incorporate concepts such as intervention analysis and counterfactual reasoning (Dablander, 2020). These approaches provide visual diagrams to interpret the statistical dependencies between independent variables and outcomes. Causal effect displays employ *do-calculus* to examine the influence of X leading to a change in Y (Cartwright, 2010). Deep learning models are often considered “black boxes” due to their complex network structures and lack of interpretability. The visualization and interpretation of CNN models can be particularly challenging due to the sequential and densely connected convolutional layers

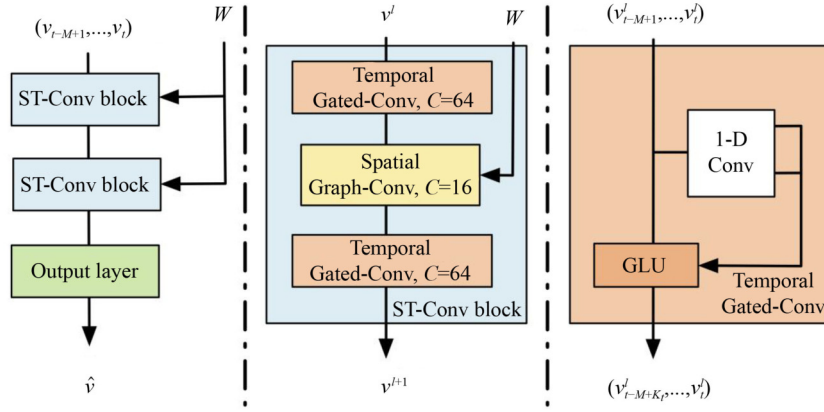


Fig. 4 Architecture of the STGCN.

(convnets). One commonly used method for understanding how input layers are transformed into successive convolutional layers is intermediate ConvNet activation visualization (ICAV) (Chollet and Allaire, 2018). By visualizing feature maps, which consist of three dimensions (width, height, and depth/channels), we can observe the independent features encoded by each channel. This allows us to see what features are learned by each layer in the CNN (Chollet and Allaire, 2018). To interpret CNN feature maps, we can apply the LIME technique. LIME assumes that the CNN model is linear on a local scale and can mimic the performance of the global model within a specific locality (Ribeiro et al., 2016). The algorithm logic is as follows (Ribeiro et al., 2016):

- (1) Permute observation to create replicated feature data with slight value randomness.
- (2) The similarity measure between the original observation and the permuted observations is computed.
- (3) At each iteration:
 - 1) A CNN is fit to a portion of the variables from the permuted data set.
 - 2) The feature importance was assessed by testing the model accuracy on another permuted data set scaled by its similarity to the original observation.
 - 3) The two previous steps are repeated by adding or removing a feature at each iteration to compare feature importance in terms of the prediction accuracy obtained.
 - 4) The resulting feature weights are used to explain local behavior that reflects the importance of individual features.

By doing so, variable importance can be inferred from the joint method of ICAV-LIME.

5 Findings

In our data preprocessing, we found that approximately 0.5% of the traffic data were missing. To address this issue, we utilized the MissForest R package, which is capable of filling in missing data considering the presence

of mixed data types. MissForest employs an iterative nonparametric imputation technique based on random forest ensemble learning (Stekhoven et al., 2012). This method imputes missing values by taking the average over numerous unpruned classification or regression trees and uses the built-in out-of-bag error to estimate the imputation errors. To evaluate the suitability of random forest imputation for missing values in our data set, we intentionally generated “NA” (missing) values for 1% of the known values in the training set. Subsequently, we employed random forest to impute the true values for the artificially generated missing values. The imputation accuracy is shown in Table 2.

As presented in Table 2, there is a close alignment between the average of the imputed values and the average of the actual values in the training set. Based on this observation, it is justifiable to utilize random forest imputation techniques to address the missing values in the test set given that the proportion of missing values is relatively low.

