Accident Analysis and Prevention 144 (2020) 105665

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models

Jiajie Hu[a], Ming-Chun Huang[b], Xiong Yu (Bill)[c,*]

[a] Department of Electrical Engineering and Computer Science, Case Western Reserve University, 2104 Adelbert Road, Bingham 279, Cleveland, OH 44106-7201, United States

[b] Department of Electrical Engineering and Computer Science, Case Western Reserve University, 10900 Euclid Avenue, Glennan 514B, Cleveland, OH 44106-7201, United States

[c] Department of Civil Engineering, Department of Electrical Engineering and Computer Science (Courtesy), Case Western Reserve University, 2104 Adelbert Road, Bingham 206, Cleveland, OH 44106-7201, United States

ARTICLE INFO

Keywords:
Connected vehicles
Multi-layer perceptron
Convolutional neural network
Crash-prone intersection
Basic safety messages
Surrogate measures of safety

ABSTRACT

Traditional methods for identifying crash-prone roadways are mainly based on historical crash data. It usually requires more than three years to collect a sufficient amount of dataset for road safety assessment. However, the emerging connected vehicles (CVs) technology generates rich instantaneous information, which can be used to identify dangerous road sections proactively. Information about the identified crash-prone intersections can be shared with the surrounding vehicles via CVs communication technology to promote cautious driving behaviors; in the longer term, such information will guide the implementation of countermeasures to prevent potential crashes. This study proposed a deep-learning based method to predict the risk level at intersections based on CVs data from the Michigan Safety Pilot program and historical traffic and intersection crash data in areas around Ann Arbor, Michigan, USA. One month of data by CVs at intersections were used for analyses, which accounts for about 3%–12% of overall trips. The risk levels of 774 intersections (i.e., low, medium and high risk) are determined by the annual crash rates. Feature extraction process is applied to both CV's data and traffic data at each intersection and 24 features are extracted. Two black-box deep-learning models, multi-layer perceptron (MLP) and convolutional neural network (CNN) are trained with the extracted features. A number of hyperparameters that affect prediction performance are fine-tuned using Bayesian optimization algorithm for each model. The performance of the two deep learning models, which are black-box models, were also compared with a decision tree model, a white-box type of simple machine learning model. The results showed that the accuracies of deep learning (DL) models were slightly better (both over 90 %) than the decision tree model (about 87 %). This indicated that the DL models were capable of uncover the inherent complexity from the dataset and therefore provided higher accuracy than the traditional machine learning model. CNN model achieves slightly higher accuracy (93.8 %) and is recommended as the classifier to predict the risk level at intersections in practice. The interpretability analysis of the CNN model is conducted to confirm the validity of the model. This study shows that combination of CVs data (V2V and V2I) and deep learning networks (i.e. MLP and CNN used in this paper) is promising to determine crash risks at intersections with high time efficiency and at low CV penetration rates, which help to deploy countermeasures to reduce the crash rates and resolve traffic safety problems.

1. Introduction

Accurate identification of crash-prone locations can help effectively allocate resources for safety improvements. Traditional methods for identifying crash-prone roadways are mainly based on collecting historical crash data, which generally requires several years. The emerging connected vehicles (CVs) technologies are enabling new applications to address real-world safety problems such as crashes. The CVs consist of hardware, software, and firmware that allows for the dynamic transmission of messages between the vehicles (V2V) and between the vehicle and infrastructure (V2I) (Lu and Cheng, 2014). The type and components of CV messages are specified in the Society of Automotive Engineers (SAE) J2735 standard and are broadcast via Dedicated Short Range Communication (DSRC) (Committee, 2009). Basic Security

* Corresponding author.
E-mail addresses: jxh919@case.edu (J. Hu), mxh602@case.edu (M.-C. Huang), xxy21@case.edu (X. Yu).

https://doi.org/10.1016/j.aap.2020.105665
Received 26 December 2019; Received in revised form 24 June 2020; Accepted 27 June 2020
Available online 16 July 2020
0001-4575/

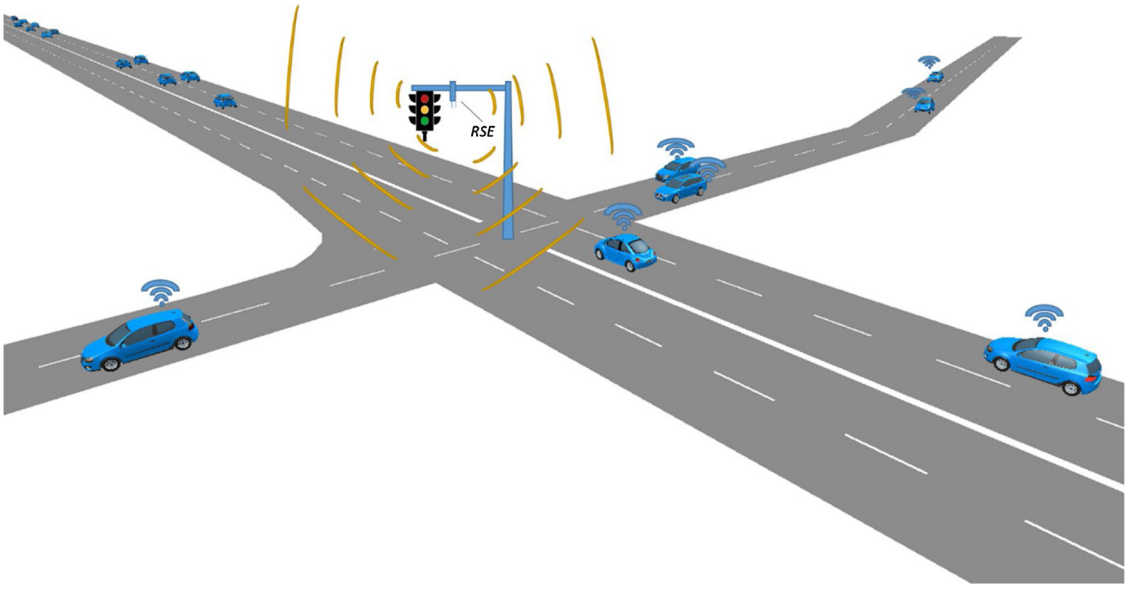

Fig. 1. Schematic of real-time intersection collision risk detection and warning system based on CVs and RSE at intersection (only CVs and RSE are shown for illustration purpose. Regular vehicles, which are a large portion of the traffic, not shown).

