



Mid-term prediction of at-fault crash driver frequency using fusion deep learning with city-level traffic violation data

Yuan-Wei Wu ^{*}, Tien-Pen Hsu

Department of Civil Engineering, National Taiwan University, Taipei, 106, Taiwan



ARTICLE INFO

Keywords:

At-fault crash driver
Traffic violation
Traffic enforcement
Deep learning

ABSTRACT

Traffic violations and improper driving are behaviors that primarily contribute to traffic crashes. This study aimed to develop effective approaches for predicting at-fault crash driver frequency using only city-level traffic enforcement predictors. A fusion deep learning approach combining a convolution neural network (CNN) and gated recurrent units (GRU) was developed to compare predictive performance with one econometric approach, two machine learning approaches, and another deep learning approach. The performance comparison was conducted for (1) at-fault crash driver frequency prediction tasks and (2) city-level crash risk prediction tasks. The proposed CNN-GRU achieved remarkable prediction accuracy and outperformed other approaches, while the other approaches also exhibited excellent performances. The results suggest that effective prediction approaches and appropriate traffic safety measures can be developed by considering both crash frequency and crash risk prediction tasks. In addition, the accumulated local effects (ALE) plot was utilized to investigate the contribution of each traffic enforcement activity on traffic safety in a scenario of multicollinearity among predictors. The ALE plot illustrated a complex nonlinear relationship between traffic enforcement predictors and the response variable. These findings can facilitate the development of traffic safety measures and serve as a good foundation for further investigations and utilization of traffic violation data.

1. Introduction

After the United Nations Road Safety Collaboration (United Nations, 2010) proposed “A Global Plan for the Decade of Action for Road Safety 2011–2020,” considerable efforts have been invested in reducing road fatalities and injuries worldwide. In Taiwan, traffic crash fatalities steadily decreased from 3,279 in 2010 to 2,780 in 2018, which is a striking decline rate of 15.2 % (Ministry of Transportation and Communications, 2020). However, over the same period, traffic crash injuries increased from 292,535 to 426,799, which is a significant increase rate of 45.9 %; further, at-fault crash driver frequency increased by 36.9 %. In 2018 alone, 85.3 % of total crashes in Taiwan were caused owing to traffic violations. Therefore, to minimize the number of deaths and injuries caused by road traffic accidents, traffic safety studies and practical safety measures need to focus on at-fault crash drivers.

Traffic violations and improper driving by at-fault crash drivers are behaviors that primarily contribute to traffic crashes and increase the risk of traffic crash for drivers (Antonopoulos et al., 2011; Al Naser et al., 2013; Xu et al., 2018; Mehdizadeh et al., 2019; Moghaddam et al.,

2017). In addition, traffic violations increase the likelihood of sustaining severe injuries (Duddu et al., 2018; Pai and Saleh, 2008; Salum et al., 2019; Savolainen and Mannering, 2007). In the past decade, several prediction models have been developed to predict crash frequency (Yannis et al., 2007; Abdel-Aty et al., 2013; Dong et al., 2014; Cai et al., 2017; Zeng et al., 2017), crash risk (Lin et al., 2015; Shaw et al., 2013), and injury severity (Duddu et al., 2018; Fountas and Anastasopoulos, 2018; Mussone et al., 2017), however, little attention has been paid to predicting at-fault crash driver frequency. Traffic violations are the major causes of traffic crashes in Taiwan, and the same phenomenon is observed in other countries as well (Fararouei et al., 2016; Penmetsa and Pulugurtha, 2017). Thus, it is believed that traffic crashes and injuries will decrease if traffic violations reduce (Factor, 2014). Moreover, by predicting at-fault crash driver frequency in the near future, traffic safety measures (e.g., reinforced traffic enforcement and public safety broadcasting) can be applied more precisely to prevent traffic crashes among risk groups. Thus, traffic violation data can be used as ideal predictors for predicting at-fault crash driver frequency, while an insight into the effect of each traffic enforcement activity on traffic safety can

* Corresponding author.

E-mail address: d04521010@ntu.edu.tw (Y.-W. Wu).

help contribute to the development of traffic safety measures.

In summary, the objectives of this study are as follows: (1) to develop effective prediction approaches for predicting at-fault crash driver frequency in the mid-term period using only traffic enforcement predictors at the city level, and (2) to investigate the contribution of each traffic enforcement activity on traffic safety in scenarios where multicollinearity exists among traffic enforcement activities.

This study demonstrates the feasibility of using traffic violation data with multicollinearity as predictors; by using a state-of-the-art approach to investigate the contribution of each traffic enforcement activity, effective traffic safety measures can be developed to reduce traffic crashes.

The rest of this paper is organized as follows. Section 2 presents a review of previous researches that focused on traffic enforcement and crash frequency prediction models. Section 3 introduces the collection and preparation of traffic crash and traffic violation data. The econometric, machine learning, and deep learning approaches implemented are described in Section 4. The results of the analysis and model comparisons are discussed in Section 5. Finally, Section 6 presents the conclusions and future scope.

2. Literature review

2.1. Traffic enforcement and traffic safety

Traffic safety can be improved by implementing appropriate traffic enforcement schemes to deter road users from committing traffic regulation offenses (European Transport Safety Council, 1999). Several studies have shown strong correlations between traffic enforcement and both driver behavior and attitude (Factor, 2014; Shaaban, 2017; Yannis et al., 2008). Meanwhile, by applying appropriate police enforcement and automated systems (e.g., cameras and point-to-point speed enforcement systems), a reduction in vehicle speeding (Ryeng, 2012; Montella et al., 2015) and the citywide rate of fatal red-light running crashes (Hu et al., 2011) and an increase in belt usage (Williams and Wells, 2004) were observed. In Hong Kong, the frequency of red-light violations significantly decreased after implementing the new penalty system (Sze et al., 2011); further, the ban on handheld cellphone use while driving in California played an important role in improving traffic safety (Liu et al., 2019). In addition, traffic violations and aggressive driver behaviors are more frequent in areas lacking traffic enforcement (Porter et al., 2013; Stanojević et al., 2013); a decrease in aggressive driving behaviors under the implementation of high visibility enforcement programs was observed by investigating the second strategic highway research program (SHRP2) naturalistic driving study (NDS) data (Pantangi et al., 2020).

The effects of traffic enforcement on controlling high-risk and dangerous driver behaviors can help decrease crashes due to drivers complying with traffic regulations. In a cross-sectional study covering ten countries, Urie et al. (2016) found that strong and efficient traffic enforcement can cause lower crash fatality rates. Blais and Gagné (2010) indicated that reducing traffic citations would increase crash injuries, while Kwon et al. (2012) found that traffic violation pardons would raise traffic accidents and the economic costs of traffic accidents in consecutive years after pardons were granted in Korea. Ali et al. (2019) inferred that poor traffic enforcement in upper middle-income countries (UMICs) is a possible factor behind the increase in road traffic fatalities. Thus, based on the successful investigation of the association of traffic enforcement and traffic safety, injudicious actions can be considered the ideal predictor of crash prediction models. However, monitoring these behaviors comprehensively and continuously is impractical because of finite resources. To address this problem, traffic violation data (e.g., citations issued by police officers) can be introduced into crash prediction models as surrogates for injudicious actions and traffic enforcement activities. However, only few studies have utilized traffic violation data as a predictor for crash prediction models (Yannis et al., 2007; Coruh

et al., 2015). Thus, the potential of using traffic violation data for traffic safety analysis deserves further attention.

2.2. Crash frequency prediction model

2.2.1. Time scale and predictors for prediction model

During the past decades, most traffic safety studies focused on crash frequency prediction models on long-term (e.g., one month/year or several months/years) (Coruh et al., 2015; Dong et al., 2014; Zhang et al., 2012, 2019), short-term (e.g., hours/days/one week) (Bao et al., 2019), and real-time scales (e.g., several minutes) (Basso et al., 2018; Li et al., 2020). Thus, there is a lack of traffic safety analyses for mid-term periods (e.g., several weeks). To achieve long-term traffic safety targets, traffic safety agencies need to set mid-term targets and frequently review safety performance. A mid-term crash frequency prediction model can enable traffic safety agencies to develop appropriate traffic safety measures for the following mid-term period. Common predictors for crash prediction models include traffic characteristics (Geedipally et al., 2010; Aguero-Valverde and Jovanis, 2006; Abdel-Aty et al., 2013; Bhowmik et al., 2019), roadway design and configuration characteristics (Geedipally et al., 2010; Abdel-Aty et al., 2013; Bhowmik et al., 2019), socio-economic characteristics (Aguero-Valverde and Jovanis, 2006; Bhowmik et al., 2019; Cai et al., 2017), and weather and environmental characteristics (Aguero-Valverde and Jovanis, 2006; Bhowmik et al., 2019; Zhao et al., 2019a, 2019b). The data for these common predictors may not always be available because of the lack of collectors, absence of related information systems or linked datasets for various data sources (Gomes et al., 2019), and high survey costs. In contrast, it may be easier to collect traffic violation data as traffic citations are recorded accurately by traffic police or other agencies. For countries or areas in which only traffic violation data are available, crash prediction approaches that can use only traffic violation data are required for traffic safety analysis. Moreover, predictors used on a long-term (e.g., roadway design and socio-economics characteristics), short-term (e.g., traffic trips and weather), and real-time scales (e.g., traffic speed and volume) may be inadequate for mid-term scale prediction. Roadway design and socio-economics characteristics are very likely to be relatively static compared with variations in crash frequency over a couple of weeks. Moreover, the association between traffic trips, weather, and traffic speed in previous mid-time periods with crash frequency in the future is not justified. Thus, these common predictors may be unable to capture the trend or variation in crash frequency in the following mid-time period. To address this issue, insight into the time halo effect of traffic enforcement can facilitate the selection of appropriate predictors for mid-term crash frequency prediction models. The time halo effect is the length of time that the effects of traffic enforcement on behavior of drivers continue after the enforcement personnel are removed (Elliott and Broughton, 2005). Studies have reported that the time halo effect can last for 1 h to 8 weeks (Vaa, 1997; Elliott and Broughton, 2005). This suggests that traffic enforcement may have effects on the behaviors and attitudes of road drivers, and to some extent on the crash frequency in the following mid-term period. Thus, it is expected that prediction models with traffic enforcement predictors can perform mid-term crash prediction tasks.

2.2.2. Predictive approaches

Statistical approaches such as the Poisson-family model (Caliendo et al., 2007; Dong et al., 2014; Qin et al., 2004), negative binomial-family model (Coruh et al., 2015; Gomes et al., 2017; Lee et al., 2017; Lord and Geedipally, 2011), Bayesian-based model (Haleem et al., 2010; Lee et al., 2015; Xie et al., 2018; Zeng et al., 2017), and time series model (Abdel-Aty and Abdelwahab, 2004; Antoniou and Yannis, 2013; Quddus, 2008; Vanlaar et al., 2014) have been utilized extensively for crash frequency analysis and prediction for decades. In recent years, crash prediction studies focused on machine learning and deep learning approaches, including boosting (Ahmed and Abdel-Aty, 2013; Zhang

et al., 2019), random forest (RF) (Lin et al., 2015; Siddiqui et al., 2012; Theofilato, 2017), support vector machine (SVM) (Basso et al., 2018; Elassad et al., 2020; Li et al., 2008), backpropagation neural network (BPNN) (Kibar et al., 2017; Wang et al., 2019), convolutional neural network (CNN) (Cai et al., 2019), long short-term memory (LSTM) neural network (Ren et al., 2018), ensemble method (Wu et al., 2019), and fusion deep learning methods (Bao et al., 2019; Dong et al., 2018; Gu et al., 2019). Owing to their capabilities such as mining high-dimensional information and extracting hierarchical feature representation in the dataset, these advanced approaches can provide desirable predictive performances. Moreover, the potential endogeneity (Chen and Tarko, 2012; Lord and Mannering, 2010), multicollinearity (Huang et al., 2016; Siddiqui et al., 2012) and complex interactions among variables in statistical models can be addressed (Abdel-Aty et al., 2005; Ülengin et al., 2007). A statistical crash prediction model with only traffic enforcement predictors would certainly suffer from significant endogeneity and multicollinearity. In this context, crash prediction tasks can be performed using machine learning and deep learning approaches to achieve desirable predictive results. Moreover, it is unnecessary to exclude any variable from the model, which means all potential predictors can be investigated in detail.

However, while data-driven approaches (e.g., machine learning and neural network-based approaches) may effectively capture associations between one variable and another, these approaches often earn a “black-box” designation because of the difficulty in unraveling how specific elements might influence predictions (Mannering et al., 2020). Hence, analytical approaches such as a partial dependent (PD) plot (Friedman, 2001) and accumulated local effects (ALE) plot (Apley and Zhu, 2019) have been developed to provide better insight into how a predictor influences the outcome when implementing machine learning or neural network-based approaches. Thus, crash prediction models using data-driven approaches can be more clearly interpreted using PD plot or ALE plot.

In terms of the geographical level for traffic safety analysis, a macro-level safety analysis suggests long-term policy-based countermeasures such as enactments of traffic rules and police enforcements (Lee et al., 2015). Several studies have focused on crash frequency prediction at the macro level (Aguero-Valverde and Jovanis, 2006; Abdel-Aty et al., 2013; Bhowmik et al., 2019; Cai et al., 2017; Lee et al., 2015; Rahman et al., 2019; Yannis et al., 2007). However, only a few studies consider the distinction between under-prediction and over-prediction that can lead to inadequate countermeasures or policies. In this study, we implemented at-fault crash driver frequency prediction tasks with city-level crash risk prediction tasks, which refer to whether the at-fault crash driver frequency in the consecutive time period is greater than that in the previous time period. Thus, performance metrics commonly used for microlevel risk analysis such as the area under the receiver-operating characteristic curve (AUC) and recall rate (sensitivity) (Basso et al., 2018; Peng et al., 2020; Yu et al., 2020) are utilized in this study.