To ensure the stability of the prediction model, a training set consisting of data from the past six months was used. During the fitting process, the model parameters were fine-tuned using three months of data, which was designated the validation set. The remaining three months of data were set aside as the test set to evaluate the accuracy of the predictions. As previously discussed, the TSGRNN model was applied to make predictions of dynamic traffic variables at intervals of 30 min from time $t-1$ to time t . The TSGRNN model trained the tuning parameters for autocorrelation lags and weights by fitting a regression model with the training data set. In the validation set, the smoothing and center parameters, which control autocorrelation and seasonality in the TSGRNN, were further fine-tuned for optimization. In the testing data set, the tuned parameters were fixed at the optimal values to make predictions of traffic variables for the next half-hour.

Initially, a WCNN model was trained on an oversampled training data set generated using the SMOTE method.

Table 2 Imputation accuracy for missing values in validation

Average	Flow	Speed	Speed variance	Occupancy	Headway
Actual	262.0	85.8	7.69	68.17	74.25
Predicted	259.3	83.8	6.59	70.21	76.55

The purpose of this was to create a robust classifier capable of handling the extreme class imbalance in the data. Subsequently, the trained model was validated using a three-month validation set. This validation set consisted of predicted dynamic traffic data obtained from the TSGRNN, which served as input variables for the weighted convolutional neural network. The validation process aimed to assess the generalizability and accuracy of the model and to fine-tune the initial weights for both class and nonclass events. The TGRCN model was built to predict crash occurrences in the following 30-min time window in the unseen test set. Sensitivity, specificity, negative predictive value (NPV), accuracy, and precision were selected as the metrics for evaluation, as described by Formosa et al. (2023). Given the imbalanced data, “precision” may be a more appropriate criterion than “accuracy” for measuring the prediction power, as suggested by Yang et al. (2023).

The results are presented in Table 3.

As indicated in Table 3, the TGRCN model demonstrates a commendable prediction rate for both crash and noncrash events in the north and south directions.

Specifically, the model achieves an approximately 78% prediction rate for crash events and 80% for noncrash events in the north direction, compared to 75% for crash events and 81% for noncrash events in the south direction. It is important to note that these results correspond to the proportion of events in the original real-world data set, not the oversampled data. Considering that more than 99.99% of the data consist of noncrash events, the model’s ability to accurately classify the majority of crash and noncrash events within the next 30 min is impressive. This performance has significant implications for real-time short-term traffic forecasting scenarios, where precise classification is essential.

For comparison purposes, the STGCN model was also applied to the same data set to test its prediction accuracy. To enhance its classification ability, the SMOTE technique was employed to generate pseudo crash data in the training set as an oversampling technique for model comparability. Additionally, binomial weights were assigned to create an attention mechanism during training of the STGCN model. However, the prediction results were not satisfactory compared to those of the TGRCN model, as shown in Table 3. While the STGCN model achieves high accuracy in noncrash prediction, it fails to capture the characteristics of crash data and misclassifies the majority of crash events. Therefore, the TGRCN model outperforms the STGCN model in short-term crash prediction. Similarly, the MLP and LSTM models

Table 3 Prediction accuracy comparison for the TGRCN and other predictive models for the northbound and southbound data of the M1 Motorway, UK

Method	Direction	Metrics employed to compare the predictive performance				
		Sensitivity	Specificity	NPV	Accuracy	Precision
H-GCN	North	0.124	0.954	0.519	0.539	0.806
	South	0.101	0.943	0.509	0.519	0.642
T-GCN	North	0.089	0.957	0.512	0.523	0.674
	South	0.088	0.969	0.515	0.529	0.739
STTNS	North	0.222	0.914	0.540	0.568	0.721
	South	0.215	0.930	0.542	0.573	0.754
STGCN	North	0.351	0.944	0.589	0.644	0.862
	South	0.292	0.958	0.572	0.622	0.874
MLP	North	0.086	0.972	0.532	0.529	0.754
	South	0.078	0.935	0.504	0.507	0.545
LSTM	North	0.382	0.862	0.582	0.622	0.735
	South	0.366	0.881	0.582	0.624	0.755
GARCH	North	0.252	0.956	0.561	0.5	0.851
	South	0.233	0.964	0.557	0.599	0.866
ARIMA	North	0.121	0.97	0.525	0.545	0.801
	South	0.118	0.962	0.522	0.54	0.756
TGRCN (Developed in paper)	North	0.781	0.802	0.786	0.792	0.798
	South	0.753	0.816	0.768	0.789	0.812