Messages (BSM) is one type of messages defined by J2735, which consists of two parts. Part 1 of BSM contains core data elements including vehicle size, position, speed, heading, acceleration, steering wheel angle, etc. Part 2 of BSM consists of a number of optional components such as air temperature, wiper and lamp status, Traction Control System (TCS) and Anti-lock Braking System (ABS) activation. Currently, AASHTO has established an infrastructure blueprint to achieve a mature transportation infrastructure system for connected vehicle by 2040. It is anticipated by 2040 most of the operating vehicles on the road will be connected vehicles, and up to 80 % of traffic signals and 25,000 roadside locations will be V2I-enabled (Wright and Garrett, 2014).

This study aims to exploit the potential of using CVs data to identify crash-prone intersections and provide guidance for proactive countermeasures to prevent crashes at intersections. The schematic of the proposed detection system is shown in Fig. 1. By deploying roadside equipment (RSE) device at the intersection, data from the surrounding CVs can be collected and processed with the pre-trained deep learning model to predict the crash risk level at the intersection. The predicted risk level can then be broadcast (by RSE via DSRC or 5 G technology or dynamic warning signs etc.) to vehicles (human operated or autonomous driving) to promote cautious driving behaviors. In the longer term, such information will guide the implementation of countermeasures to improve the intersection design and promote safe driving behaviors at intersections.

Connected vehicle data used in this study is from the Safety Pilot Model Deployment (SPMD) program, which was conducted by U.S. Department of Transportation (USDOT) in Ann Arbor, Michigan. The data collected on April 2013 is used to train and evaluate the performance of the proposed deep-learning models. The field test program was conducted on 75 miles of instrumented roadway with about 3 %–12 % CV market penetration rate. 28 RSE devices and around 3000 connected vehicles which capable of V2V and V2I communication were deployed throughout the road network. There are 774 intersections in the test area. Those intersections can be classified into three categories of risk levels (low, medium, high risk) based on the annual crash rate. By common practice, priorities are set for countermeasures to be deployed at high-risk intersections. The problem to identify or assess the crash-prone intersection can be formulated as a task to classify or predict the risk level at a specific intersection. This can be effectively accomplished by the emerging deep learning method. Deep learning, also known as deep neural networks, is one of the most promising techniques in the emerging machine learning field (LeCun and Bengio, 2015). It overcomes the shortcomings of traditional machine learning methods and has been implemented successfully in many areas, such as computer vision, natural language processing, and speech recognition (Collobert and Weston, 2008, 2011; Hinton and Deng, 2012; Graves and Mohamed, 2013; Manning and Surdeanu, 2014; Goodfellow and Bengio, 2016; Szegedy and Vanhoucke, 2016; Zhong and Hu, 2020). There have also been studies on the implementation of deep learning in transportation applications such as traffic flow prediction, road condition detection, and congestion recognition (Lv and Duan, 2015; Ma and Yu, 2015; Zhang and Yang, 2016; Polson and Sokolov, 2017; Zhang and Li, 2019). With the “big data” generated by CVs technology, deep learning model is advantageous in extracting features hidden in the data and learn the knowledge and relations from the data. Therefore, integration of big data from CVs technology with emerging deep learning algorithms is well-suited to provide new solutions to transportation safety issues. Therefore two deep-learning models including multi-layer perceptron (MLP) and convolutional neural network (CNN) are proposed in this paper to classify the collision risk levels at the intersections.

The rest of this paper is structured as follows: a literature review is firstly introduced to provide an overview on previous related work. This is followed by description of CV data acquisition and preprocessing. The theoretical basis of two proposed deep learning models is then described, followed by presentation of modeling results and discussions of interpretability. The conclusion of the study and future work are provided in the last section.

2. Literature review

The traditional method for crash-prone road identification generally requires collecting historical crash data for a few years. To overcome the limitations, many previous studies focused on developing more time-efficient methods for crash-prone road determination. Among these, the Surrogate Safety Measures (SSMs) approach has received extensive attention. The idea behind this approach is that SSMs are causally related to the likelihood of crashes, and therefore, they are indicators of road safety performance by understanding the process that leads to crashes (Davis and Hourdos, 2011; Yang, 2012; Laureshyn and Johnsson, 2016; Gao and Ozbay, 2018).