The objective of this study is to develop effective approaches for predicting at-fault crash driver frequency in the near-mid-term period using only traffic enforcement predictors at the city level. Thus, a fusion deep learning approach combining two deep neural networks was developed. One econometric approach, two machine learning approaches, and the other deep learning approach were employed to compare the predictive performance of the proposed fusion deep learning approach. Crash and traffic violation data were collected from Taoyuan City, which is one of the six metropolitan cities in Taiwan. We further investigated the contribution of each traffic enforcement activity on at-fault crash driver frequency using the ALE plot, which is based on the built machine learning models. Results from this study can contribute to the development of mid-term traffic safety measures to which insufficient attention has been paid, especially for areas where only traffic violation data is available.

3. Data preparation

The at-fault crash driver frequency and traffic citation data were considered response variable and predictors, respectively. The major source of data for this study is the crash and traffic citation database maintained by the Traffic Police Corps, Taoyuan Police Department (TPC-TYPD). Crashes and traffic citations issued from 2010 to 2018 in Taoyuan City were collected. Notably, only fatal and injury crashes are reported by traffic police officers in Taiwan. Thus, property damage-only crashes are under-reported owing to the absence of regulatory requirements. Hence, drivers involved in property damage-only crashes are not included in the collected dataset. We investigated only at-fault crash drivers, including vehicle drivers, motorcyclists, and bicyclists involved in fatal or injury crashes. During the nine-year period between 2010 and 2018, a total of 177,484 at-fault crash drivers were involved in fatal or injury crashes.

In Taiwan, 67 categories of traffic crash causes can be reported by traffic police officers. These categories include tens of traffic violations, inattention or distraction, road condition, and vehicle malfunction. Meanwhile, the five major crash causes are violations such as failure to give way, illegal turn, illegal lane change or usage, speeding, and signal violation. As listed in Table 1, more than 70 % of at-fault crash drivers were guilty of one of these five major traffic crash causes. To help traffic safety agencies utilize limited resources optimally to reduce fatalities and injuries, this study focused on at-fault crash drivers with these five major crash causes. Thus, we used six response variables: the at-fault crash driver counts for the five major crash causes and the total of crash-involved road user counts.

Furthermore, traffic violation data were considered as predictors. The “Road Traffic Management and Penalty Act”—the primary traffic law in Taiwan—comprises over 200 traffic violations. In Taiwan, traffic citations are issued by physical policing or direct accusation. Physical policing refers to the use of police units for stationary or mobile enforcement on the road (Elliott and Broughton, 2005). Direct accusation in this study refers to the issuing of traffic citations for violation behaviors without physical policing. For instance, traffic citation issued by automated enforcement systems (e.g., cameras) will be mailed to the vehicle owner or registered vehicle driver within days. For the simplicity of the prediction models, all traffic violations are grouped into 27 categories in this study based on the properties and features of each traffic violation. In the collected violation dataset, 27 categories of physical policing violations and 24 categories of directly accused violations were obtained as traffic enforcement predictors.

The description of response variables and predictors are presented in Table 2. The time required for presenting the effect of traffic enforcement on the attitude and behavior of the driver is determined based on the assumption that physical policing has an immediate effect, whereas the effect of directed accusation can only be presented when the violator is noticed by the mail of traffic citation. A 14-day lag of the effect of directly accused citations was selected based on the suggestion from TPC-TYPD officers considering the operational procedure and delivery time of directly accused citations.

The data aggregation and manipulation process is illustrated in Fig. 1. The at-fault crash driver frequency was aggregated at three

Table 1
At-fault crash driver frequency by committed violation.

Violation	At-fault crash driver frequency	Percentage
<i>The five major traffic crash causes</i>		
Failure to give way	45,817	25.81%
Illegal turn	39,587	22.30%
Illegal lane changing or usage	14,420	8.12%
Speeding	13,792	7.77%
Signal violation	10,933	6.16%
<i>Other causes</i>		
Other causes	52,935	29.83%

Table 2

Description of response variables and predictors.

Variables	Description
<i>Response variables</i>	
Total road users in crashes	Number of road users involved in crashes
Not Give way drivers	Number of drivers who contributed to crashes because of failing to give way
Illegal turn drivers	Number of drivers who contributed to crashes because of illegal turn
Illegal lane usage drivers	Number of drivers who contributed to crashes because of illegal lane usage
Speeding drivers	Number of drivers who contributed to crashes because of speeding
Signal violation drivers	Number of drivers who contributed to crashes because of disobeying traffic signal
<i>Predictors</i>	
Vehicle plate	Failure to correctly display vehicle plate
Fatigue driving*	Fatigue driving
DUI	Driving under influence
Speeding	Exceeding speed limit
Vehicle light	Failure to correctly use headlight or turn signal
Dangerous driving	Driving in a dangerous or an aggressive way
Offence on pedestrian	Not giving priority to pedestrians
Failure to reduce speed	Failure to reduce speed before entering a junction or at specific segments
Wrong way	Driving the wrong way down a one-way street or driving against traffic flow
Illegal lane driving	Illegally entering an exclusive lane or disobeying lane markings
Illegal overtaking	Overtaking at segments where overtaking is prohibited
Vehicle license	Driving an unlicensed vehicle
Illegal turn	Turning in an improper way or without the permission of traffic signal/sign
Red light running	Entering a junction after the signal light has turned red
Illegal parking	Parking in places where stopping/parking is prohibited
Too close	Following too closely
Marking/sign violation	Disobeying traffic markings or signs
Refuse inspection	Refusing to be inspected by police officers
Bicyclist violation*	Violation committed by bicyclists
Pedestrian violation*	Violation committed by pedestrians
Vehicle modification	Modifying vehicle components without permission
Safety lesson	Failure to attend traffic safety lesson without a legitimate reason
Driver license	Driving a vehicle without a qualified license
Overloading	Carrying goods that exceed the prescribed specifications or weight limit
Safety appliance	Failure to use helmet, seatbelt, or other safety apparatus
Violations on freeway	Disobeying special regulations for freeways
Other violations	Other violations on local roads/streets

* Only cited by physical policing.

different temporal levels: weekly (7 days), semimonthly (15 days), and monthly (30 days). The time periods of the three temporal levels were defined as mid-term periods. For weekly aggregation, the daily at-fault crash driver frequencies in the following 7-day period were aggregated for the first day in the 7-day period. For example, the daily at-fault crash driver frequencies from July 1 to July 7 were aggregated for the weekly frequency on July 1. The weekly at-fault crash driver frequency on July 2 was obtained by similarly aggregating data from July 2 to July 8. This method for aggregating data was further implemented to each date during the study period for all temporal levels.

In this study, traffic violation data in a recent period of 30 days were utilized as model inputs. This 30-day period was selected based on a conservative speculation of the time halo effect of traffic enforcement in Taiwan. Based on different prediction models, there are two approaches to aggregate and manipulate traffic violation data. First, for recurrent neural network (RNN) based deep learning models, traffic violation data were collected on a daily basis. The lookback parameter for RNN-based deep learning models was determined as 30, which implies the daily

traffic violation data in the previous 30-day period were used as model inputs. For example, the daily 51 categories of traffic violation data from June 1 to June 30 are used as predictors for at-fault crash driver frequency on July 1. Second, for tree-based machine learning models, 51 categories of traffic violation data in the previous 30-day period are aggregated on a monthly scale for the last day to prevent model complexity and ensure that the models are easy to interpret. For example, the daily traffic violation data from June 1 to June 30 are aggregated for the monthly traffic violation data on June 30. Then, the monthly 51 categories of traffic violation data are used as predictors for the at-fault crash driver frequency on July 1.

The numbers of observations for the three temporal levels were the same: 3,287 (days from 2010 to 2018). The descriptive statistics of response variables at four different temporal levels are presented in Table 3. The descriptive statistics of predictors at the daily and monthly levels are presented in Tables 4 and 5.

Multicollinearity between traffic enforcement predictors at the monthly level is illustrated in Fig. 2. The Pearson correlation between each pair of predictors is depicted in different colors. Significant widespread multicollinearity is observed among predictors; this may be a problem in statistical modeling. If some traffic enforcement predictors were dropped to eliminate multicollinearity in the prediction model, insight cannot be obtained on the corresponding traffic enforcement activities of the dropped predictors. Therefore, as discussed in Section 2.2.2, machine learning and deep learning approaches are suitable in this scenario. A total of 51 traffic enforcement predictors were considered in the prediction models for a detailed investigation of all traffic enforcement predictors.

4. Methodology

In this study, we developed a fusion deep learning approach that can extract high-dimensional features and spatial and temporal dependencies in traffic violation and crash datasets; this approach is aimed at predicting at-fault crash driver frequency. Further, we compared the proposed fusion deep learning approach with several approaches. The selected comparison approaches include a representative econometric approach (autoregressive integrated moving average (ARIMA)), two popular machine learning approaches (random forest (RF) and the gradient boosting regression tree (GBRT)), and one state-of-the-art deep learning approach (gated recurrent units (GRU)). This section presents the prediction approaches and the evaluation metrics employed.

4.1. Econometric approach

Time series approaches have been utilized commonly in traffic safety research to predict traffic fatalities and crash count (Abdel-Aty and Abdelwahab, 2004; Antoniou and Yannis, 2013; Quddus, 2008; Vanlaar et al., 2014). Among the time series approaches, autoregressive integrated moving average (ARIMA) is a parametric time series method that is widely employed in prediction tasks. A univariate ARIMA model describes the behavior of a variable in terms of its past value (Washington et al., 2020). The standard notation for an ARIMA model is ARIMA (p, d, q), where p refers to the auto regressive order; d , the differencing order; and q , the moving average order. Despite the recent popularity of machine learning and deep learning approaches in the transportation field, ARIMA has been employed frequently as a representative comparison approach to other state-of-the-art approaches in transportation research such as traffic crash risk (Bao et al., 2019) and traffic speed (Gu et al., 2019; Ma et al., 2015). Because widespread multicollinearity is observed among predictors, this study uses ARIMA rather than the statistical model (e.g., Poisson-based model or negative binomial-based model) as the benchmark model for model comparison.

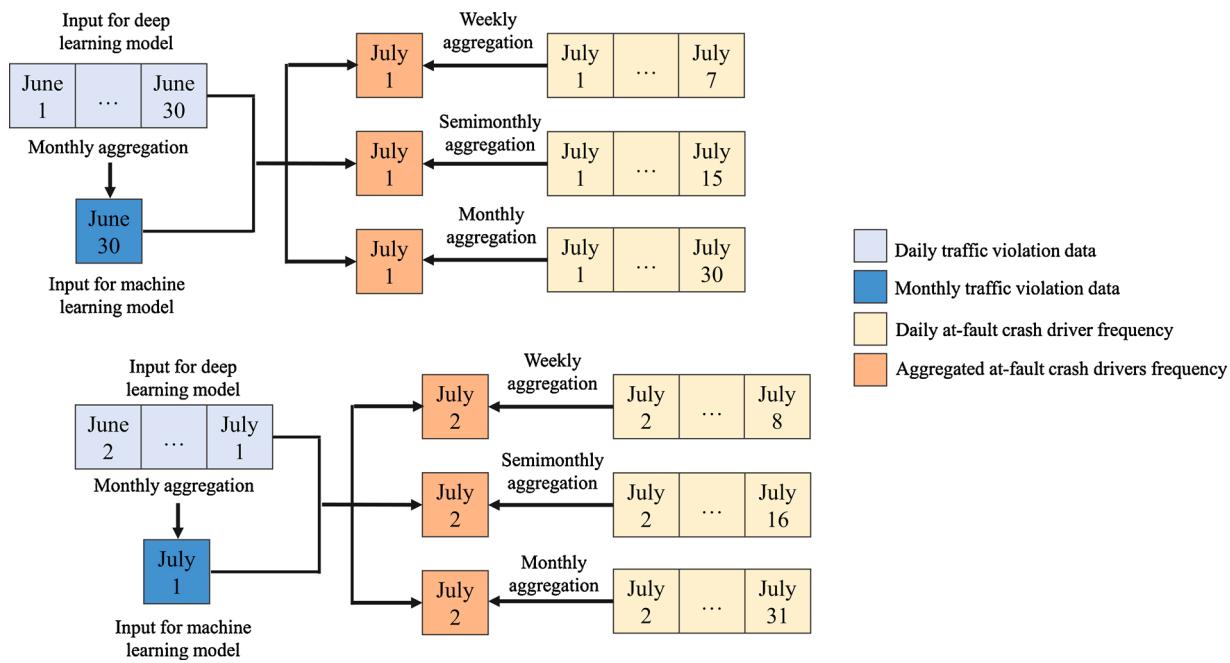


Fig. 1. Illustration of data aggregation and manipulation process.

Table 3

Descriptive statistics of response variables.

Variables	Daily level				Weekly level			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Not Give way drivers	0	42	8.90	5.13	28	146	84.43	19.71
Illegal turn drivers	0	1	0.00	0.02	18	134	71.42	14.95
Illegal lane usage drivers	4	157	34.39	16.85	4	49	25.43	7.59
Speeding drivers	0	64	1.84	4.66	3	81	24.86	11.00
Signal violation drivers	0	4	0.06	0.29	4	55	19.29	6.59
Total road users in crashes	0	43	0.91	3.70	307	1,816	994.76	250.19

Variables	Semimonthly level				Monthly level			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Not Give way drivers	84	267	168.93	35.92	209	545	362.45	72.29
Illegal turn drivers	76	242	142.91	25.71	189	468	306.55	48.78
Illegal lane usage drivers	17	90	50.89	12.84	49	183	109.25	24.31
Speeding drivers	16	150	49.78	20.63	44	272	106.91	42.47
Signal violation drivers	14	86	38.58	11.05	37	147	82.64	20.78
Total road users in crashes	834	3,440	1,991.14	484.64	2,052	7,061	4,275.52	1,014.72

4.2. Machine learning approaches

Ensemble tree-based machine learning methods including random forest (RF) and gradient boosting regression tree (GBRT) comprise weak learners to generate more reliable and better results compared to a single decision tree. The RF and GBRT are two of the most representative ensemble tree-based machine learning approaches that are popular in the transportation field. Moreover, approaches such as accumulated local effects (ALE) plot, can describe on average the influence of predictors on the response variable. In this study, we used RF and GBRT for prediction tasks and utilized the ALE plot to investigate the effect of traffic enforcement activities on predictive at-fault crash driver frequency.