accurately predict noncrash data, but they lack the ability to differentiate crash data by utilizing dynamic traffic parameters. Traditional time-series models such as the ARIMA and GARCH models were also used to test the prediction accuracy in near-future traffic crash analysis. However, their ability to predict near-future traffic crashes is inferior to that of deep learning-based neural network models. This may be attributed to the limited feature capturing ability of conventional time-series models in scenarios characterized by an excess of zeros. Although the TGRCN model sacrifices a slight decrease in its ability to accurately predict noncrash data, it achieves a good trade-off by forecasting more crash incidents.

Both the H-GCN and T-GCN models struggle to classify crash and noncrash data accurately. The STTNS model performs better than the H-GCN and T-GCN models in crash prediction, but it still falls short compared to the TGRCN model. One possible explanation for the inferior performance of the STTNS model to that of the TGRCN model is that traffic crashes are rare incidents with high uncertainty and randomness. Although transformer-related models are widely used for predicting traffic patterns such as traffic flows and speeds, they may not be suitable for handling imbalanced classification problems. Very limited literature is available on the use of GNNs or transformer-related networks for short-term traffic crash prediction.

The ICAV-LIME method was utilized to analyze the importance of dynamic traffic variables in crash occurrence. The results demonstrate that speed variance is the most crucial factor in crash prediction. In addition to speed variances, speed, headway, occupancy, and traffic flow are all contributing factors and have similar effects on crash occurrence. The ranking order of variable contributions is as follows: speed variances, speed, headway, occupancy, and traffic flow (refer to Fig. 5).

6 Discussion and conclusions

This paper developed an integrated method consisting of a time series generalized regression neural network and a weighted convolutional neural network for short-term crash prediction within 30-min intervals. The joint model,

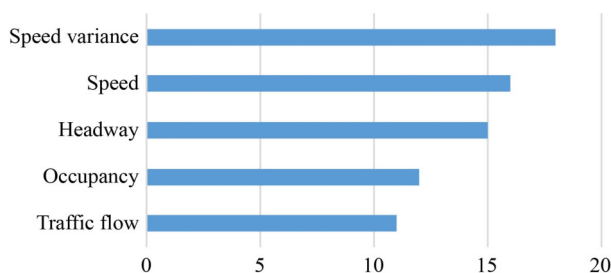


Fig. 5 Variable importance for explanatory variables.

named the TGRCN, effectively captures both spatial and temporal features present in segment-based crash data. The results indicate that the TGRCN model outperforms other state-of-the-art models, such as the STGCN, MLP, and LSTM, in predicting near-future crashes. Speed variances emerge as the most critical contributing factor to crash occurrence. While the literature has focused primarily on traffic conflicts due to limitations in collecting historical crash data, predicting real crashes has proven to be more challenging due to the significantly lower likelihood of occurrence. Furthermore, the TGRCN model offers increased applicability to real-world scenarios.

The integrated method developed in this study, the TGRCN, is a novel approach that combines the TSGRNN with the WCNN to address crash features. Nevertheless, further evaluation of various and generalized CNN structures should be undertaken to determine their suitability for crash prediction. Future research should also focus on studying the dynamic threshold for classifying crash versus noncrash data in the test set, considering that the middle value of 0.5 may not be the optimal choice for predicting crashes with imbalanced data.

Furthermore, the ICAV-LIME method was employed for model interpretability. The analysis revealed that speed variance, speed, and headway are the three most significant variables related to traffic crashes. The significance of speed variance as a traffic characteristic affecting road safety becomes evident, in addition to the importance of speed and headway. Previous studies have confirmed the positive relationship between speed variance and accident rates (Nicholas and Ravi, 1989). Solomon (1964) and Cirillo (1968) found that crash involvement rates increased as vehicle speeds deviated from the average speed of the traffic stream.