Time to collision (TTC) is one of the most widely used SSMs. TTC is defined as the time required for two vehicles to collide if they continue driving on the same path at their present speeds. A car-following scenario is considered unsafe if its TTC value is less than a certain threshold time because there will be insufficient amount of time for driver of the following vehicle to respond and take preventative actions such as braking or changing lanes to avoid collision. Previous studies also proposed different TTC threshold values. Horst (1991) suggested a TTC value of 4 s to distinguish between safe and uncomfortable situations while driving on the roads. Hogema and Janssen (1996) suggested a minimum TTC value of 3.5 s for drivers without an automatic cruise control system and 2.6 s for drivers with equipped vehicles. Xie and Yang (2018) found the optimal TTC thresholds of 2.3 s through sensitivity analyses. Due to its simplicity, TTC has been widely used to evaluate the collision risk of car-following scenarios (van der Horst and Hogema, 1993; Yang, 2012). However, the TTC equation assumes the following vehicle keeps its speed until the collision and did not consider the effects of acceleration or deceleration processes. A modified TTC (MTTC) was proposed by Ozbay and Yang (2008) by taking acceleration/deceleration rate of leading and following vehicles into consideration. Results by traffic simulations illustrated that the MTTC is correlated to the actual crash and features the capability to highlight the high-risk locations. MTTC with a threshold value of 4 s was used to identify the potential collision. Deceleration rate to avoid a crash (DRAC) is another widely used safety surrogate measure. It is defined as the minimum deceleration rate required by the following vehicle to come to a timely stop (or match the leading vehicle's speed) and hence to avoid a crash (Cooper and Ferguson, 1976). A vehicle is at risk if its DRAC value exceeds a certain threshold value. Archer (2005) suggested that a given vehicle is likely to collide if its DRAC exceeds a threshold braking value of 3.35 m/s^2. The American Association of State Highway and Transportation Officials (AASHTO) (Officials, 2011) recommends a slightly higher threshold of 3.40 m/s^2 for most drivers. Xie and Yang (2018) determined the threshold value of 3.0 m/s^2 to achieve the highest correlation coefficients with actual rear-end crashes. It has been confirmed that DRAC is an effective measure of safety performance and can be used to identify high-risk road sections for rear end collisions (Guido and Saccomanno, 2010). Besides these three widely used surrogate measure which described above, many other useful surrogate measures have been derived such as time exposed TTC (TET) and time-integrated TTC (TIT) (Minderhoud and Bovy, 2001), crash potential index (CPI) (Cunto and Saccomanno, 2008), maximum available deceleration rate (MADR) (Cunto and Saccomanno, 2008), aggregated crash index (ACI) (Kuang and Qu, 2015), etc. However, there are three major shortcomings of the SSMs approach. Firstly, it is difficult to determine the threshold values that are used to distinguish different risk levels. The arbitrary selection of threshold might result in inaccurate outcomes since the definition of risk levels are threshold sensitive and are determined by different thresholds. Although many studies proposed different threshold values, standard threshold values have not been widely accepted. Secondly, most of the criteria used in the derivations of SSMs are primarily reliant on the car following model causing the safety surrogate measures to achieve a high correlation with rear-end crashes only. Other types of crashes such as side-impact crashes, head-on crashes, and single-vehicle crashes are not captured by the SSMs. As a result, SSMs approach has major limitations in predicting the overall crash rate at intersections. Thirdly, data used to calculate SSMs, such as the relative distance, velocity, and acceleration of vehicles, is not easily collected. Additional sensors are needed such as Mobileyes, which incur expensive installation cost and is hard to implement on a large scale.

Increasing cases of recent studies have explored machine learning technology into crash risk prediction. Compared with conventional methods, machine learning methods do not need to set any threshold, instead it allows the computer to learn automatically and adjust accordingly without human intervention. From these, an optimal decision can be made based on data. This method is capable of finding hidden patterns in the provided data, making it highly generalized rather than limited to a particular crash scenery (i.e., rear-end crashes). Bao and Liu (2019) proposed a spatiotemporal convolutional long short-term memory network (STCL-Net) for the prediction of city-wide short-term crash risk. The dataset was collected from Manhattan in New York City, which includes crash data, taxi GPS data, road network attributes, land use features, population data and weather data. The analysis results showed that the proposed deep learning model can better capture the spatiotemporal characteristics and achieved a higher prediction accuracy rate and a lower false alarm rate than the traditional econometric models. Theofilatos and Chen (2019) used both state-of-art machine learning models and emerging deep learning models to predict the real-time crash occurrence and compared the perdition performances. The dataset includes historical crash data, real-time traffic and weather data collected from Attica Tollway in Greece. The testing results showed that the deep learning model outperformed all of the candidate machine learning models and achieved a balanced performance among accuracy, precision, recall and AUC. Many other studies also used machine learning/deep learning methods with multi-source dataset to predict crash risk and obtained good results (Iranitalab and Khattak, 2017; Chen and Wu, 2019; Yuan and Abdel-Aty, 2019). However, few studies used data collected from connected vehicles for crash risk prediction. This study aims to fill this important gap by proposing a new approach to identify crash-prone intersections with CVs technology and deep learning method. Moreover, by sharing the information of likelihood of crash-prone intersections with the surrounding vehicles via V2V or I2V communications, driver behaviors can be adjusted to mitigate traffic crashes. This also helps to guide the implementation of traffic countermeasures to prevent potential crashes.

3. Data acquisition and processing

3.1. Data assembling, cleaning and preprocessing

Data from the SPMD program is used in this study. As described in the earlier context, the program is conducted by USDOT in Ann Arbor, Michigan between August 18, 2011 and August 29, 2014. SPMD is one of the first pilot projects conducted by FHWA with comprehensive CVs data collection under realistic conditions. The distributed data by this project contains eight datasets, among which the DAS1 dataset, collected during April 2013 by the University of Michigan Transportation Research Institute (UMTRI), is used in this study. In the DAS1 dataset, the motion information of host vehicles, including position, epoch time, speed, longitudinal acceleration, angular velocity, etc., was collected via vehicle onboard Wireless Safety Units (WSUs) at a sample rate of 10 Hz. The total dataset contains 62,589,725 messages, collected by 90 equipped vehicles. For simplicity, the dataset is called CVs data in the subsequent section of the paper.

Data quality check and cleaning are performed before further analysis. Firstly, the records with invalid WSU, CAN Bus and GPS messages are filtered out. Then erroneous data points are removed based on the abnormal values of speed, longitudinal acceleration, and angular velocity, etc. Records with vehicle speed higher than 220 kph (140 mph), the absolute values of acceleration rate greater than 10 m/s^2, and the absolute values of angular velocity larger than 180 deg/s are removed as outliers. Finally, a Hampel filter (Pearson and Neuvo, 2016) is applied to the vehicle kinetic data such as speed, acceleration and angular velocity to remove the noise due to the instrument issues. This filter works by using a moving window to replace the central value in the window with its median if it is an outlier (i.e., the value is significantly different from the median). The working process is as follows:

A moving window of length $2K+1$ is applied to select a section of data (Eq. (1))

Compute the medium value of selected data (Eq. (2))

1) Criteria S_i is defined based on the difference between the data and medium value (Eq. (3)) and used subsequently to determine outliers
2) The raw data is processed based on filtering criteria (Eq. (4)). Data in the center of the window is replaced with the median value if it is an outlier (i.e., its value is significantly different from the median value of the selected section of data).