4.2.1. Random forest

The RF is an ensemble tree-based learning algorithm that has been used widely since it was proposed by Breiman (2001). In an ensemble tree-based learning algorithm, decision trees are built on a host of bootstrapped datasets obtained from the original training set. The

predictions of each tree are gathered for a majority vote for the classification outcome. In terms of the regression model, a quantitative outcome can be obtained by averaging the predictions of each tree. However, an ensemble learning algorithm may be ineffective if all predictors are considered at each split when building trees (bagging method). Strong predictors dominate top splits and cause similar trees. The RF addresses this problem and provides an outstanding predictive power by decorrelating the trees. Only a random subset of the full set of predictors is selected as split candidates in the RF. With a higher chance of being selected in splits for normal predictors, the decorrelation technique can address the problem of highly correlated trees in the bagging method. Consequently, averaging the prediction of trees with a high variance but low bias can yield a more reliable and better predictive performance (James et al., 2013).

There are no criteria for determining the subset size m of the predictor. As a rule of thumb, $m \approx \sqrt{p}$ is suggested in the literature (James et al., 2013), while $m = p/3$ is the default setting in some software packages (e.g., R package "randomForest"), where p is the number of total predictors. In this present study, 7 and 17 random predictors were

Table 4

Descriptive statistics of predictors at the daily level.

Predictors	Physical policing				Directed accusation			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Vehicle plate	0	42	8.90	5.13	0	55	1.82	3.80
Fatigue driving	0	1	0.00	0.02	–	–	–	–
DUI	4	157	34.39	16.85	0	10	0.02	0.24
Speeding	0	64	1.84	4.66	0	9,358	467.73	742.02
Vehicle light	0	4	0.06	0.29	0	57	0.47	1.76
Dangerous driving	0	43	0.91	3.70	0	73	2.50	5.30
Offence on pedestrian	0	19	0.51	1.34	0	56	0.46	2.28
Failure to reduce speed	0	5	0.22	0.56	0	1	0.00	0.06
Wrong way	0	104	19.63	14.01	0	357	10.37	22.43
Illegal lane driving	0	102	9.64	7.80	0	415	19.63	34.64
Illegal overtaking	0	5	0.22	0.50	0	23	0.70	1.77
Vehicle license	0	27	1.45	3.13	0	3	0.02	0.15
Illegal turn	0	854	195.89	118.84	0	1,506	32.72	63.67
Red light running	4	1,758	143.68	105.93	0	4,727	134.22	215.98
Illegal parking	0	86	10.62	6.71	0	10,700	664.03	734.98
Too close	0	6	0.28	0.60	0	11	0.03	0.32
Sign/marking violation	0	171	19.67	14.76	0	533	30.79	49.68
Refuse inspection	0	10	1.42	1.40	0	3	0.04	0.21
Bicyclist violation	0	20	0.41	0.96	–	–	–	–
Pedestrian violation	0	46	1.48	3.07	–	–	–	–
Vehicle modification	0	78	2.44	4.53	0	7	0.19	0.61
Safety lesson	0	102	13.58	11.71	0	7	0.03	0.25
Driver license	2	143	38.28	17.52	0	19	0.04	0.46
Overloading	0	63	10.61	7.40	0	7	0.06	0.30
Safety appliance	0	184	9.71	12.09	0	114	3.16	5.69
Violations on freeway	0	5	0.04	0.26	0	262	8.67	21.35
Other violations	0	58	10.38	7.86	0	7	0.04	0.29

Table 5

Descriptive statistics of predictors at the monthly level.

Predictors	Physical policing				Directed accusation			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Vehicle plate	106	617	266.90	101.21	3	218	53.81	50.29
Fatigue driving	0	1	0.02	0.13	–	–	–	–
DUI	494	2,045	1,033.01	342.85	0	24	0.56	2.09
Speeding	0	489	56.12	89.09	717	36,291	13,984.35	4,577.59
Vehicle light	0	12	1.92	2.25	0	87	13.84	17.29
Dangerous driving	0	385	27.30	59.47	9	317	74.74	44.36
Offence on pedestrian	0	85	15.37	15.16	0	163	13.54	24.72
Failure to reduce speed	0	35	6.39	8.95	0	4	0.12	0.41
Wrong way	108	1,232	588.00	234.26	4	1,441	307.58	286.89
Illegal lane driving	34	856	288.37	128.49	22	1,954	586.21	402.60
Illegal overtaking	0	27	6.49	5.78	0	101	20.60	23.24
Vehicle license	0	281	44.42	68.40	0	6	0.59	0.96
Illegal turn	1,231	13,860	5,871.48	2,182.67	4	5,004	968.59	889.67
Red light running	1,711	14,903	4,318.08	1,841.41	590	11,040	4,012.02	1,907.28
Illegal parking	125	720	317.92	111.35	4,949	47,093	19,830.87	6,952.83
Too close	0	42	8.29	7.39	0	16	0.88	2.20
Sign/marking violation	133	1,499	590.76	285.00	97	2,504	928.01	423.86
Refuse inspection	13	86	42.41	16.89	0	7	1.15	1.29
Bicyclist violation	0	75	12.02	13.70	–	–	–	–
Pedestrian violation	1	344	44.34	49.13	–	–	–	–
Vehicle modification	11	1,027	73.62	91.02	0	28	5.63	7.05
Safety lesson	96	1,003	409.39	196.22	0	15	0.97	1.74
Driver license	504	2,290	1,148.51	365.80	0	62	1.21	5.14
Overloading	49	627	319.11	126.65	0	9	1.89	1.98
Safety appliance	26	1,202	294.17	265.24	10	398	94.48	68.20
Violations on freeway	0	16	1.29	2.01	2	826	261.89	143.72
Other violations	89	984	311.03	185.11	0	11	1.17	1.76

selected as subsets of the 51 predictors.

4.2.2. Gradient boosting regression tree

Unlike RF or the bagging method that average the predictions of multiple trees as the model outcome, trees in GBRT are built sequentially, and the model outcome is produced by the combination of predictions from sequential trees. For the regression scheme, trees are built based on residuals from previously built trees. These trees are relatively

small, which is done to maintain few terminal nodes in a tree. Small and sequential trees provide the capability of overcoming potential overfitting by learning slowly (James et al., 2013). An additive model was developed to combine a series of trees; it is expressed as (Hastie et al., 2009):

$$f(x) = \sum_m f_m(x) = \sum_m \beta_m b(x; \gamma_m) \quad (1)$$

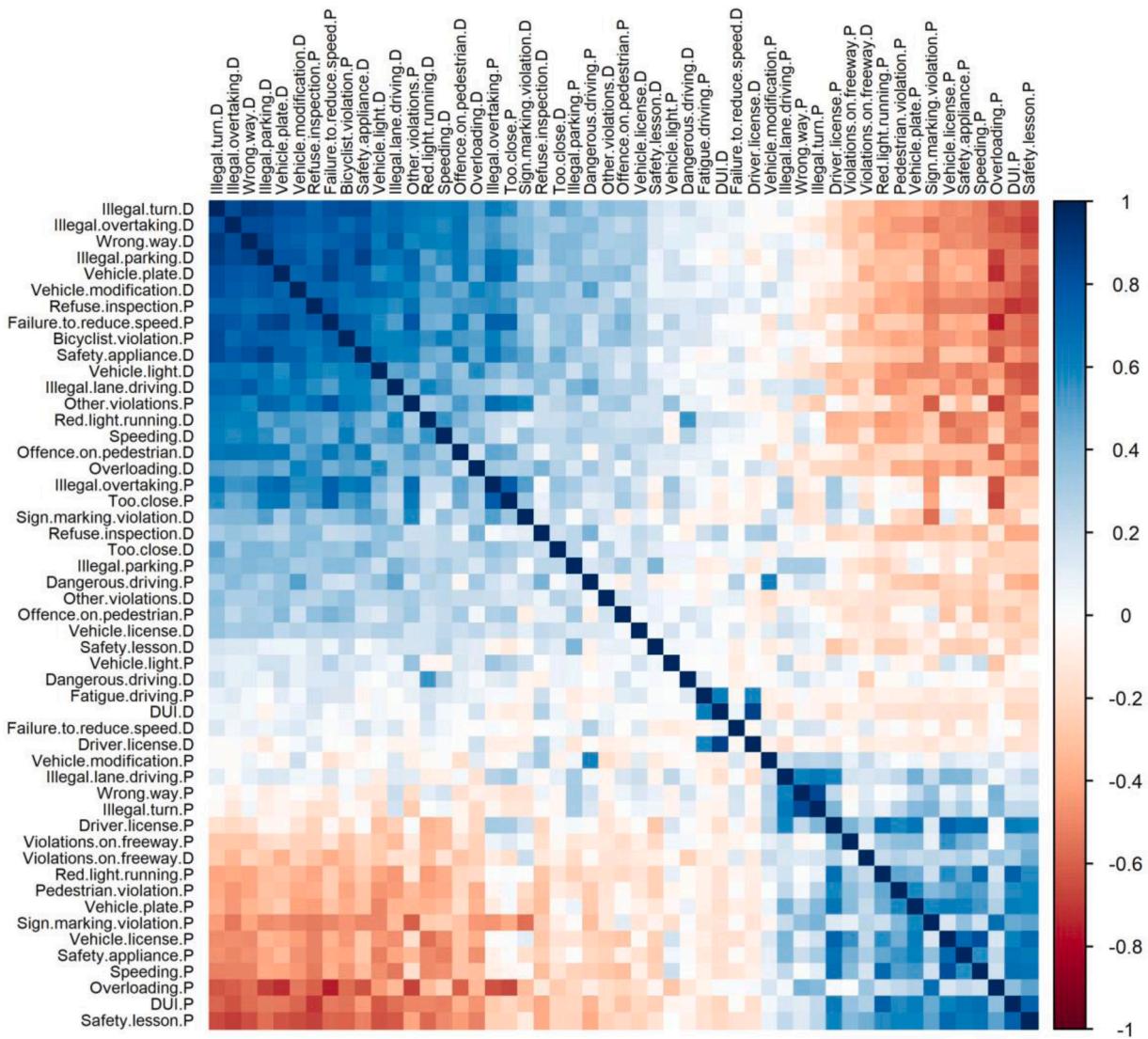


Fig. 2. Correlations among predictors. Blue grids represent strong positive correlation and red grids represent strong negative correlation. “D” and “P” refer to “Directed accusation” and “Physical policing,” respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

where $f(x)$ is a function that approximates response variable y ; x refers to a set of predictors; the function $b(x; \gamma_m)$ refers to individual trees with a set of split variables γ_m in the m -th iteration; and the parameter β_m is the weight estimated by the gradient boosting algorithm to determine how predictions from the individual trees are combined. A gradient boosting algorithm was proposed to estimate β_m and γ_m (Friedman (2001); De'ath (2007)). As mentioned above, the GBRT learns slowly to address overfitting. Approaches that learn slowly tend to perform effectively but require more trees (De'ath, 2007; James et al., 2013). A shrinkage parameter η , which can be viewed as the learning rate, can be introduced into the model to control how trees in the GBRT fit the residuals. A small value of η is preferable but should be determined based on the problem (James et al., 2013). The GBRT models with various shrinkage values and different depth of tree were used for the subsequent analysis in this study.

4.2.3. Accumulated local effects plot

To interpret ensemble tree-based learning methods, the average decrease in the residual sum of squares (RSS) contributed by a given predictor for all trees is calculated as a measure of relative importance of the given predictor. The higher the relative importance value, the more

the predictor contributes to the model outcome. However, the relative importance of predictors is insufficient to provide insight into how a predictor influences a model outcome. Thus, Friedman (2001) proposed a partial dependence (PD) plot to investigate the marginal effect of a feature on the response variable. However, independence is assumed for predictors in the PD plot. When this assumption of independence is violated, the ALE plot (Apley and Zhu, 2019) is an alternative to investigate the marginal effect of a predictor on the response variable.

Let X_j be the predictor of interest and X_c be the set of other predictors; let a function $f(\cdot)$ represent the fitted model that predicts the response variable. The local effect of x_j on a differentiable function $f(\cdot)$ is defined as:

$$f^j(x_j, x_c) \equiv \frac{\partial f(x_j, x_c)}{\partial x_j} \quad (2)$$

ALE calculates the local effect of a given predictor X_j at $(x_j = z_j, x_c)$ and averages this local effect across all values of x_c grids with weight $p_{c|j}(x_c | z_j)$. Finally, it accumulates the averaged local effect over all values of z_j up to x_j . The ALE main effect can be expressed as:

$$\begin{aligned} g_j ALE(x_j) &= \int_{X_{\min,j}}^{X_j} E[f^j(X_j, X_C) | X_j = z_j] dz_j \\ &= \int_{X_{\min,j}}^{X_j} \int p_{c|j}(X_C | z_j) f^j(z_j, X_C) dx_c dz_j \end{aligned} \quad (3)$$

In Eq. (3), $g_j ALE(x_j)$ denotes the uncentered ALE main effect of predictor X_j . The centered ALE main effect of predictor X_j , denoted as $f_j ALE(x_j)$, differs from $g_j ALE(x_j)$ because it has a mean of zero with respect to the marginal distribution of X_j . Apley and Zhu (2019) defined $f_j ALE(x_j)$ as (Please refer to Apley and Zhu (2019) for detailed information):

$$\begin{aligned} f_j ALE(x_j) &= g_j ALE(x_j) - E[g_j ALE(X_j)] \\ &= g_j ALE(x_j) - \int p_j(z_j) g_j ALE(z_j) dz_j \end{aligned} \quad (4)$$

Compared to the PD plot, the ALE plot addresses the problem of correlation between predictors by averaging the predictions on each grid value of the predictor of interest instead of assuming the marginal distribution at each grid value (Molnar, 2019). The ALE plot has considerable potential for analyzing the contribution of a predictor on the response variable in a machine learning model because of the following two advantages: (1) The ALE plot can work with correlated features, and (2) it shows better computational efficiency than the PD plot. In this study, we utilized the ALE plot to investigate how each predictor of the violation affects the prediction of at-fault crash drivers with the RF and GBRT models. The results provide insight into the contribution of each traffic enforcement activity on traffic safety. To the best of our knowledge, this is the first time that the ALE plot is employed in traffic safety analysis.