To assess the robustness of the proposed TGRCN model across different data sets, hourly traffic data from the Sutong Bridge Freeway were utilized. The Sutong Bridge, which is situated on the Yangtze River and connects Nantong and Suzhou in China, is a cable-stayed bridge spanning 32.4 km (20.2 mi) (Cai et al., 2024). The extracted hourly traffic data cover a 9-month period from July 2020 to April 2021, with 91% of the data being noncrash related. A range of econometric and deep learning-based prediction and classification methods were tested and compared. The sensitivity, specificity, precision, NPV, and accuracy are employed as metrics for comparison. The findings are presented in Table 4.

Consistent with previous findings, the TGRCN model continues to outperform other candidate methods, demonstrating its transferability and robustness across different scenarios. Furthermore, deep learning neural network models exhibit superior prediction and classification capabilities compared to conventional econometric time series methods.

Several limitations of the TGRCN model have been identified, namely, (i) the dynamic traffic parameters

Table 4 Prediction accuracy comparison for the TGRCN and other prediction models for Sutong Bridge traffic data

Methods	Sensitivity	Specificity	Precision	NPV	Accuracy
H-GCN	0.158	0.888	0.585	0.513	0.523
T-GCN	0.141	0.890	0.562	0.509	0.516
STTNS	0.295	0.842	0.651	0.544	0.569
MLP	0.028	0.98	0.583	0.502	0.504
GARCH	0.393	0.917	0.826	0.602	0.655
ARIMA	0.101	0.89	0.479	0.497	0.496
STGCN	0.455	0.891	0.807	0.621	0.673
LSTM	0.567	0.902	0.852	0.676	0.735
TGRCN	0.667	0.909	0.879	0.732	0.788

predicted by the GRNN may contain a margin of error. Although predictors are converted to the same dimension, utilizing predictors with a degree of error to forecast the near future could introduce uncertainty and noise that are difficult to measure; (ii) While this study employed the WCNN, the structure of the CNN is fairly simple. Deeper networks are more suitable for other input data types, such as images or audio data; however, for numerical data, deeper networks might result in overfitting. It would be beneficial to explore the suitability of generalized parallel network structures, such as Inception Net, or networks with attention mechanisms such as “Transformers”; (iii) Currently, the data set lacks a sufficient number of crashes; thus, it is crucial to collect and utilize additional data across a broader range of scenarios to further enhance the performance of the TGRCN model.

In summary, this paper proposes a novel approach to crash prediction by incorporating a time-series GRNN and a binomial weighted CNN model. Unlike previous studies that treated crash prediction as a classification problem without considering the temporal structures in crash data, the proposed method takes into account the temporal dependencies. The GRNN is applied to capture the temporal features of the data, while the binomial weighted CNN model optimizes gradient descent by learning spatial features. This application of a time-series-based GRNN in crash prediction is a unique contribution.

The joint model, known as the TGRCN, utilizes dynamic traffic data at time t to predict real-time crashes in the same time dimension. It transfers $t - 1$ traffic data to predict values at time t using the GRNN. The insights gained from this research can assist highway operators in predicting the likelihood of a crash in advance and implementing interventions such as enforcing speed limits.

Author Statement The authors confirm contribution to the paper as follows: study conception and design, data collection, analysis and interpretation of results, draft manuscript preparation: Bowen Cai, Léah Camarcat, Mohammed Quddus. All authors reviewed the results and approved the final version of the manuscript.