$$W_i^K = \{x_{i-K},...,x_i,...,x_{i+K}\} \quad (1)$$

$$m_i = median\{x_{i-K},...,x_i,...,x_{i+K}\} \quad (2)$$

$$S_i = \alpha \times median_{j\in[-K,K]}\{|x_{i-j} - m_i|\} \quad (3)$$

$$y_i = \begin{cases} x_i & |x_i - m_i| \le tS_i \\ m_i & |x_i - m_i| > tS_i \end{cases} \quad (4)$$

where, W_i^K is the selected data vector by applying a moving window of length $2K+1$ (K is a parameter that affects the length of the window), x_i is the data at the center of the window, m_i is the median value of the selected data vector, α and t are parameters to define the criteria for outliers, α is set as a constant 1.4826, t is a tuning parameter that defines the filtering criteria. The Hampel filter reduces to the standard median filter when $t = 0$, and is more versatile in preprocessing the data to reduce the effects of outliers. It delivers less effects on the raw signal with increasing t. The parameters affecting the length of data window (K) and tunning parameter (t) are selected based on recommendations from Pearson and Neuvo (2016).

34,644,861 data points are retained after the data cleaning process. The descriptive statistics of the selected variables are shown in Table 1.

Intersection-related traffic data such as the annual average number of crashes (2008–2012), number of approaches, annual average daily traffic (AADT) of major roads and traffic signal light are obtained from Southeast Michigan Council of Governments (SEMCOG, https://semcog.org). In this study, the territory of the intersections is limited to 30-meter along each approach from the center of the intersection, referring to the definition of high-frequency crash locations by SEMCOG. This is set because 30-meter is sufficiently large to cover crashes at an intersection or caused by an intersection while small enough to distinguish two adjacent intersections. Moreover, it is noticed that 30-meter can capture the frequent starting and braking of vehicles very well which is an important feature for risk level prediction. With this criteria, 774 intersections are selected in the Ann Arbor area. A total of 14,205 recorded crashes occurred on the selected intersections, among which 45 % are rear-end crashes, 24 % are angle crashes, 15 % are side-wipe crashes, 5 % are head-on crashes, and 11 % are other crashes. For all intersections, the mean annual crash rate is 3.67 with a standard deviation of 5.64. The distribution of total crash rates (below 20) is shown in Fig. 2. Additionally, there are 21 intersections with a crash rate exceeding 20. Referred to the study of SEMCOG and predefined intersection territory, the risk level of intersections is divided into three categories based on annual crash rate. The intersections with a crash rate less than or equal to 2 are classified as low-risk. The intersections with a crash rate greater than 2 and less than or equal to 7 are classified as medium-risk. The intersections with a crash rate exceeding 7 are classified as high-risk. The total numbers of low, medium and high risk intersections based on this criteria are 474, 183, and 117, respectively. Fig. 3 shows the distribution of risk level at the selected intersections in Ann Arbor area. The red lines on the map indicate the major roads. It is noted that crashes tend to be concentrated at the intersection of major roads, especially in the downtown areas.

Table 1
Descriptive statistics of selected variables.

Category	Variable	Mean	SD	Min	Max
Position	Latitude (°)	42.28	0.02	42.23	42.32
	Longitude(°)	−83.72	0.02	−83.8	−83.68
Motion	Speed (kph)	38.90	25.95	0.16	140
	Longitudinal acceleration (m/s^2)	0	0.76	−9.97	9.95
	Angular velocity (°/s)	−0.01	5.24	−179.5	154.9

*SD: standard deviation.

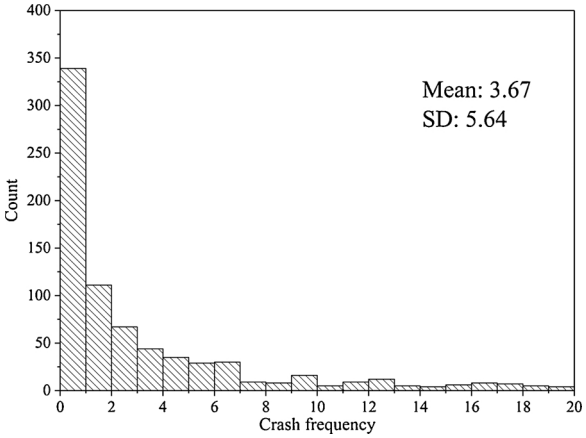

Fig. 2. Histogram of annual crash rates at intersections.

3.2. Feature extraction

For each intersection, a radius of 30 m is used to allocate pre-processed CVs data to the corresponding intersection, based on simple geometry calculations using the corresponding longitudinal and latitude values. These data points allocated to each intersection are used to calculate 19 different variables representing the features of each intersection. To distinguish these variables, a notation system is used where variables have two terms in their names separated by a dash "-". The first term indicates the element to which the variable is applied from the data points, which is shown below:

- Speed: vehicular driving speed
- Yaw: both positive and negative values of vehicle angular velocity
- YawPos: positive values of vehicle angular velocity (right turn)
- YawNeg: negative values of vehicle angular velocity (left turn)
- AccDec: both positive and negative values of acceleration
- Acc: positive values of acceleration
- Dec: negative values of acceleration
- SudAccDec: either sudden acceleration or deceleration (True = 1, False = 0)
- SudAcc: sudden acceleration (True = 1, False = 0)
- SudDec: sudden deceleration (True = 1, False = 0)
- WaitTime: duration vehicles waiting at an intersection
- WaitRatio: the proportion of vehicle waiting time to total passage time at an intersection

Here the sudden acceleration/deceleration is used to evaluate the dangerous driving behavior. The definition of sudden acceleration/deceleration is based on different acceleration thresholds that vary with speed of the vehicle. The thresholds measured by speed section are shown in Table 2 (Han and Yang, 2009; Kim and Choi, 2013).

The second term shows what measures are applied to obtain the variables include sum, mean and standard deviation. For example, if the standard deviation is applied to the deceleration, the variable will be named as "Dec-S_{dev}".

Apart from CVs data, the data from SEMCOG are also used for feature extraction. Therefore, there are up to 24 total number of features at each intersection. The descriptive statistics of these features are shown in Table 3. The standardization process is then applied to the whole dataset to normalize the ranges of different variables to the same scale, before the data are used for training the deep-learning models.