However, there are two reasons that the ALE plot is not suitable of RNN-based (e.g., RNN, LSTM, and GRU) and CNN models. First, for RNN-based and CNN models, each batch input in one training iteration contains several values of the predictor of interest at different timesteps, while only one value of the predictor is served as input at a time for machine learning and normal artificial neural networks (ANN) models. Second, the structure of RNN-based models and CNN models attempt to extract complex temporal dependency and local representation in the dataset while machine learning and normal ANN models consider all input predictors as independent and equivalent. Thus, this study only applies the ALE plot with machine learning models.

4.3. Deep learning approaches

Deep learning approaches, which are based on deep neural networks, have recently attracted considerable interest in almost every field. Several recent studies have applied deep learning approaches such as BPNN, CNN, and LSTM in traffic safety analysis for crash frequency prediction (Bao et al., 2019; Cai et al., 2019; Dong et al., 2018), crash risk prediction (Li et al., 2020; Ren et al., 2018), and other safety-related prediction tasks (Wang et al., 2019). The successful application of a single deep neural network has been reported in the literature. Some studies combined multiple deep neural networks to develop fusion deep learning architectures that can perform various high-dimensional feature extraction and are compatible with multiple data sources (Bao et al., 2019; Li et al., 2020).

To predict at-fault crash driver frequency in the following mid-term period, GRU, which is a variant of LSTM, was selected in this study to capture temporal dependency in the dataset. In addition to the GRU, a fusion model comprising a CNN and GRU was developed to extract both spatial and temporal features from the collected dataset.

4.3.1. Gated recurrent units

The GRU, proposed by Cho et al. (2014), is a variant of both RNN and LSTM. Similar to LSTM, the GRU can address long temporal dependency issues caused by the vanishing and exploding gradient problems in RNN.

Moreover, the GRU can achieve equal or better performances than LSTM with a lower computation cost. The GRU revises hidden units in the LSTM by reducing the gates from four to two. The gating unit, which comprises one update gate and one reset gate, can modulate the information flow inside the GRU without a separate memory cell. The structure of a single cell of the GRU is shown in Fig. 3. The computation of information flow in the GRU can be expressed as:

$$r_t = \sigma(U_r h_{t-1} + W_r x_t) \quad (5)$$

$$\tilde{h}_t = \tanh(U(r_t \otimes h_{t-1}) + Wx_t) \quad (6)$$

$$z_t = \sigma(U_z h_{t-1} + W_z x_t) \quad (7)$$

$$h_t = (1 - z_t)h_{t-1} + z_t \tilde{h}_t \quad (8)$$

In Eq. (5), r_t denotes the reset gate; x_t denotes the input vectors; σ is a logistic sigmoid function; h_{t-1} denotes the activation from a previous time step; U_r and W_r are weight matrices for activation and input data, respectively, in the reset gate. In Eq. (6), \tilde{h}_t is the candidate activation; \otimes denotes an element-wise multiplication; U and W are weight matrices for activation and input data, respectively. In Eq. (7), z_t is the update gate, which decides the amount by which the unit updates its activation or content. Similarly, U_z and W_z are weight matrices for activation and input data, respectively, in the update gate. Finally, in Eq. (8), the activation of the GRU at time t is a linear interpolation between the previous activation h_{t-1} and candidate activation \tilde{h}_t . The logistic sigmoid function σ and tanh function are calculated as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (10)$$

4.3.2. Fusion deep learning model

Recently, models that combine two or more deep learning approaches have attracted widespread interest. Given the various strengths of each deep neural network, a fusion deep learning model extracts multiple features in a dataset for desirable predictive performances. Previous studies implemented various fusion deep learning architectures including CNN-RNN (Canizo et al., 2019; Kong et al., 2020), LSTM-CNN (Li et al., 2020), and ConvLSTM (Bao et al., 2019) for prediction or classification tasks. Among the various fusion deep learning approaches, CNN-RNN architecture, which sequentially combines CNN with one deep neural network of the RNN family (e.g., RNN, LSTM, or GRU), has been utilized in several studies (Canizo et al., 2019; Kong et al., 2020; Sun et al., 2019; Zhao et al., 2019a, 2019b). The CNN-RNN can learn local and hierarchical features and then extract temporal dependencies from the learned features (Zhao et al., 2019a, 2019b). In this study, daily violation data can be regarded as time series data with one dimension, while the filter of the CNN can be specified as one dimension instead of two for the image (Hu et al., 2020; Li et al., 2020). Thus, this study developed a CNN-GRU approach, which is adequate for predicting at-fault crash driver frequency with traffic violation data in previous time periods. To the best of our knowledge, this is the first study to introduce the CNN-GRU approach in traffic safety analysis.

The CNN-GRU first learns local representations of violation data using CNN, and then, the temporal dependencies from a sequence of local features are extracted by GRU. In the CNN, convolution kernels convolve input data sequences, and each patch is then transformed into a vector. The vectors assemble the output feature map by a computation of dot product between the entries of the kernel and the input, using an activation function. The computation of the CNN with multiple hidden layers can be expressed as (Abdoli et al., 2019):

$$y = F(X|\Theta) = f_L(\dots f_2(f_1(X|\Theta_1)|\Theta_2)|\Theta_L) \quad (11)$$

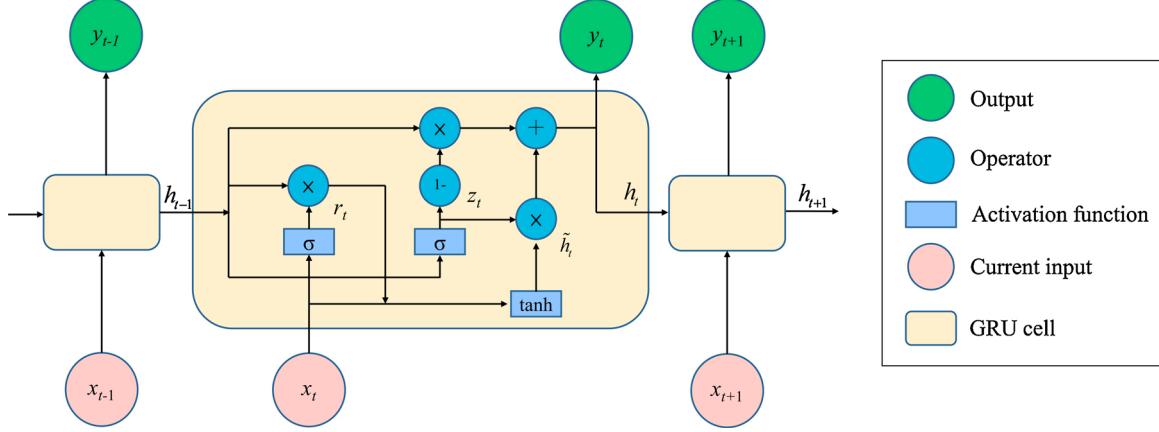


Fig. 3. Structure of GRU.

where y denotes the output of the CNN section, X denotes the input array, Θ denotes the set of parameters to be learned, and L is the number of hidden layers of the CNN. The operation of the l -th layer can be expressed as:

$$y_l = f_l(X_l | \Theta_l) = p(h(W \otimes X_l + b)), \Theta_l = [W, b] \quad (12)$$

where \otimes denotes the convolution operation, X_l is a two-dimensional input matrix of N feature maps, W is a set of N one-dimensional kernels used for extracting a new set of features from the input array, b is the bias vector, $h(\cdot)$ is the activation function, and $p(\cdot)$ is the pooling method that can downsample representations. In most cases, a rectified linear unit (ReLU) is used as the activation function and max-pooling is used for downsampling. The ReLU function and max-pooling are expressed respectively as:

$$h(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad (13)$$

$$p(x) = \max_{\forall p \in \Omega_k} y_l^p \quad (14)$$

where Ω_k denotes the pooling region with index k , and y_l^p is the p -th input feature of the l -th max-pooling layer. The output of the CNN section is then used as the input in the GRU section to extract temporal features through steps discussed in Section 4.3.1. The proposed fusion CNN-GRU architecture is illustrated in Fig. 4.

4.4. Approaches for performance comparison

The mean absolute percentage error (MAPE) was selected to measure the effectiveness of ARIMA, RF, GBRT, GRU, and CNN-GRU in this study. MAPE is defined as:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (15)$$

where y_i denotes the observed at-fault crash driver frequency in the i -th period and \hat{y}_i denotes the estimated at-fault crash driver frequency in the i -th period. The low values of MAPE indicate better model performance. We defined the prediction tasks that used MAPE as the evaluation metric for the at-fault crash driver frequency prediction tasks.

Although MAPE is a popular evaluation metric, its property of ignoring the distinction between under-prediction and over-prediction can lead to inadequate countermeasures in traffic safety analysis. For instance, assuming 105 and 100 at-fault crash drivers are observed in the i -th and $(i+1)$ -th periods, respectively, the MAPE values for both estimated 90 and 110 at-fault crash drivers for the $(i+1)$ -th period are the same at 10 %. However, the corresponding traffic safety measures for 90 and 110 at-fault crash drivers in the $(i+1)$ -th period should be different. The forecast of 90 at-fault crash drivers implies an improved traffic safety condition in the near future, while a worse condition is expected if the predictive at-fault crash driver frequency is greater than 105. To facilitate the development of traffic safety measures under limited resources in practice and to evaluate the robustness of prediction approaches, city-level crash risk prediction tasks were performed in

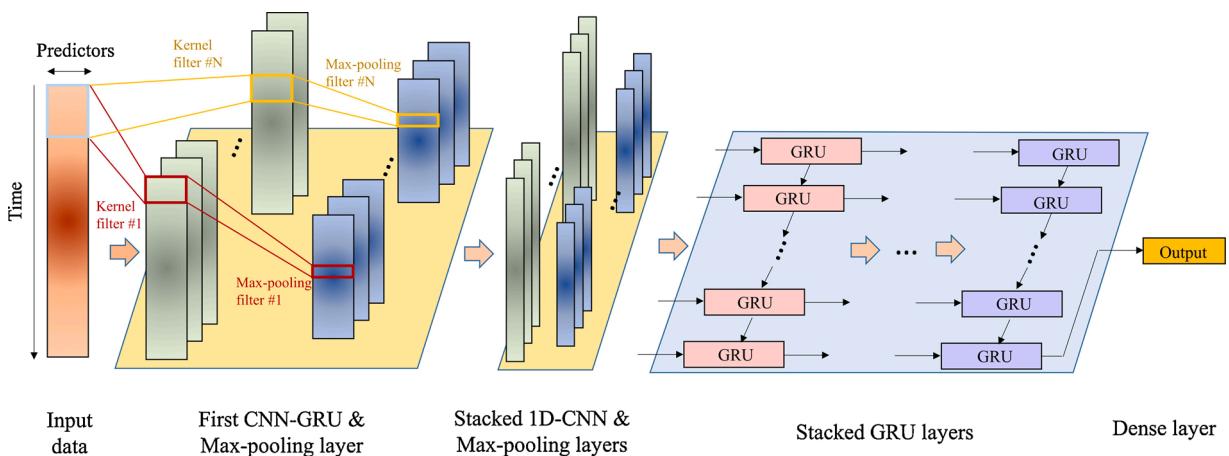


Fig. 4. Proposed CNN-GRU architecture.

addition to the at-fault crash driver frequency prediction tasks. The city-level crash risk refers to whether at-fault crash driver frequency in the $(i+1)$ -th period is greater than that in the i -th time period. Therefore, city-level crash risk prediction was formulated as a classification task. The confusion matrix of the city-level crash risk prediction tasks is shown in Table 6. We adopted an indicator of the area under the receiver-operating characteristic curve (AUC) as the evaluation metric for the city-level crash risk prediction tasks. Moreover, the type II error may mislead authorities and analysts to believe that the at-fault crash driver frequency decreases. Consequently, inadequate countermeasures or lower enforcement levels would be implemented, and this would lead to a loss of the opportunity to prepare for the worse condition in advance. Recall rate—the proportion of correctly predicted increase in at-fault crash driver frequency when the actual condition is increased—can serve as a good metric to evaluate model performance in terms of the type II error. Recall rate is defined as:

$$\text{Recall rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

The higher the recall rate, the higher is the confidence that the predictive result does not cause the type II error. This study uses recall rate as another evaluation metric for model comparison.

Further, this study calculates the mean and standard deviation values for six response variables in training sets (data from 2010 to 2016), validation sets (data from 2017 to 2018), and the predictive results (from 2017 to 2018) to investigate possible approaches to improve accuracy in future research.

4.5. Experiments

The ARIMA, RF, GBRT, GRU, and proposed CNN-GRU were implemented to predict at-fault crash drivers in 7, 15, and 30 days using traffic enforcement predictors at the city level. Prediction tasks were grouped into two categories: (1) at-fault crash driver frequency prediction and (2) city-level crash risk prediction. In addition, the ALE plots of the predictors were further investigated.

The augmented Dickey–Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test were applied on the series for the six response variables before performing ARIMA. The results of the ADF test indicate that the null hypotheses are rejected, which implies the unit root is not present; however, the presence of trend and drift is inconclusive. The results of the KPSS test show that the null hypotheses of the trend or level stationarity for KPSS are also rejected. Though the stationarity of the series may be inconclusive, this study utilizes the *auto.arima* function (Hyndman and Khandakar, 2008) in the R package “forecast” to determine the optimal hyperparameter values for the ARIMA model based on the Akaike information criterion value, while the Ljung–Box test is used to test for white noise residuals. The results of the Ljung–Box tests show that the residual series after ARIMA analysis is a white noise series.