Competing Interests The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Athiappan K, Karthik C, Rajalaskshmi M, Subrata C, Dastjerdi H, Liu Y, Campusano C, Gheisari M (2022). Identifying influencing factors of road accidents in emerging road accident blackspots. *Advances in Civil Engineering*, 2022: 1–10
- Basso F, Basso L, Bravo F, Pezoa R (2018). Real-time crash prediction in an urban expressway using disaggregated data. *Transportation Research Part C, Emerging Technologies*, 86: 202–219
- Basso F, Pezoa R, Varas M, Villalobos M (2021). A deep learning approach for real-time crash prediction using vehicle-by-vehicle data. *Accident Analysis & Prevention*, 162: 106409, ISSN 0001-4575
- Cai B, Quddus M, Ma L (2022). High accurate deep learning models for estimating traffic characteristics from video data. Presented in 101st Transportation Research Board
- Cai B, Quddus M, Wang X, Miao Y (2024). New modeling approach for predicting disaggregated time-series traffic crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2678(3): 637–648
- Cartwright N (2010). Hunting causes and using them: Approaches in philosophy and economics: summary. *Analysis*, 70(2): 307–310, ISSN 0003-2638
- Chollet F, and Allaire J (2018). Deep Learning in R. *R-bloggers*, 1(7080): 360.
- Cirillo J (1968). Interstate System Accident Research Study II, Interim Report II. *Public Roads*, 35(3)
- Dablander F (2020). An introduction to causal inference
- Deva Hema D, Ashok Kumar K (2022). Novel algorithm for multivariate time series crash risk prediction using CNN-ATT-LSTM model. *Journal of Intelligent & Fuzzy Systems*, 43(4): 4201–4213
- Elyassami S, Hamid Y, Habuza T (2020). Road crashes analysis and prediction using gradient boosted and Random Forest Trees. In: 6th IEEE Congress on Information Science and Technology (CiSt), 520–525
- Formosa N, Quddus M, Man C K, Timmis A (2023). Appraising machine and deep learning techniques for traffic conflict prediction with class imbalance. *Data Science for Transportation*, 5(2): 4:25
- Ghanipoor Machiani S, Abbas M. (2016). Safety Surrogate