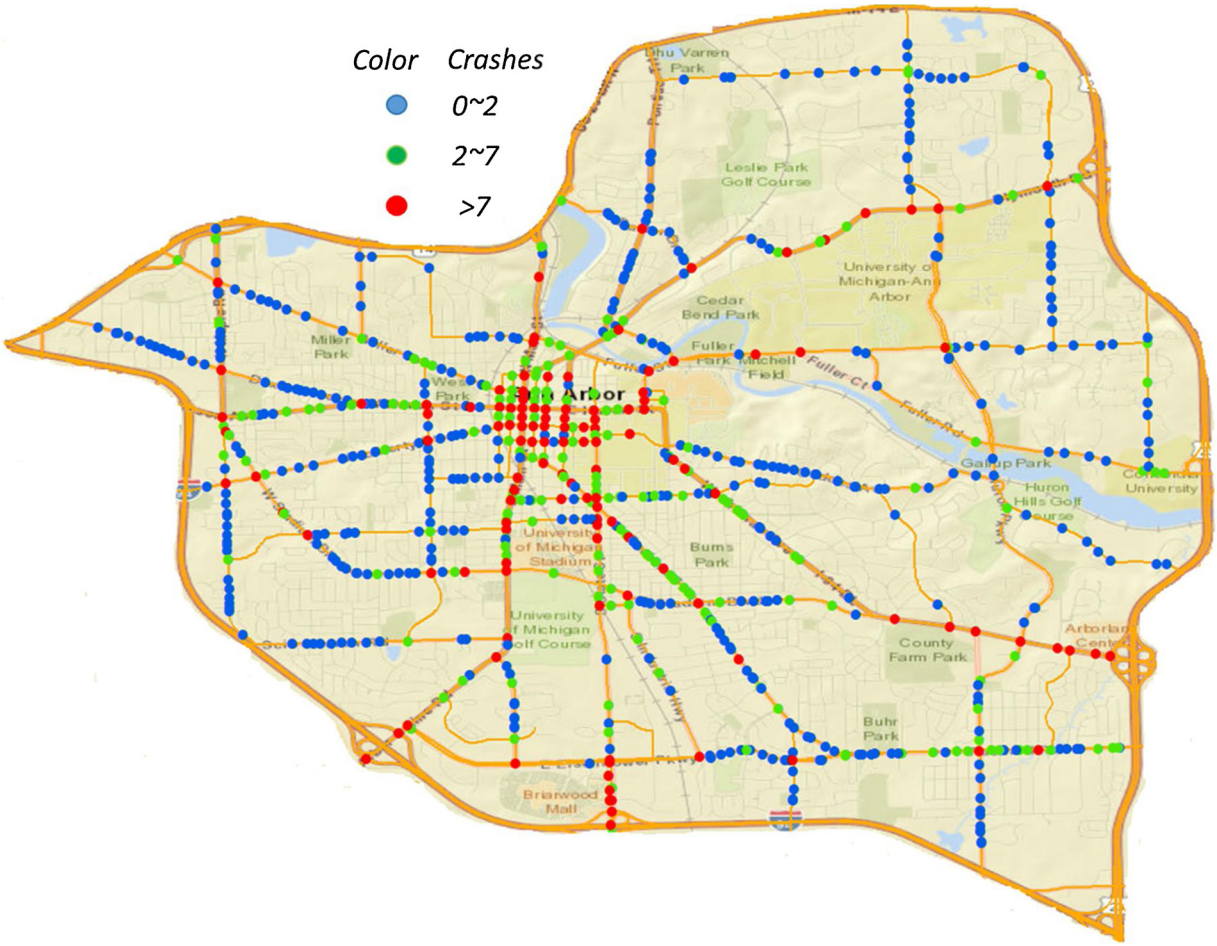

Fig. 3. Distribution map of risk levels at intersections, Ann Arbor, Michigan.

Table 2
Threshold for sudden acceleration/de-acceleration measured by vehicle speed intervals.

Vehicle speed (kph)	Acceleration (m/s^2)	Deceleration (m/s^2)
< 20	2.2	3.1
20 ~ 29	2.1	3.1
30 ~ 39	2.0	3.1
40 ~ 49	1.9	2.8
50 ~ 69	1.5	2.8
70 ~ 79	1.4	2.5
80 ~ 89	1.3	2.5
> 90	1.2	2.4

4. Deep learning models

4.1. Multi-layer perceptron (MLP) model and optimization of model structure

MLP network is a class of feedforward artificial neural network. An MLP consists of, at least, three layers of nodes: an input layer, a hidden layer, and an output layer, as shown in Fig. 4. The node can imitate the process of neuron in the human brain, which fires when it encounters sufficient stimuli. Therefore, except for the input nodes, each node is just a computational unit for data analysis and signal processing. A node combines input from the data with a set of coefficients (i.e. weights) that either amplify or dampen that input, thereby assigning significance to inputs with regard to the task the algorithm is trying to learn. These input-weight products are summed and passed through an activation function to determine whether and to what extent that signal should progress further through the network to affect the ultimate result. The characteristics of multiple-layer nodes and non-linear

Table 3
Descriptive statistics of features of intersection (n = 774).

Data source	Feature	Mean	SD	Min	Max
SEMCOG	AADT of major road (vehicle/day)	13,793	8817	301	44,718
	Number of approaches	3.31	0.53	2	6
	Number of major arterials	2.29	0.84	0	6
	Number of local roads	0.91	0.72	0	4
	Signalized (yes = 1)	0.19	0.39	0	1
CVs data	Speed-mean (m/s)	44.28	13.87	9.45	71.41
	Speed-Sdev (m/s)	11.48	4.83	1.61	26.35
	Yaw-mean (o/s)	0.06	1.49	−10.80	12.92
	Yaw- Sdev (o/s)	3.45	3.23	0.23	14.67
	YawPos-mean (o/s)	1.99	2.29	0.15	18.7
	YawPos- Sdev (o/s)	2.68	2.88	0.12	11.92
	YawNeg-mean (o/s)	−1.92	2.23	−14.72	−0.17
	YawNeg- Sdev (o/s)	2.53	2.68	0.14	17.26
	AccDec-mean (m/s2)	0.03	0.19	−0.65	0.77
	AccDec-Sdev (m/s2)	0.63	0.32	0.161	1.76
	Acc-mean (m/s2)	0.46	0.28	0.09	1.43
	Acc- Sdev (m/s2)	0.39	0.18	0.08	0.91
	Dec-mean (m/s2)	−0.49	0.30	−1.79	−0.1
	Dec- Sdev (m/s2)	0.46	0.22	0.08	1.10
	SudAccDec-sum	74	205	0	3689
	SudAcc-sum	66	193	0	3614
	SudDec-sum	8	20	0	195
	WaitTime-mean (s)	1.1	2.5	0	26
	WaitRatio-mean (%)	3	6	0	61

* The unit of analysis is the intersection.