The hyperparameters to be tuned for the machine learning and deep learning models are summarized in Tables 7 and 8, respectively. Three principles were utilized to determine the hyperparameter values: (1) enhancement of predictive performance, (2) prevention of overfitting, and (3) trade-off between predictive performance and computation cost. For the RF model, we utilized two alternative values (Section 4.2.1) for

Table 6
Confusion matrix of the city-level crash risk prediction tasks.

Change in at-fault crash driver frequency in the following period		Actual condition	
		Increase	Decrease
Predicted condition	Increase	True positive (TP)	False positive (FP) (Type I error)
	Decrease	False negative (FN) (Type II error)	True negative (TN)

Table 7
Hyperparameters for machine learning models.

RF		GBRT	
Hyperparameter	Value	Hyperparameter	Value
Max number of trees	2,000	Number of trees	2,000
Number of predictors	7, 17	Depth of each tree	2, 3, 4, 5
		Shrinkage rate	0.01, 0.025, 0.05, 0.1

Table 8
Hyperparameters for deep learning models.

GRU		CNN-GRU		
		GRU part	CNN part	
Hyperparameter	Value	Value	Hyperparameter	Value
Optimization function	RMSprop	RMSprop	Optimization function	RMSprop
Learning rate	0.001, 0.0005, 0.0002	0.001, 0.0005, 0.0002		
Activation function	tanh	tanh	Activation function	ReLU
Dropout rate	0.1–0.9	0.1–0.9	Kernel size	10
Recurrent dropout rate	0.1–0.9	0.1–0.9	Padding method	Same
Max number of units	512	128	Max number of filters	128
Max number of layers	8	3	Max number of CNN layers	4
Lookback	30	30	Pooling size	3
			Max number of pooling layers	3

the number of input predictors. For the GBRT model, the grid search over the depth of trees and the shrinkage rate were tested. For the GRU and CNN-GRU models, a grid search over three hyper-parameters—(1) dropout rate, (2) recurrent dropout rate, and (3) learning rate—was implemented to achieve better prediction performance. To save computation cost, the optimization function, activation function, look-back, kernel size, padding method, and pooling size were specified without tuning. The models were all fine-tuned and only the best results were reported in later discussions.

All models were implemented using R 3.6.1 in a Windows 10 (Microsoft) environment with an Intel Core i9-9820X@3.30 GHz processor, 64 GB RAM, and an RTX 2080Ti graphics processing card with CUDA 10.1. The R packages “forecast,” “randomForest,” and “gbm” were used to build ARIMA, RF, and GRBT models, respectively. The ALE plot was implemented using the R package “DALEX.” The GRU and CNN-GRU models were implemented on Keras using Tensorflow 1.15.0 GPU version as the backend.

All prediction tasks were conducted at three temporal levels: weekly, semimonthly, and monthly. At each temporal level, prediction models were developed using each of the five prediction approaches to predict the frequency of (1) total road users involved in crashes (TRU), (2) not give way drivers (NGWD), (3) illegal turn drivers (ITD), (4) illegal lane usage drivers (ILUD), (5) speeding drivers (SD), and (6) signal violation drivers (SVD). The data of the first 7 years of the study’s focus period, which covers 2,557 days from 2010 to 2016, was used as the training set. The remaining data from 2017 to 2018, which covers 730 days, was used to test predictive performances. The MAPE, AUC and recall rate were used as indexes to measure the performances of models for at-fault crash driver frequency and city-level crash risk prediction tasks.

5. Results and discussions

5.1. Predictive results

The results of at-fault crash driver frequency prediction tasks are summarized in Table 9. It is evident that our proposed CNN-GRU significantly outperforms other approaches by 22.64 %, 28.02 %, and 38.06 % in terms of the mean MAPE value for the weekly, semimonthly, and monthly prediction tasks, respectively. Further, GRU outperforms ARIMA and machine learning approaches on prediction tasks at all three temporal levels. In contrast, ARIMA exhibits relatively poor prediction accuracies on weekly prediction tasks in terms of the mean MAPE value but outperforms RF on semimonthly and monthly prediction tasks. Machine learning approaches RF and GBRT perform similarly; both are slightly superior to ARIMA on average on all prediction tasks. Although ARIMA seems to be inferior to other approaches in terms of the mean MAPE value, it exhibits outstanding prediction accuracy (9.048 %) in terms of the grand mean of the MAPE value for NGWD. The grand means of the MAPE values for all approaches on weekly, semimonthly, and monthly prediction tasks are 20.828 %, 18.418 %, and 16.031 %, respectively. Thus, there is an obvious general trend of the prediction accuracies for all approaches increases with the temporal level. The same findings are observed on all six response variables. This is in good agreement with the study conducted by Bao et al. (2019), where the model performance degraded sharply as the spatio-temporal resolution of the prediction tasks increased.

In terms of the response variables, the grand means of the MAPE values for TRU, NGWD, ITD, ILUD, SD and SVD are 17.336 %, 10.807 %, 16.015 %, 20.615 %, 28.657 %, and 17.123 %, respectively. The grand mean of the MAPE values for NGWD is a remarkable prediction accuracy of 10.807 %, ranging from an accuracy improvement of 36.887%–62.288% for other response variables. Although more challenges seem to exist when predicting SD, a satisfactory accuracy can still be achieved. To improve the prediction accuracy on SD, the 14-day lag assumption of the effect of directly accused citations should be further investigated because most of speeding citations are accused directly. The ARIMA model outperforms other approaches on NGWD. Machine learning approaches achieve better performances than ARIMA only on ITD and SVD. Deep learning approaches are generally superior to ARIMA and

machine learning approaches on all response variables. The RF and GBRT, as discussed previously, perform similarly on the different response variables while the proposed CNN-GRU achieves better performances than GRU in most cases.

In Table 10, we summarize the predictive performances for city-level crash risk prediction tasks, where AUC is utilized as the evaluation metric. It is found that deep learning approaches outperform ARIMA and machine learning approaches in almost all cases, whereas machine learning approaches exhibit slightly better results than ARIMA in general. Our CNN-GRU achieves much better performances compared with other approaches in almost all cases. Moreover, the CNN-GRU provides desirable prediction accuracies on ILUD and SVD as the grand means of AUC value are higher than 0.7. The ARIMA and RF perform fairly sufficiently; however, the latter achieved slightly better performances for more cases. Interestingly, the model performance decreases as the temporal level increases. The means of the AUC values for weekly, semimonthly, and monthly prediction tasks are 0.641, 0.614, and 0.582, respectively. This result is inconsistent with the results of at-fault crash driver frequency prediction tasks because the general trend of AUC for all approaches decreases with the temporal level. In addition, the results of model comparison on at-fault crash driver frequency prediction tasks are consistent with those on the city-level crash risk prediction tasks in general. However, a portion of the results of model comparison on the at-fault crash driver frequency prediction tasks may be inconsistent with those on the city-level crash risk prediction tasks. For instance, at the monthly level, the GRU outperforms CNN-GRU on SVD in terms of MAPE values; however, CNN-GRU scores higher AUC values than GRU on SVD. This provides evidence that model comparison can be enhanced by implementing both regression and classification analyses.

The grand means of the AUC values for TRU, NGWD, ITD, ILUD, SD, and SVD are 0.529, 0.662, 0.610, 0.619, 0.576, and 0.678, respectively. Apparently, the crash risks for NGWD and SVD are easier to predict. The prediction accuracies on ITD and ILUD are also good; however, all approaches present relatively poor prediction accuracies on TRU and SD. In addition to traffic violation, several factors contribute to traffic crashes, and therefore, a better performance on TRU can be expected by introducing more significant factors as model input. Similar to at-fault crash driver frequency prediction tasks, further investigation is required to determine the time lag of the effect of directly accused

Table 9
Predictive performances of at-fault crash driver frequency prediction tasks.

Temporal level	Approach	Performance (MAPE%)					
		TRU	NGWD	ITD	ILUD	SD	SVD
Weekly	ARIMA	19.852	11.926	27.576	24.124	34.375	23.167
	RF	22.848	15.756	19.422	25.180	35.059	19.512
	GBRT	20.736	13.762	18.156	27.049	36.806	21.675
	GRU	13.506	11.936	15.869	20.303	24.510	21.511
	CNN-GRU	12.913	12.831	15.307	19.558	19.820	19.783
	Mean	17.971	13.242	19.266	23.243	30.114	21.130
Semimonthly	ARIMA	18.489	8.479	19.190	21.899	33.773	20.089
	RF	20.349	12.173	16.144	23.668	35.110	15.972
	GBRT	23.779	11.675	15.511	25.011	33.822	16.615
	GRU	13.825	10.301	12.333	19.137	22.130	19.586
	CNN-GRU	11.859	9.075	11.704	16.170	21.674	12.994
	Mean	17.660	10.341	14.976	21.177	29.302	17.051
Monthly	ARIMA	18.046	6.738	18.356	19.391	34.159	19.596
	RF	19.572	11.679	15.813	20.471	34.510	14.525
	GBRT	22.182	12.130	14.005	22.099	32.888	10.757
	GRU	10.722	6.699	11.356	14.246	18.492	10.347
	CNN-GRU	11.363	6.944	9.479	10.918	12.730	10.719
	Mean	16.377	8.838	13.802	17.425	26.556	13.189
Grand mean	ARIMA	18.796	9.048	21.707	21.805	34.102	20.951
	RF	20.923	13.203	17.126	23.106	34.893	16.670
	GBRT	22.232	12.522	15.891	24.720	34.505	16.349
	GRU	12.684	9.645	13.186	17.895	21.711	17.148
	CNN-GRU	12.045	9.617	12.163	15.549	18.075	14.499
	Mean	17.336	10.807	16.015	20.615	28.657	17.123

Table 10

Predictive performances of city-level crash risk prediction tasks using AUC.

Temporal level	Approach	Performance (AUC)						
		TRU	NGWD	ITD	ILUD	SD	SVD	
Weekly	ARIMA	0.516	0.732	0.558	0.628	0.564	0.680	0.613
	RF	0.542	0.640	0.613	0.607	0.570	0.744	0.619
	GBRT	0.531	0.673	0.630	0.584	0.552	0.732	0.617
	GRU	0.587	0.708	0.675	0.671	0.562	0.759	0.660
	CNN-GRU	0.581	0.720	0.707	0.748	0.638	0.785	0.697
	Mean	0.551	0.695	0.637	0.648	0.577	0.740	0.641
Semimonthly	ARIMA	0.533	0.692	0.557	0.607	0.542	0.602	0.589
	RF	0.552	0.623	0.573	0.583	0.548	0.660	0.590
	GBRT	0.571	0.628	0.601	0.567	0.557	0.638	0.594
	GRU	0.532	0.668	0.655	0.670	0.623	0.661	0.635
	CNN-GRU	0.536	0.713	0.685	0.733	0.571	0.731	0.662
	Mean	0.545	0.665	0.614	0.632	0.568	0.658	0.614
Monthly	ARIMA	0.456	0.689	0.536	0.535	0.524	0.528	0.545
	RF	0.468	0.591	0.555	0.546	0.542	0.591	0.549
	GBRT	0.517	0.591	0.567	0.557	0.563	0.656	0.575
	GRU	0.522	0.635	0.582	0.600	0.588	0.667	0.599
	CNN-GRU	0.486	0.629	0.657	0.651	0.693	0.731	0.641
	Mean	0.490	0.627	0.579	0.578	0.582	0.635	0.582
Grand mean	ARIMA	0.502	0.704	0.550	0.590	0.543	0.603	0.582
	RF	0.521	0.618	0.580	0.579	0.553	0.665	0.586
	GBRT	0.540	0.631	0.599	0.569	0.557	0.675	0.595
	GRU	0.547	0.670	0.637	0.647	0.591	0.696	0.631
	CNN-GRU	0.534	0.687	0.683	0.711	0.634	0.749	0.666
	Mean	0.529	0.662	0.610	0.619	0.576	0.678	0.612

citations. Our CNN-GRU achieves an excellent prediction accuracy of 0.785 on SVD at the weekly level. At the same temporal level, GRU, GBRT, and RF exhibit their best performances with accuracies of 0.759, 0.732, and 0.744, respectively. The best score of ARIMA is 0.732 on NGWD at the weekly level. Although ARIMA is generally inferior to other approaches, it presents better performances than machine learning approaches on NGWD and ILUD in most cases.

This study applies the recall rate to evaluate the model performance in terms of the type II error. Table 11 presents the predictive performances for city-level crash risk prediction tasks using the recall rate. Deep learning approaches outperform ARIMA and machine learning models significantly, while ARIMA exhibit better results than machine learning models. Most of the highest recall rates were obtained by GRU

with a performance over 0.8. The proposed CNN-GRU model appears to be the best approach with respect to the means of recall rates at all temporal levels. ARIMA is superior to machine learning approaches because machine learning approaches can only achieve poor accuracy in terms of recall rate. The general trend of prediction accuracies for all approaches decreases with the temporal level. This implies the relative difficulty for the monthly level prediction task. The grand means of the recall rates for TRU, NGWD, ITD, ILUD, SD, and SVD are 0.345, 0.446, 0.445, 0.492, 0.320, and 0.556, respectively. The results are in line with those of MAPE and AUC evaluation that NGWD, ITD and SVD are relatively easier to predict.

The results of model evaluation using three metrics provide strong evidence that implementing both regression and classification analyses

Table 11

Predictive performances of city-level crash risk prediction tasks using the recall rate.