- Histograms (SSH): A novel real-time safety assessment of dilemma zone related conflicts at signalized intersections. *Accident Analysis and Prevention*, 96: 361–370
- Han S, Jeong J (2020). An weighted CNN ensemble model with small amount of data for bearing fault diagnosis. *Procedia Computer Science*, 175: 88–95
- Hassouna F, Al-Sahili K (2020). Practical minimum sample size for road crash time-series prediction models. *Advances in Civil Engineering*, 2020(1): 6672612
- Hossain M, Abdel-Aty M, Quddus M, Muromachi Y, Sadeek S (2019). Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. *Accident Analysis and Prevention*, 124: 66–84
- Hossain M, Muromachi Y (2012). A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident Analysis and Prevention*, 45: 373–381
- Huang T, Wang S, Sharma A (2020). Highway crash detection and risk estimation using deep learning. *Accident Analysis and Prevention*, 135: 105392
- Lee C, Hellinga B, Saccomanno F (2003). Real-time crash prediction model for application to crash prevention in freeway traffic. *Transportation Research Record*, 1840(1): 67–77
- Lecun Y, Bengio Y (1995). Convolutional networks for images, speech, and time-series. In M. A. Arbib (Ed.), *The Handbook of Brain theory and Neural Networks*. MIT Press
- Li P, Abdel-Aty M, Yuan J (2020). Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis and Prevention*, 135: 105371
- Martínez F, Charte F, Frías M P, Martínez-Rodríguez A M (2022). Strategies for time series forecasting with Generalized Regression Neural Networks. *Neurocomputing*, 491: 509–521
- Martínez-Blanco M R, Ornelas-Vargas G, Solís-Sánchez L O, Castañeda-Miranada R, Vega-Carrillo H R, Celaya-Padilla J M, Garza-Veloz I, Martínez-Fierro M, Ortiz-Rodríguez J M (2016). A comparison of back propagation and generalized regression neural networks performance in neutron spectrometry. *Applied Radiation and Isotopes*, 117: 20–26
- Martinez C, Heucke M., Wang, F, Gao B, Cao D (2018). Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 19(3): 666–676
- Man C K, Quddus M, Theofilatos A (2022). Transfer learning for spatio-temporal transferability of real-time crash prediction models. *Accident Analysis and Prevention*, 165: 106511
- Man C, Quddus M, Theofilatos A, Yu R, Marianna-Ioanna Imprialou (2019). Utilising Generative Adversarial Network (GAN) to address the imbalanced data issue in real-time crash risk prediction. In: *Transportation Research Board 99th Annual Meeting*, Washington D.C., USA, 2020
- Nicholas J, Ravi G (1989). Factors affecting speed variance and its influence on accidents transportation research record. *Journal of the Transportation Research Board*, (1213): 64–71
- Pew T, Warr R L, Schultz G G, Heaton M (2020). Justification for considering zero-inflated models in crash frequency analysis. *Transportation Research Interdisciplinary Perspectives*, 8: 100249
- Quddus M A, Ochieng W Y, Zhao L, Noland R B (2003). A general map matching algorithm for transport telematics applications. *GPS Solutions*, 7(3): 157–167
- Quddus M A (2008). Time series count data models: An empirical application to traffic accidents. *Accident Analysis and Prevention*, 40(5): 1732–1741
- Rahim M A, Hassan H M (2021). A deep learning based traffic crash severity prediction framework. *Accident Analysis and Prevention*, 154: 106090
- Rolison J, Regev S, Moutari S, Feeney A (2018). What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records. *Accident Analysis and Prevention*, 115: 11–24
- Ribeiro M T, Singh S, Guestrin C (2016). Why should I trust you? Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, 1135–1144
- Stekhoven D, Bühlmann P (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 8: 1201, 112–118
- Song J, Huang B, Wang Y, Wu C, Zou X, Zhang J, Li J (2021). A zero-inflated negative binomial crash prediction model for freeway bridge sections. In: *CICTP 2021: Advanced Transportation, Enhanced Connection—Proceedings of the 21st COTA International Conference of Transportation Professionals*, 1227–1236
- Solomon D (1964). *Accidents on main rural highways related to speed, driver, and vehicle*. Federal Highway Administration, Washington, DC (Reprinted 1974)
- Specht D F (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6): 568–576
- Sun J, Sun J (2015). A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. *Transportation Research Part C, Emerging Technologies*, 54: 176–186
- Theofilatos A, Chen C, Antoniou C (2019). Comparing machine learning and deep learning methods for real-time crash prediction. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(8): 169–178
- World Health Organisation (WHO). *Road Traffic Injuries*. Available at the website of WHO
- Xu M, Dai W, Liu C, Gao X, Lin W, Qi G, X H (2020). Spatial-temporal transformer networks for traffic flow forecasting, arXiv: 2001.02908
- Yang K, Quddus M, Antoniou M (2022). Developing a new real-time traffic safety management framework for urban expressways utilizing Reinforcement Learning Tree. *Accident Analysis and Prevention*, 178: 106848
- Yang Y, Rasouli S, Liao F (2023). Effects of life events and attitudes on vehicle transactions: A dynamic Bayesian network approach. *Transportation Research Part C, Emerging Technologies*, 147: 103988
- Yu R, Wang Y, Zou Z, Wang L (2020). Convolutional neural

- networks with refined loss functions for the real-time crash risk analysis. *Transportation Research Part C, Emerging Technologies*, 119: 102740
- Yuan J, Abdel-Aty M, Gong Y, Cai Q (2019). Real-time crash risk prediction using long short-term memory recurrent neural network. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(4): 314–326
- Zhang C, Geng T, Guo A, Tian J, Herboldt M, Li A, Tao D (2022). H-GCN: A graph convolutional network accelerator on versal ACAP architecture. arXiv: 2206.13734
- Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2018). T-GCN: A temporal graph convolutional network for traffic prediction. arXiv: 1811.05320
- Zheng L, Sayed T (2020). A novel approach for real time crash prediction at signalized intersections. *Transportation Research Part C, Emerging Technologies*, 117: 102683