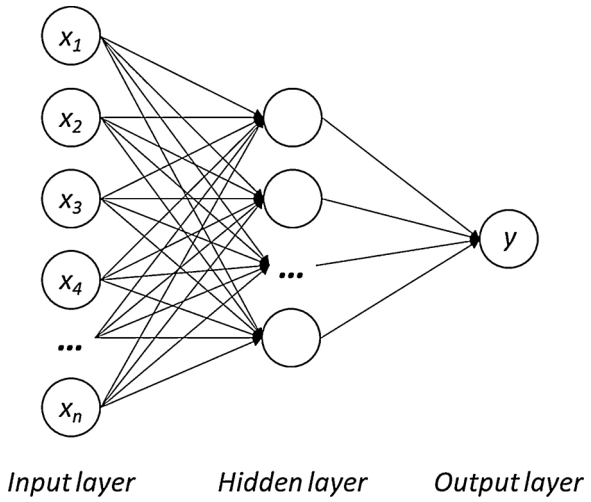

Fig. 4. MLP model with 1 hidden layer.

activation enable MLP to approximate complex continuous functions and solve many linearly inseparable problems.

The training process of MLP can be divided into two phases: forward and backward (Gardner and Dorling, 1998; Lee and Choeh, 2014). In the forward phase, the input values forward from the input layer through the hidden layers to the output layer to generate the predicted value(s). The error between predicted value(s) and ground truth can then be measured in a variety of ways, including the mean squared error (MSE) for regression problems and minimum error entropy (MEE) for classification problems. In the backward phase, the backpropagation algorithm and the chain rule of calculus are used. The partial derivatives of the error function are calculated. The various weights and biases are back-propagated through the networks. A gradient can be obtained from the differentiation using a gradient-based optimization algorithm such as stochastic gradient descent, along which the parameters can be adjusted as the error move closer to the minimum. The updated parameters can then be used to output a new prediction value in the forward phase. The network will keep repeating this process until a minimum is found or convergence criterion is met.

In this study, the MLP model includes 24 nodes in the input layer that provides inputs of all extracted features and 3 nodes in the output layer representing the predicted probability of each risk level. The risk level with the largest probability is the predicted risk level. The loss function, nonlinear transfer function and optimization function used in the MLP training are cross-entropy, rectified linear unit (ReLU) and Adam, respectively. The cosine annealing learning rate method that reduces the learning rate according to a half cosine curve is used. This allows for relatively larger weight changes at the beginning of the learning process and smaller changes or fine-tuning towards the end of the learning process, thereby tends to achieve better performance. The initial learning rate is set as 0.01. Due to the limited dataset, the model is trained and tested via the 5-fold cross-validation method. This method firstly randomly stratified partitions the data into five equal-sized subgroups. Of the 5 subgroups, a single subgroup is retained for testing, and the remaining 4 subgroups are used for training. This process is then repeated 5 times, with each of the subgroup used exactly once for model testing. As there are much more low-risk intersections than the numbers of medium and high risk intersections, the over-sampling technique is applied to the training data to generate a balanced dataset. The prediction results of each testing subgroup are obtained and the overall classification accuracy is calculated. The training iteration is set as 500 to ensure convergence. Pytorch is used to develop the proposed deep learning networks. It is a Python library that performs immediate execution of dynamic tensor computations with automatic differentiation and GPU acceleration. The model training is performed on a workstation with one Intel Core i7 − 8700 CPU and one NVIDIA GTX 1070 GPU.

The structure of the hidden layer (i.e. the number of layers and neurons) and dropout rate (percentage of neurons to be ignored during training to avoid over-fitting) are determined by Bayesian optimization (Brochu and Cora, 2010; Snoek and Larochelle, 2012). Bayesian optimization works by constructing a posterior distribution of functions (Gaussian process) that best describes the target function (i.e. the function of validation accuracy with different hyperparameter combinations in this study). With the increase of the number of hyperparameter combinations observed, the posterior distribution is improved and the algorithm becomes more certain of which hyperparameter combination should be picked for the next observation. This process is designed to minimize the number of steps required to find a combination of hyperparameters that are close to the optimal combination. The candidate set of each hyperparameter to be optimized is as follows:

- The number of hidden layers ∈ [2,3,4,5,6,7,8,9]
- The number of neurons in each hidden layer ∈ [16,32,64,128,256,512]
- Dropout rate ∈ [0,0.1,0.2,0.3,0.4,0.5]

There is a total of 288 hyperparameter combinations. Fig. 5 shows the observed highest accuracy over iterations. The accuracy is improved and converged at the 36th step. The optimum values of layer number, neurons number and dropout rate obtained are 2, 16 and 0.5, respectively.