Temporal level	Approach	Performance (Recall rate)						
		TRU	NGWD	ITD	ILUD	SD	SVD	
Weekly	ARIMA	0.29	0.301	0.356	0.517	0.218	0.272	0.326
	RF	0.196	0.306	0.316	0.254	0.156	0.642	0.312
	GBRT	0.188	0.435	0.391	0.228	0.134	0.606	0.330
	GRU	0.535	0.644	0.640	0.734	0.576	0.818	0.658
	CNN-GRU	0.523	0.704	0.677	0.731	0.594	0.746	0.663
	Mean	0.346	0.478	0.476	0.493	0.336	0.617	0.458
Semimonthly	ARIMA	0.434	0.331	0.388	0.806	0.223	0.373	0.426
	RF	0.251	0.284	0.274	0.169	0.104	0.349	0.239
	GBRT	0.233	0.276	0.348	0.155	0.122	0.360	0.249
	GRU	0.439	0.626	0.538	0.802	0.546	0.807	0.626
	CNN-GRU	0.514	0.703	0.624	0.686	0.571	0.727	0.638
	Mean	0.374	0.444	0.434	0.524	0.313	0.523	0.435
Monthly	ARIMA	0.159	0.275	0.555	0.720	0.278	0.394	0.397
	RF	0.204	0.201	0.247	0.165	0.082	0.207	0.184
	GBRT	0.180	0.363	0.291	0.165	0.144	0.499	0.274
	GRU	0.496	0.618	0.402	0.575	0.421	0.779	0.549
	CNN-GRU	0.527	0.619	0.622	0.668	0.628	0.759	0.637
	Mean	0.313	0.415	0.423	0.459	0.311	0.528	0.408
Grand mean	ARIMA	0.294	0.302	0.433	0.681	0.240	0.346	0.383
	RF	0.217	0.263	0.279	0.196	0.114	0.399	0.245
	GBRT	0.200	0.358	0.343	0.183	0.133	0.488	0.284
	GRU	0.490	0.629	0.527	0.704	0.514	0.801	0.611
	CNN-GRU	0.521	0.675	0.641	0.695	0.598	0.744	0.646
	Mean	0.345	0.446	0.445	0.492	0.320	0.556	0.434

is essential for model comparison. For instance, ARIMA is slightly superior to CNN-GRU on NGWD at the semimonthly level in terms of MAPE; however, CNN-GRU outperforms ARIMA by 3.035 % and 112.387 % in terms of AUC and recall rate, respectively.

Tables 12 and 13 presents the mean and standard deviation values for six response variables in the training sets, validation sets, and predictive results. The mean and standard deviation values for the validation sets are higher than those for the training sets. For the prediction models, knowledge learned from the training sets may not be sufficient to predict at-fault crash driver frequency perfectly. The mean and standard deviation values for deep learning models are closer to the ground truth for the validation sets than those of the ARIMA and machine learning models. Among the deep learning approaches, CNN-GRU outperforms GRU by capturing the variation of response variables more precisely. To improve prediction accuracy, predictors correlated with response variables at the mid-term period can be introduced into the model. In addition, predictive results for potential predictors can be obtained—such as for weather forecast reports and predictive socio-economic characteristics in the near future—such that the variation in the dataset could be captured. This study serves as a good starting point for future studies that tend to investigate or utilize traffic violation data for traffic safety analysis and prediction tasks.

Based on the results above, several observations are highlighted:

- (1) In general, deep learning approaches outperform other approaches on both at-fault crash driver frequency and city-level crash risk prediction tasks. The proposed CNN-GRU is superior to other approaches as a fusion deep learning architecture that can extract both local and temporal features in crash and violation data. Machine learning approaches regularly exhibit better performances than ARIMA for at-fault crash driver frequency prediction tasks and city-level crash risk prediction tasks in terms of AUC, while ARIMA provides better prediction accuracies than machine learning approaches for city-level crash risk prediction tasks in terms of recall rate. Among machine learning approaches, GBRT achieved slightly better performance than RF.
- (2) Results at higher temporal levels tend to have better accuracy on at-fault crash driver frequency prediction tasks. However, the model performances decrease as the temporal level increases on

city-level crash risk prediction tasks. Moreover, the results of model comparison for at-fault crash driver frequency prediction tasks are invariably inconsistent with those for city-level crash risk prediction tasks. Consequently, we suggest that the comparison of prediction models and development of traffic safety measures should consider the predictive results from both crash frequency and crash risk prediction tasks for effective and appropriate traffic safety measures.

- (3) Among six response variables, it seems to be relatively easy to extract features from NGWD and SVD data because all approaches exhibit remarkable predictive performances on these two response variables. The prediction accuracies on ITD and ILUD are both generally good. In contrast, satisfactory prediction accuracies were obtained on TRU and SD by all approaches. Thus, we suggest that more significant factors can be introduced into the prediction model, and the time lag of the effect of directly accused citations should be further investigated.
- (4) The challenge for prediction tasks increases with an increase in the frequency and variation in the dataset. Deep learning models outperform ARIMA and machine learning approaches by capturing the trend of frequency and variation in the dataset more precisely. To improve prediction accuracy, future researches can introduce more predictors correlated with response variables in the mid-time period.

5.2. Results of ALE plot

The ALE plot was utilized in this study to investigate how each predictor influences the predictive at-fault crash driver frequency with RF and GBRT models. To make this discussion brief, we focus only on the results of the red-light running enforcement with physical policing at various temporal levels on six response variables. From Fig. 5, it is evident that as the number of red light-running citations increase, the predicted at-fault crash driver frequency decreases. Complex nonlinear relationships can be observed in Fig. 5 between the predictor and response variable, which statistical models cannot describe easily. Therefore, it is confirmed that machine learning and deep learning approaches are preferred in this scenario to extract high-dimensional features and spatio-temporal dependency in the dataset.

Table 12
Mean and standard deviation values for TRU, NGWD, and ITD.

Temporal level	Year	TRU		NGWD		ITD	
		Mean	SD	Mean	SD	Mean	SD
Weekly							
	Training set	908.33	91.28	78.92	9.92	67.33	9.94
	Validation set	1,297.53	216.08	103.74	14.66	85.75	15.46
	ARIMA	1,223.32	45.03	90.47	2.51	75.97	0.27
	RF	1,050.63	32.42	90.04	2.83	73.86	1.86
	GBRT	1,028.65	58.28	94.39	5.77	76.14	4.74
	GRU	1,308.54	109.89	101.67	7.09	85.34	2.06
	CNN-GRU	1,300.00	92.06	102.59	9.21	84.60	10.31
Semimonthly							
	Training set	1,817.93	346.66	157.91	31.06	134.73	19.01
	Validation set	2,597.83	404.83	207.51	22.78	171.58	25.61
	ARIMA	2,565.89	99.29	189.88	3.36	158.85	0.01
	RF	2,144.45	49.68	185.96	2.94	149.42	3.62
	GBRT	2,019.82	102.95	182.57	11.32	153.37	10.05
	GRU	2,585.15	147.09	204.87	18.84	166.49	11.93
	CNN-GRU	2,630.75	203.49	205.54	16.83	168.49	18.69
Monthly							
	Training set	3,902.01	710.47	338.65	61.27	288.94	32.19
	Validation set	5,583.85	818.21	445.81	38.15	368.23	46.89
	ARIMA	4,532.03	281.11	410.59	2.13	326.32	8.19
	RF	4,612.86	106.27	397.00	6.67	320.67	6.06
	GBRT	4,363.27	196.30	410.23	23.87	320.79	18.79
	GRU	5,627.01	470.89	442.86	17.46	351.81	6.81
	CNN-GRU	5,667.56	353.53	446.17	14.50	366.94	27.71

Table 13

Mean and standard deviation values for ILUD, SD, and SVD.

Temporal level	Data	ILUD		SD		SVD	
		Mean	SD	Mean	SD	Mean	SD
Weekly							
	Training set	23.45	5.68	20.60	6.00	17.92	8.52
	Validation set	32.38	7.32	39.80	10.71	24.08	5.89
	ARIMA	29.56	0.91	28.80	1.19	18.07	0.70
	RF	25.26	0.88	24.51	0.79	22.11	0.87
	GBRT	24.33	2.68	24.28	1.43	21.66	2.54
	GRU	35.73	2.71	40.50	2.87	24.54	1.26
	CNN-GRU	32.47	3.97	37.87	2.12	22.57	1.74
Semimonthly							
	Training set	46.93	10.12	41.23	10.75	35.81	10.14
	Validation set	64.78	11.65	79.74	18.95	48.30	8.28
	ARIMA	67.58	2.10	59.83	1.69	40.54	1.96
	RF	50.54	1.66	49.00	1.38	41.11	1.44
	GBRT	49.17	2.61	50.76	2.56	41.42	4.15
	GRU	70.55	4.08	76.51	21.39	52.28	8.95
	CNN-GRU	62.42	8.75	67.64	14.15	47.16	4.99
Monthly							
	Training set	100.73	17.78	88.47	20.14	76.54	18.24
	Validation set	139.07	20.30	171.51	36.60	104.01	13.87
	ARIMA	142.17	4.75	129.77	1.85	91.59	1.87
	RF	109.77	4.30	106.63	3.60	87.73	2.08
	GBRT	105.44	8.92	108.73	6.23	97.27	6.13
	GRU	138.54	11.47	155.91	24.29	103.34	2.25
	CNN-GRU	139.41	20.50	170.14	26.76	100.83	9.42

The effect of red light-running enforcement on the different response variables can also be observed from Fig. 5. It is observed that red light running enforcement is an effective traffic safety measure as its effect on deterrence is not limited to traffic signal violation but also other major violations; this is consistent with Schneider et al.'s study (2012), wherein they suggest that educational and enforcement strategies aimed at addressing any one of the high-risk behaviors are likely to have tangential effects on other high-risk behaviors. Among the different response variables, SVD is affected the most by red light running enforcement while TRU and ILUD experience weaker effects. Based on the discussion in Section 5.1, GBRT outperforms RF in general; therefore, the relationships revealed by ALE plots for GBRT model are more convincing.

This study shows the potential of the ALE plot for evaluating the power of various traffic enforcement activities among which multicollinearity may exist. By investigating the ALE plots, insight on the effect of specific traffic enforcement schemes on traffic safety can be obtained. Traffic safety agencies can analyze the ALE plots to organize highly effective traffic enforcement schemes to decrease at-fault crash driver frequency for specific driver groups and prevent traffic crashes.

5.3. Contributions

To differentiate our study from existing traffic safety analyses in the literature, the contributions of our study are summarized below.

- (1) Results from this study can contribute to the development of mid-term traffic safety measures to which insufficient attention has been paid, especially for areas where only traffic violation data is available. Using machine learning and deep learning approaches, we confirmed the feasibility of using only multicollinearity traffic violation data as predictors.
- (2) This study utilized at-fault crash driver frequency as response variable rather than crash frequency. Consequently, traffic safety measures and interventions, such as traffic enforcement and public safety broadcasting, can be developed in advance and more precisely for specific groups.
- (3) To the best of our knowledge, this is the first time that CNN-GRU, which is a fusion deep learning approach, and ALE plot are

employed in traffic safety analysis. Remarkable predictive performance was achieved by CNN-GRU while insight on the contribution of each traffic enforcement activity on traffic safety was provided by the ALE plot.

- (4) This study demonstrated the benefit of evaluating model performance by considering both crash frequency and crash risk prediction tasks. Thus, traffic safety analysts can determine the most effective prediction approach based on the performances of regression and classification tasks.

6. Conclusions

The objective of this study was to develop effective prediction approaches for predicting at-fault crash driver frequency in a mid-term period using only traffic enforcement predictors at the city level. A fusion deep learning approach combining the CNN and GRU was developed to compare the predictive performance with one econometric approach (ARIMA), two machine learning approaches (RF and GBRT), and a deep learning approach (GRU). The performances of the five approaches were tested on crash and traffic citation datasets collected from TPC-TYPD. The performance comparison was conducted on two categories of prediction tasks: (1) at-fault crash driver frequency prediction tasks, which use MAPE as the evaluation metric, and (2) city-level crash risk prediction tasks, which use AUC and recall rate. For each prediction task, prediction models were developed for TRU, NGWD, ITD, ILUD, SD, and SVD at weekly, semimonthly, and monthly levels. The experimental results demonstrate that the proposed CNN-GRU achieves remarkable prediction accuracy and outperforms other approaches because of its capability to extract both spatial and temporal features from crash and traffic violation data. The RF, GBRT, and GRU also exhibited excellent performances in general. Therefore, the desirable results of traffic safety analysis could be obtained using machine learning and deep learning approaches on only traffic crash and violation data.

Furthermore, we suggest that effective prediction approaches and appropriate traffic safety measures can be developed by considering the predictive results from both crash frequency and crash risk prediction tasks. The ALE plot was utilized to provide insight into the contribution of each traffic enforcement activity on traffic safety in scenarios where multicollinearity exists among traffic enforcement predictors. The

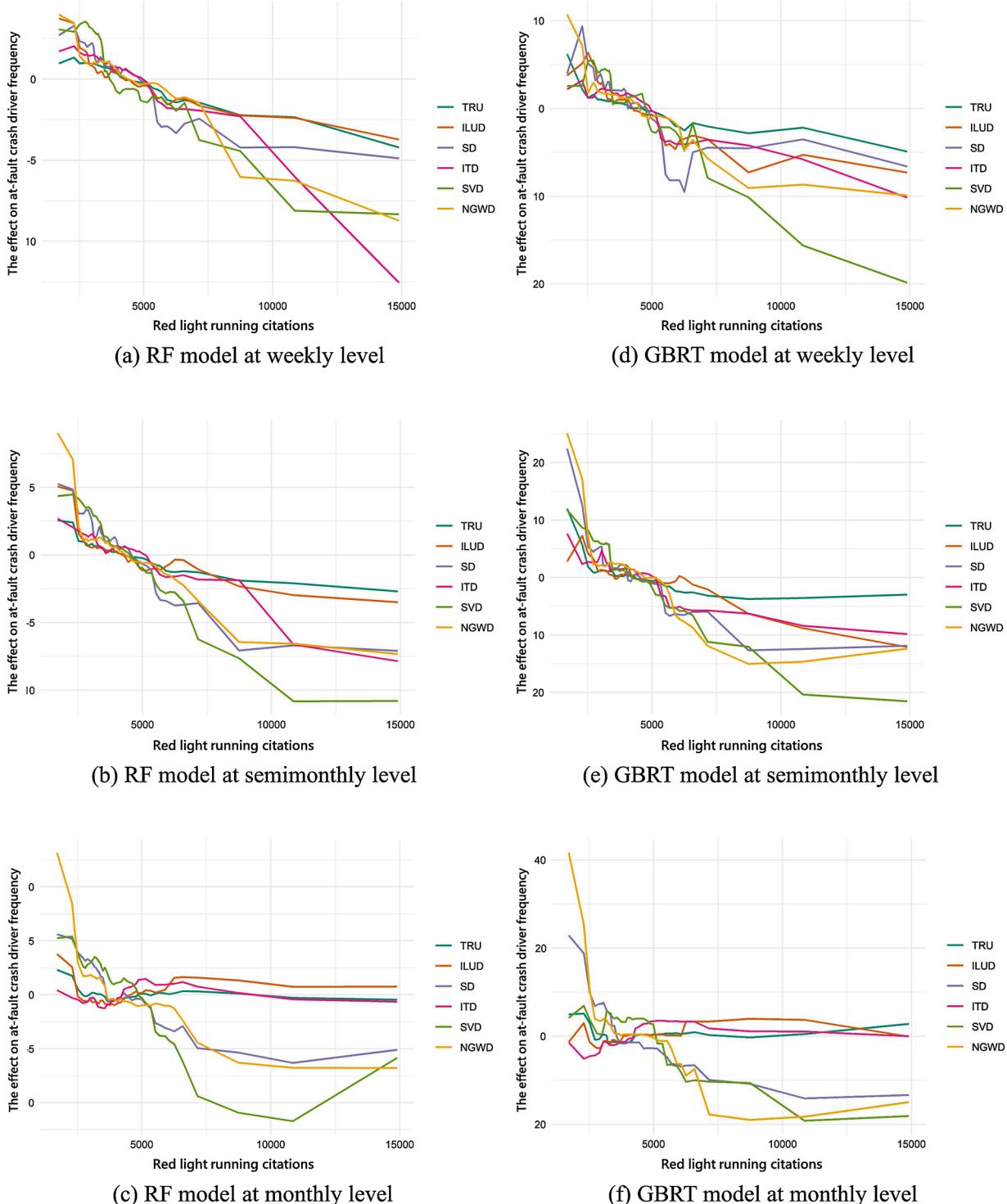


Fig. 5. ALE plots for red light running at different temporal levels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

results show that the ALE plot successfully captures complex nonlinear relationships between the number of traffic citations and predictive at-fault crash driver frequency.

There are some limitations and scope for future studies. First, the assumed 14-day time lag of the effect of directly accused traffic citations may be incorrectly determined in addition to the 30-day conservative speculation of time halo effect in Taiwanese traffic enforcement. A more precise specification of these parameters can improve predictive performances. Therefore, further investigations of these parameters are

required in future studies. Second, although the potential of the traffic violation data for traffic safety analysis has been demonstrated in this study, an enhanced predictive performance can be achieved by introducing more factors corresponding to response variables in the mid-term period as model inputs to capture increasing/decreasing frequency trends and variations in the response variable. In summary, this study serves as a good starting point for future studies that tend to investigate or utilize traffic violation data in detail.

CRediT authorship contribution statement

Yuan-Wei Wu: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Validation, Writing - original draft, Writing - review & editing. **Tien-Pen Hsu:** Conceptualization, Methodology, Supervision, Investigation, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This article was subsidized by Ministry of Science and Technology and National Taiwan University (NTU), Taiwan.

References

- Abdel-Aty, M., Abdelwahab, H., 2004. Analysis and prediction of traffic fatalities resulting from angle collisions including the effect of vehicles' configuration and compatibility. *Accid. Anal. Prev.* 36 (3), 457–469. [https://doi.org/10.1016/S0001-4575\(03\)00041-1](https://doi.org/10.1016/S0001-4575(03)00041-1).
- Abdel-Aty, M., Keller, J., Brady, P.A., 2005. Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. *Transp. Res. Rec.* 1908 (1), 37–45. <https://doi.org/10.1177/0361198105190800105>.
- Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. *Transp. Res. Part A Policy Pract.* 49, 62–75. <https://doi.org/10.1016/j.tra.2013.01.030>.
- Abdoli, S., Cardinal, P., Koerich, A.L., 2019. End-to-end environmental sound classification using a 1D convolutional neural network. *Expert Syst. Appl.* 136, 252–263. <https://doi.org/10.1016/j.eswa.2019.06.040>.
- Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accid. Anal. Prev.* 38 (3), 618–625. <https://doi.org/10.1016/j.aap.2005.12.006>.
- Ahmed, M., Abdel-Aty, M., 2013. A data fusion framework for real-time risk assessment on freeways. *Transp. Res. Part C Emerg. Technol.* 26, 203–213. <https://doi.org/10.1016/j.trc.2012.09.002>.
- Al Naser, N.B., Hawas, Y.E., Maraqa, M.A., 2013. Characterizing driver behaviors relevant to traffic safety: a multistage approach. *J. Transp. Saf. Secur.* 5 (4), 285–313. <https://doi.org/10.1080/19439962.2013.766291>.
- Ali, Q., Yaseen, M.R., Khan, M.T.I., 2019. The causality of road traffic fatalities with its determinants in upper middle income countries: a continent-wide comparison. *Transp. Res. Part A Policy Pract.* 119, 301–312. <https://doi.org/10.1016/j.tra.2018.12.002>.
- Antoniou, C., Yannis, G., 2013. State-space based analysis and forecasting of macroscopic road safety trends in Greece. *Accid. Anal. Prev.* 60, 268–276. <https://doi.org/10.1016/j.aap.2013.02.039>.
- Antonopoulos, C.N., Germeni, E., Bacopoulou, F., Kalampoki, V., Maltezos, S., Skalkidis, I., Daskalopoulou, S., Negri, E., Petridou, E., 2011. Assessing the impact of risk-taking behavior on road crash involvement among University students residing in two Mediterranean countries. *Saf. Sci.* 49 (6), 933–938. <https://doi.org/10.1016/j.ssci.2011.03.013>.
- Apley, D.W., Zhu, J., 2019. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *arXiv preprint arXiv: 1612.08468*.
- Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accid. Anal. Prev.* 122, 239–254. <https://doi.org/10.1016/j.aap.2018.10.015>.
- Basso, F., Basso, L.J., Bravo, F., Pezoa, R., 2018. Real-time crash prediction in an urban expressway using disaggregated data. *Transp. Res. Part C Emerg. Technol.* 86, 202–219. <https://doi.org/10.1016/j.trc.2017.11.014>.
- Bhowmik, T., Yasmin, S., Eluru, N., 2019. Do we need multivariate modeling approaches to model crash frequency by crash types? A panel mixed approach to modeling crash frequency by crash types. *Anal. Methods Accid. Res.* 24, 100107 <https://doi.org/10.1016/j.amar.2019.100107>.
- Blais, É., Gagn, M., 2010. The effect on collision with injuries of a reduction in traffic citations issued by police officers. *Inj. Prev.* 16 (6), 393–397. <https://doi.org/10.1136/ip.2009.025379>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017. Comparative analysis of zonal systems for macro-level crash modeling. *J. Saf. Res.* 61, 157–166. <https://doi.org/10.1016/j.jscr.2017.02.018>.
- Cai, Q., Abdel-Aty, M., Sun, Y., Lee, J., Yuan, J., 2019. Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data. *Transp. Res. Part A Policy Pract.* 127, 71–85. <https://doi.org/10.1016/j.tra.2019.07.010>.
- Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multiline roads. *Accid. Anal. Prev.* 39, 657–670. <https://doi.org/10.1016/j.aap.2006.10.012>.
- Canizo, M., Triguero, I., Conde, A., Onieva, E., 2019. Multi-head CNN-RNN for multi-time series anomaly detection: an industrial case study. *Neurocomputing* 363, 246–260. <https://doi.org/10.1016/j.neucom.2019.07.034>.
- Chen, E., Tarko, A.P., 2012. Analysis of crash frequency in work zones with focus on police enforcement. *Transp. Res.* 2280 (1), 127–134. <https://doi.org/10.3141/2280-14>.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv preprint arXiv: 1406.1078*.
- Coruh, E., Bilgic, A., Tortum, A., 2015. Accident analysis with aggregated data: the random parameters negative binomial panel count data model. *Anal. Methods Accid. Res.* 7, 37–49. <https://doi.org/10.1016/j.amar.2015.07.001>.
- De'ath, G., 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88 (1), 243–251. [https://doi.org/10.1890/0012-9658\(2007\)88%5B243:BTFEMA%5D2.0.CO;2](https://doi.org/10.1890/0012-9658(2007)88%5B243:BTFEMA%5D2.0.CO;2).
- Dong, C., Richards, S.H., Clarke, D.B., Zhou, X., Ma, Z., 2014. Examining signalized intersection crash frequency using multivariate zero-inflated Poisson regression. *Saf. Sci.* 70, 63–69. <https://doi.org/10.1016/j.ssci.2014.05.006>.
- Dong, C., Shao, C., Clarke, D.B., Nambisan, S.S., 2018. An innovative approach for traffic crash estimation and prediction on accommodating unobserved heterogeneities. *Transp. Res. Part B: Methodol.* 118, 407–428. <https://doi.org/10.1016/j.trb.2018.10.020>.
- Duddu, V.R., Pennetsa, P., Pulugurtha, S.S., 2018. Modeling and comparing injury severity of at-fault and not-at-fault drivers in crashes. *Accid. Anal. Prev.* 120, 55–63. <https://doi.org/10.1016/j.aap.2018.07.036>.
- Elassad, Z.E.A., Mousannif, H., Moatassime, H.A., 2020. Class-imbalanced crash prediction based on real time traffic and weather data: a driving simulator study. *Traffic Inj. Prev.* 21 (3), 201–208. <https://doi.org/10.1080/15389588.2020.1723794>.
- Elliott, M.A., Broughton, J., 2005. How Methods and Levels of Policing Affect Road Casualty Rates. *TRL Report TRL637*. TRL Limited, Wokingham.
- European Transport Safety Council, 1999. Police Enforcement Strategies to Reduce Traffic Casualties in Europe. European Transport Safety Council, Brussels.
- Factor, R., 2014. The effect of traffic tickets on road traffic crashes. *Accid. Anal. Prev.* 64, 86–91. <https://doi.org/10.1016/j.aap.2013.11.010>.
- Fararouei, M., Sedaghat, Z., Sadat, S.J., Shahroki, G., 2016. Risk factors for being the at-fault driver: a case-control study. *Traffic Inj. Prev.* 18 (3), 262–266. <https://doi.org/10.1080/15389588.2016.1244604>.
- Fountas, G., Anastasopoulos, P.C., 2018. Analysis of accident injury-severity outcomes: the zero-inflated hierarchical ordered probit model with correlated disturbances. *Anal. Methods Accid. Res.* 20, 30–45. <https://doi.org/10.1016/j.amar.2018.09.002>.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29 (5), 1189–1232.
- Geedipally, S.R., Patil, S., Lord, D., 2010. Examination of methods to estimate crash counts by collision type. *Transp. Res. Rec.: J. Transp. Res. Board* 2165 (1), 12–20. <https://doi.org/10.3141/2165-02>.
- Gomes, M.J.T.L., Cunto, F., da Silva, A.R., 2017. Geographically weighted negative binomial regression applied to zonal level safety performance models. *Accid. Anal. Prev.* 106, 254–261. <https://doi.org/10.1016/j.aap.2017.06.011>.
- Gomes, M.M., Pirdavani, A., Brijs, T., Pitombo, C.S., 2019. Assessing the impacts of enriched information on crash prediction performance. *Accid. Anal. Prev.* 122, 162–171. <https://doi.org/10.1016/j.aap.2018.10.004>.
- Gu, Y., Lu, W., Qin, L., Li, M., Shao, Z., 2019. Short-term prediction of lane-level traffic speeds: a fusion deep learning model. *Transp. Res. Part C Emerg. Technol.* 106, 1–16. <https://doi.org/10.1016/j.trc.2019.07.003>.
- Haleem, K., Abdel-Aty, M., Mackie, K., 2010. Using a reliability process to reduce uncertainty in predicting crashes at unsignalized intersections. *Accid. Anal. Prev.* 42, 654–666. <https://doi.org/10.1016/j.aap.2009.10.012>.
- Hastie, T., Tibshirani, R.J., Friedman, J.H., 2009. The Elements of Statistical Learning, 2nd ed. Springer-Verlag, New York, NY, USA. <https://doi.org/10.1007/978-0-387-84858-7>.
- Hu, W., McCourt, A.T., Teoh, E.R., 2011. Effects of red light camera enforcement on fatal crashes in large US cities. *J. Saf. Res.* 42 (4), 277–282. <https://doi.org/10.1016/j.jscr.2011.06.002>.
- Hu, J., Huang, M.C., Yu, X., 2020. Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models. *Accid. Anal. Prev.* 144, 105665 <https://doi.org/10.1016/j.aap.2020.105665>.
- Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. *J. Transp. Geogr.* 54, 248–256. <https://doi.org/10.1016/j.jtrangeo.2016.06.012>.
- Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. *J. Stat. Software* 27 (3), 1–22. URL: <https://www.jstatsoft.org/index.php/jss/article/view/v027i03/v27i03.pdf>. (Accessed 14 August 2020).
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction of Statistical Learning, 2nd ed., 112. Springer-Verlag, New York, NY, USA, pp. 3–7. <https://doi.org/10.1007/978-1-4614-7138-7>.
- Kibar, F.T., Celik, F., Wegman, F., 2017. Analyzing truck accident data on the interurban road Ankara-Aksaray-Eregli in turkey: comparing the performances of negative binomial regression and the artificial neural networks models. *J. Transp. Saf. Secur.* 11 (2), 129–149. <https://doi.org/10.1080/19439962.2017.1363841>.
- Kong, Z., Tang, B., Deng, L., Liu, W., Han, Y., 2020. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. *Renew. Energy* 146, 760–768. <https://doi.org/10.1016/j.renene.2019.07.033>.