4.2. CNN model and optimization of model structure

The architecture of a CNN is inspired by the organization of the visual cortex and is analogous to that of the connectivity pattern of neurons in the human brain. The visual cortex has small regions of cells that are sensitive to specific regions of the visual field. Some individual neuronal cells in the brain responded (or fired) only in the presence of edges of a certain orientation. For example, some neurons fire when exposed to vertical edges and some when shown horizontal or diagonal edges. All of these neurons are organized in a columnar architecture and that together, they are able to produce visual perception. This idea of specialized components is the basis behind CNN. Typically, CNN is very useful if input data exhibits certain patterns, for example, the ordering of image pixels in a grid and the temporal structures of an audio signal. Therefore, CNN is powerful in analyzing visual imagery such as image recognition, image classifications, and object detections, etc. (Krizhevsky and Sutskever, 2012; Kim, 2014). The sensor data from CVs can also be regarded as a one-dimensional image and the features

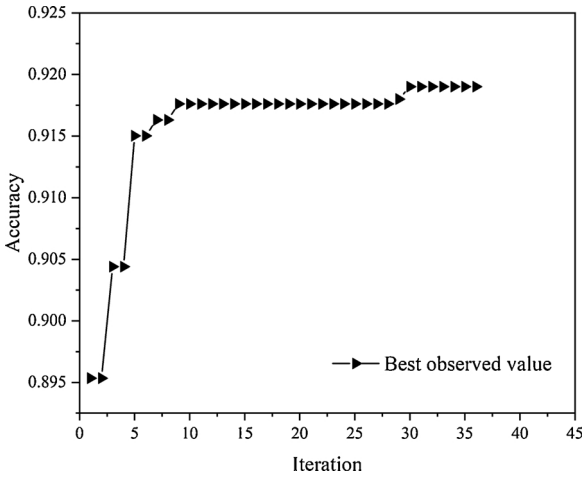

Fig. 5. Best observed results over iterations under MLP optimization.

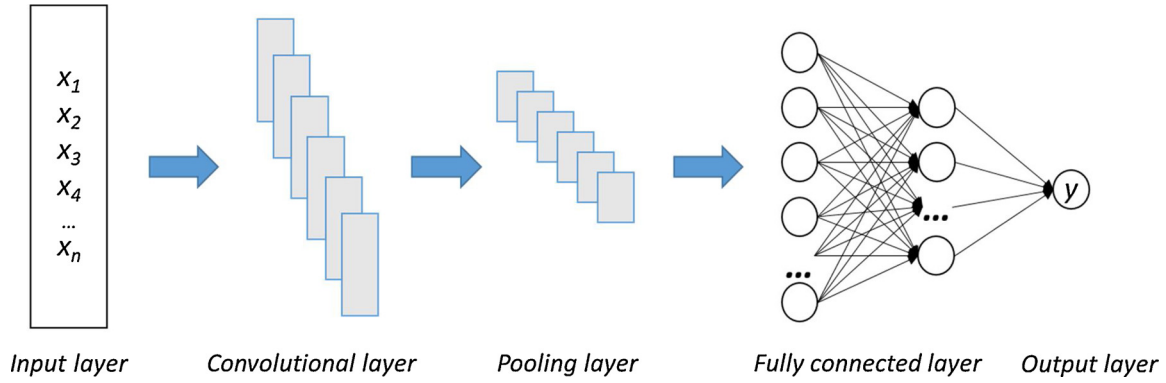

Fig. 6. A schematic overview of a CNN network.

hidden in the vehicle kinetic and traffic data can be extracted and learnt by CNN model, which can yield a good prediction. The CNN network for one-dimensional data is shown in Fig. 6.

Similar to MLP, CNN contains one input layer, multiple hidden layers, and one output layer. The hidden layers of a CNN typically consist of a series of convolutional layers, pooling layers and fully connected layers. A convolutional layer is able to successfully capture the spatial and temporal dependencies in an image through the convolution operations with relevant filters. The objective of the convolution operation is to extract high-level features such as edges, from the input image. Normally, CNN needs a series of convolutional layers. The first few convolutional layers are responsible for capturing the low-level features such as color, shape, gradient orientation, etc. The deeper convolutional layers can gradually extract high-level features. The last convolutional layer contains the richest information and a complete understanding of the input image. This mechanism is similar to how the human brain cognizes the images. Compared with the perceptron layer, the convolutional layer can better understand the sophistication of the input image, reduce the number of parameters involved and greatly accelerate the process. The pooling layer is responsible for reducing the spatial size of the convolved image feature. This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model. The fully connected layer (i.e. perceptron layer) can learn non-linear combinations of the high-level image features output from the last convolutional layer. The image features shall be flattened into a column vector and then fed to fully connected layers to generate an outcome. The process is the same as MLP.

The input layer, output layer, loss function, nonlinear transfer function, optimization function, learning rate and training process of the CNN model are similar as used for the MLP model. For convolutional layers, the hyperparameters include the layer number, filter number, kernel size and dilation rate. While for the fully connected layers, the hyperparameters include layer number, neuron number and dropout rate. All hyperparameters are selected by Bayesian optimization, as described in section 4.1. The candidate set of each hyperparameter to be optimized is as follows:

- Number of convolutional layers ∈ [2,3,4]
- Number of filters ∈ [32,64,128,256,512]
- Kernel size ∈ [3,4,5,6]
- Dilation rate ∈ [1,2,3,4,5]
- Number of fully connected layers ∈ [1,2,3]
- Number of neurons ∈ [256,512,1024]
- Dropout rate ∈ [0,0.1,0.2,0.3,0.4,0.5]

There are a total of 16,200 hyperparameter combinations. Fig. 7 shows the observed highest accuracy over iterations. The accuracy is improved and converged at the 37th step. The optimized model has 4 convolution layers and 2 fully connected layers. In each convolution layer, the number of filters, kernel size and dilation rate are 64, 3 and 3, respectively. In each fully connected layer, the number and dropout rate of neurons are 1024 and 0.1, respectively.

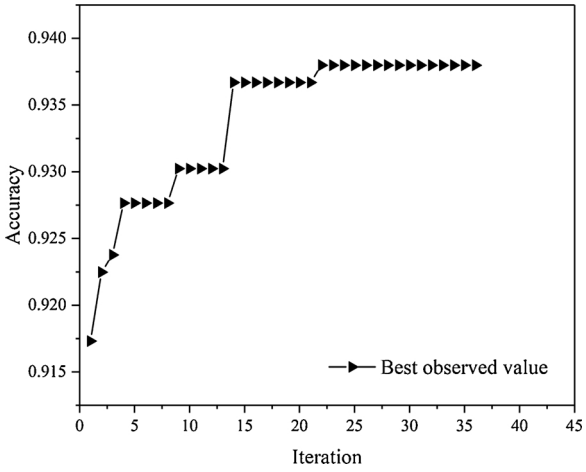

Fig. 7. Best observed results over iterations under CNN optimization.