- Kwon, Y., Han, S.H., Nam, C., 2012. Estimating the costs of political populism: traffic violation pardons in Korea. *Transp. Res. Part A Policy Pract.* 46, 1449–1457. <https://doi.org/10.1016/j.tra.2012.07.003>.
- Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accid. Anal. Prev.* 78, 146–154. <https://doi.org/10.1016/j.aap.2015.03.003>.
- Lee, J., Abdel-Aty, M., Cai, Q., 2017. Intersection crash prediction modeling with macro-level data from various geographic units. *Accid. Anal. Prev.* 102, 213–226. <https://doi.org/10.1016/j.aap.2017.03.009>.
- Li, X., Lord, D., Zhang, Y., Xie, Y., 2008. Predicting motor vehicle crashes using Support Vector Machine models. *Accid. Anal. Prev.* 40, 1611–1618. <https://doi.org/10.1016/j.aap.2008.04.010>.
- Li, P., Abdel-Aty, M., Yuan, J., Stipancic, J., 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. *Accid. Anal. Prev.* 135, 105371. <https://doi.org/10.1016/j.aap.2019.105371>.
- Lin, L., Wang, Q., Sadek, A.W., 2015. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. *Transp. Res. Part C Emerg. Technol.* 160, 1–16. <https://doi.org/10.1016/j.trc.2015.03.015>.
- Liu, C., Lu, C., Wang, S., Sharma, A., Shaw, J., 2019. A longitudinal analysis of the effectiveness of California's ban on cellphone use while driving. *Transp. Res. Part A Policy Pract.* 124, 456–467. <https://doi.org/10.1016/j.tra.2019.04.016>.
- Lord, D., Geedipally, S.R., 2011. The negative binomial – Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. *Accid. Anal. Prev.* 43, 1738–1742. <https://doi.org/10.1016/j.aap.2011.04.004>.
- Lord, D., Mannerling, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transp. Res. Part A Policy Pract.* 44 (5), 291–305. <https://doi.org/10.1016/j.tra.2010.02.001>.
- Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y., 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C Emerg. Technol.* 54, 187–197. <https://doi.org/10.1016/j.trc.2015.03.014>.
- Mannerling, F., Bhat, C.R., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Anal. Methods Accid. Res.* 25, 110113. <https://doi.org/10.1016/j.amar.2020.100113>.
- Mehdizadeh, M., Shariat-mohaymany, A., Nordfjærn, T., 2019. Driver behaviour and crash involvement among professional taxi and truck drivers: light passenger cars versus heavy goods vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* 62, 86–98. <https://doi.org/10.1016/j.trf.2018.12.010>.
- Ministry of Transportation and Communications, 2020. Taipei, Taiwan. URL: <https://stat.motc.gov.tw/mocdb/stmain.jsp?sys=100&funid=defjsp>. (Accessed 4 September 2020).
- Moghaddam, A.M., Tabibi, Z., Sadeghi, A., Ayati, E., Ravandi, A.G., 2017. Screening out accident-prone Iranian drivers: are they at-fault accidents related to driving behavior? *Transp. Res. Part F Traffic Psychol. Behav.* 46, 451–461. <https://doi.org/10.1016/j.trf.2016.09.027>.
- Molnar, C., 2019. A Guide for Making Black Box Models Explainable. URL: <https://christophm.github.io/interpretable-ml-book/>. (Accessed 31 January 2020).
- Montella, A., Imbriani, L.L., Marzano, V., Mauriello, F., 2015. Effects on speed and safety of point-to-point speed enforcement systems: evaluation on the urban motorway A56 Tangenziale di Napoli. *Accid. Anal. Prev.* 75, 164–178. <https://doi.org/10.1016/j.aap.2014.11.022>.
- Mussone, L., Bassani, M., Masci, P., 2017. Analysis of factors affecting the severity of crashes in urban road intersections. *Accid. Anal. Prev.* 103, 112–122. <https://doi.org/10.1016/j.aap.2017.04.007>.
- Pai, C.W., Saleh, W., 2008. Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: focusing on the effects of driver's failure to yield and junction control measures. *Accid. Anal. Prev.* 40 (2), 479–486. <https://doi.org/10.1016/j.aap.2007.08.003>.
- Pantangi, S.S., Fountas, G., Anastasopoulos, P.C., Pierowicz, J., Majka, K., Blatt, A., 2020. Do high visibility enforcement programs affect aggressive driving behavior? An empirical analysis using naturalistic driving study data. *Accid. Anal. Prev.* 138, 105361. <https://doi.org/10.1016/j.aap.2019.105361>.
- Peng, Y., Li, C., Wang, K., Gao, Z., Yu, R., 2020. Examining imbalanced classification algorithms in predicting real-time traffic crash risk. *Accid. Anal. Prev.* 144, 105610. <https://doi.org/10.1016/j.aap.2020.105610>.
- Pennetta, P., Pulugurtha, S.S., 2017. Risk drivers pose to themselves and other drivers by violating traffic rules. *Traffic Inj. Prev.* 18 (1), 63–69. <https://doi.org/10.1080/15389588.2016.1177637>.
- Porter, B.E., Johnson, K.L., Bland, J.F., 2013. Turning off the cameras: red light running characteristics and rates after photo enforcement legislation expired. *Accid. Anal. Prev.* 50, 1104–1111. <https://doi.org/10.1016/j.aap.2012.08.017>.
- Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accid. Anal. Prev.* 36, 183–191. [https://doi.org/10.1016/S0001-4575\(02\)00148-3](https://doi.org/10.1016/S0001-4575(02)00148-3).
- Quddus, M.A., 2008. Time series count data models: an empirical application to traffic accidents. *Accid. Anal. Prev.* 40, 1732–1741. <https://doi.org/10.1016/j.aap.2008.06.011>.
- Rahman, M.S., Abdel-Aty, M., Hasan, S., Cai, Q., 2019. Applying machine learning approaches to analyze the vulnerable road-users' crashes at statewide traffic analysis zones. *J. Saf. Res.* 70, 275–288. <https://doi.org/10.1016/j.jscr.2019.04.008>.
- Ren, H., Song, Y., Wang, J., Hu, Y., Lei, J., 2018. A deep learning approach to the citywide traffic accident risk prediction. In: Proc. 2018 21st Int. Conf. Intell. Transp. Syst. (ITSC). Maui, HI, USA, pp. 3346–3351. <https://doi.org/10.1109/ITSC.2018.8569437>.
- Ryeng, O.E., 2012. The effect of sanctions and police enforcement on drivers' choice of speed. *Accid. Anal. Prev.* 45, 446–454. <https://doi.org/10.1016/j.aap.2011.08.010>.
- Salum, J.H., Kitali, A.E., Bwire, H., Sando, T., Alluri, P., 2019. Severity of motorcycle crashes in Dar es Salaam, Tanzania. *Traffic Inj. Prev.* 20 (2), 189–195. <https://doi.org/10.1080/15389588.2018.1544706>.
- Savolainen, P., Mannerling, F., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. *Accid. Anal. Prev.* 39 (5), 955–963. <https://doi.org/10.1016/j.aap.2006.12.016>.
- Schneider, W.H., Savolainen, P., Boxel, D.V., Beverley, R., 2012. Examination of factors determining fault in two-vehicle motorcycle crashes. *Accid. Anal. Prev.* 45, 669–676. <https://doi.org/10.1016/j.aap.2011.09.037>.
- Shaaban, K., 2017. Assessment of drivers' perceptions of various police enforcement strategies and associated penalties and rewards. *J. Adv. Transp.* <https://doi.org/10.1155/2017/5169176>. Article Id: 5169176.
- Shew, C., Pande, A., Nuworsoo, C., 2013. Transferability and robustness of real-time freeway crash risk assessment. *J. Saf. Res.* 46, 83–90. <https://doi.org/10.1016/j.jstr.2013.04.005>.
- Siddiqui, C., Abdel-Aty, M., Huang, H., 2012. Aggregate nonparametric safety analysis of traffic zones. *Accid. Anal. Prev.* 45, 317–325. <https://doi.org/10.1016/j.aap.2011.07.019>.
- Stanojević, P., Jovanović, D., Lajunen, T., 2013. Influence of traffic enforcement on the attitudes and behavior of drivers. *Accid. Anal. Prev.* 52, 29–38. <https://doi.org/10.1016/j.aap.2012.12.019>.
- Sun, Y., Lo, F.P.W., Lo, B., 2019. EEG-based user identification system using 1D-convolutional long short-term memory neural networks. *Expert Syst. Appl.* 125, 259–267. <https://doi.org/10.1016/j.eswa.2019.01.080>.
- Sze, N.N., Wong, S.C., Pei, X., Choi, P.W., Lo, Y.K., 2011. Is a combined enforcement and penalty strategy effective in combating red light violations? An aggregate model of violation behavior in Hong Kong. *Accid. Anal. Prev.* 45, 317–325. <https://doi.org/10.1016/j.aap.2010.08.020>.
- Theofilato, A., 2017. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. *J. Saf. Res.* 61, 9–21. <https://doi.org/10.1016/j.jstr.2017.02.003>.
- Ülengin, F., Önsel, Ş., Topcu, Y.İ., Aktaş, E., Kabak, Ö., 2007. An integrated transportation decision support system for transportation policy decisions: the case of Turkey. *Transp. Res. Part A Policy Pract.* 41 (1), 80–97. <https://doi.org/10.1016/j.tra.2006.05.010>.
- United Nations, 2010. Global Plan for the Decade of Action for Road Safety 2011–2020. UN Road Safety Collaboration. Available at: https://www.who.int/roadsafety/de_code_of_action/plan/en/. (Accessed 16 January 2020).
- Urie, Y., Velaga, N.R., Maji, A., 2016. Cross-sectional study of road accidents and related law enforcement efficiency for ten countries: a gap coherence analysis. *Traffic Inj. Prev.* 17 (7), 686–691. <https://doi.org/10.1080/15389588.2016.1146823>.
- Vaa, T., 1999. Increased police enforcement: effects on speed. *Accid. Anal. Prev.* 29 (3), 373–385. [https://doi.org/10.1016/S0001-4575\(97\)00003-1](https://doi.org/10.1016/S0001-4575(97)00003-1).
- Vanlaar, W., Robertson, R., Marcoux, K., 2014. An evaluation of Winnipeg's photo enforcement safety program: results of time series analyses and an intersection camera experiment. *Accid. Anal. Prev.* 62, 238–247. <https://doi.org/10.1016/j.aap.2013.09.023>.
- Wang, J., Liu, B., Fu, T., Liu, S., Stipancic, J., 2019. Modeling when and where a secondary accident occurs. *Accid. Anal. Prev.* 130, 160–166. <https://doi.org/10.1016/j.aap.2018.01.024>.
- Washington, S.P., Karlaftis, M.G., Mannerling, F.L., Anastasopoulos, P., 2020. Statistical and Econometric Methods for Transportation Data Analysis, 2nd ed. Chapman Hall/CRC, Boca Raton, FL, USA.
- Williams, A.F., Wells, J.K., 2004. The role of enforcement programs in increasing seat belt use. *J. Saf. Res.* 35 (2), 175–180. <https://doi.org/10.1016/j.jscr.2004.03.001>.
- Wu, P., Meng, X., Song, L., 2019. A novel ensemble learning method for crash prediction using road geometric alignments and traffic data. *J. of Transp. Saf. Secur.* 1–19. <https://doi.org/10.1080/19439962.2019.1579288>.
- Xie, S.Q., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. *Accid. Anal. Prev.* 121, 285–294. <https://doi.org/10.1016/j.aap.2018.09.030>.
- Xu, C., Bao, J., Wang, C., Liu, P., 2018. Association rule analysis of factors contributing extraordinarily severe traffic crashes in China. *J. Traffic Saf. Res.* 67, 65–75. <https://doi.org/10.1016/j.jtsr.2018.09.013>.
- Yannis, G., Papadimitriou, E., Antoniou, C., 2007. Multilevel modelling for the regional effect of enforcement on road accidents. *Accid. Anal. Prev.* 39 (4), 818–825. <https://doi.org/10.1016/j.aap.2006.12.004>.
- Yannis, G., Papadimitriou, E., Antoniou, C., 2008. Impact of enforcement on traffic accidents and fatalities: a multivariate multilevel analysis. *Saf. Sci.* 46, 738–750. <https://doi.org/10.1016/j.ssci.2007.01.014>.
- Yu, R., Wang, Y., Zou, Z., Wang, L., 2020. Convolutional neural networks with refined loss functions for the real-time crash risk analysis. *Transp. Res. Part C Emerg. Technol.* 119, 102740. <https://doi.org/10.1016/j.trc.2020.102740>.
- Zeng, Q., Sun, J., Wen, H., 2017. Bayesian hierarchical modeling monthly crash counts on freeway segments with temporal correlation. *J. Adv. Transp.* <https://doi.org/10.1155/2017/5391054>. Article Id: 5391054.
- Zhang, Y., Xie, Y., Li, L., 2012. Crash frequency analysis of different types of urban roadway segments using generalized additive model. *J. Saf. Res.* 43 (2), 107–114. <https://doi.org/10.1016/j.jstr.2012.01.003>.

- Zhang, X., Waller, S.T., Jiang, P., 2019. An ensemble machine learning-based modeling framework for analysis of traffic crash frequency. Computer-Aided Civil Infrastruct. Eng. 35 (3), 258. <https://doi.org/10.1111/mice.12485>.
- Zhao, F., Mao, X., Chen, L., 2019a. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomed. Signal Process. Control 47, 312–323. <https://doi.org/10.1016/j.bspc.2018.08.035>.
- Zhao, S., Wang, K., Liu, C., Jackson, E., 2019b. Investigating the effects of monthly weather variations on Connecticut freeway crashes from 2011 to 2015. J. Saf. Res. 71, 153–162. <https://doi.org/10.1016/j.jsr.2019.09.011>.