5. Results of model prediction

Four metrics that are frequently used in classification studies are chosen to evaluate the performance of the machine learning models in classifying the risk level of intersections. These metrics include the *precision*, *recall*, F_1-*score* and *accuracy*, as defined below:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where *TP* is the true positive; *TN* is the true negative; *FP* is the false positive; *FN* is the false negative.

The *TP, TN, FP* and *FN* of each risk level are shown in Table 4 which contains prediction of all validation sets in the cross-validation process (as described in section 4.1). For example, in terms of the low-risk intersections predicted by MLP model, 394 of low-risk intersections are correctly classified (*TP*) and 80 of them are misclassified (*FN*); 43 of

Table 4
Model confusion matrix on validation sets by cross-validation method.

N = 774		Predicted								
		MLP			CNN			Decision Tree		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
Actual	Low	394	71	9	401	69	4	384	78	12
	Medium	41	111	31	48	107	28	74	70	39
	High	2	29	86	2	23	92	11	29	77

medium and high risk intersections are misclassified as low-risk (*FP*) and 257 of them are classified as either medium or high risk (*TN*). It is noted that some medium-risk intersections are misclassified and some low and high risk intersections are misclassified as medium-risk intersections. This may be caused by the threshold value of intersection risk level division. For example, two intersections with an annual crash rate difference of 0.2 may be classified into two different risk levels, even if they have similar features. Therefore, considering the fuzzy or ambiguity in the characteristics of risk level definition for intersections, a range of recorded crashes is set for each intersection category. The following false-positive samples are regarded as acceptable predictions: low-risk intersections with crash rate between 1 and 2 predicted as medium-risk, medium-risk intersections with crash rate between 2 and 3 predicted as low-risk, medium-risk intersections with crash rate between 5 and 7 predicted as high-risk, high-risk intersections with crash rate between 7 and 9 predicted as medium-risk.

The deep learning models are black box models, which means they can achieve satisfactory results from data but are lacking in terms of interpretability. Improving the interpretability of machine learning model is an active area of machine learning research. For comparison purposes, we also used the same data to train a Decision Tree model, which is a simple white-box type of machine learning model. Decision tree model is a flowchart-like tree structure where an internal node represents an input feature, the branch represents a decision rule, and each leaf node represents the outcome. It's visualization of a flowchart diagram, which is relatively easy to understand and interpret.

The validation metrics by different ML models are summarized in Table 5. The results indicate that the performances of the two deep learning models are much better than that of the decision tree model in all risk level predictions, and the performance of CNN model is slightly better than that of MLP model.

The annual average number of crashes between 2012 and 2014 were also collected as the test dataset to verify the reliability of the model. These independent crash data, which were not used in training the machine learning models, provided an unbiased evaluation of the final models. The testing metrics are shown in Table 6. The accuracies of the two deep learning models are both over 90 % while the accuracy of the decision tree model is 87 %, which shows that the deep learning models are capable of modeling the inherent complexity of the dataset and providing higher accuracy when compared with the traditional machine learning models. Meanwhile, the CNN model outperformed the MLP model in predicting the risks at different intersections. In terms of the CNN model, 98 % of the low-risk intersections are correctly detected with an accuracy of 94.9 %, 86.8 % of the medium-risk intersections are correctly detected with an accuracy of 91.5 %, and 90.2 % of the high-risk intersections are correctly detected with an accuracy of 92.4 %. Therefore, CNN model is recommended to predict the risk level for a specific intersection. Also, it was noticed that all three models perform better for the low-risk intersections. This is possibly due to the fact that most road intersections are low-risk intersections and there are relatively smaller amount of data samples from intersections with medium and high risks.

6. Interpretability of CNN model

Deep learning models are actually black-box models that do not explain their predictions in a way that humans can understand. Basically, you just need to give a set of inputs to black-box models and then get a set of corresponding output without seeing how the model works and understanding why a particular decision was made. Sometimes you do not even know if the model is reliable or not. The lack of transparency and accountability of predictive models has already had severe consequences for high-stakes applications such as healthcare and criminal justice (Varshney and Alemzadeh, 2017) (Rudin, 2019). Similar problems exist in the use of MLP and CNN models to predict crash risk. Therefore, the interpretability of the proposed black-box models is studied to understand why a model makes a certain prediction so as to better apply them in transportation safety field. The SHAP (SHapley Additive exPlanations) method is used to measure feature importance. SHAP is a unified framework for interpreting predictions especially for complex black-box models (Lundberg and Lee, 2017). It assigns each feature an importance value for a particular prediction indicating how much a model prediction relies on each feature, in other words, how much each feature contributes to the prediction. This is helpful to understand the impacts of input variables on prediction performance; the importance values by SHAP also helps to provide guidance in safety measures to reduce the crash rate at intersections. It is observed that the interpretability results of MLP and CNN by SHAP are similar and both of them can properly identify the important factors contributing to risk level decisions. Since the CNN model achieves higher accuracy, only the interpretability of the CNN model is discussed.

The output layer of CNN contains 3 nodes representing the predicted probability of each risk level. The one with the greatest probability will be the predicted risk level. As an example, Fig. 8 shows the predicted probability of each risk level by the trained CNN model for

Table 5
Model classification metrics on validation sets.

	MLP			CNN			Decision Tree		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Precision	0.955	0.860	0.833	0.967	0.898	0.874	0.902	0.846	0.735
Recall	0.987	0.842	0.769	0.992	0.869	0.829	0.975	0.689	0.759
F_1	0.968	0.851	0.800	0.979	0.883	0.851	0.735	0.709	0.722
Accuracy	0.916			0.938			0.867		

Table 6
Model classification metrics on independent test dataset from year 2012 to 2014.

	MLP			CNN			Decision Tree		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Precision	0.927	0.900	0.915	0.949	0.915	0.924	0.901	0.838	0.764
Recall	0.982	0.807	0.885	0.980	0.868	0.902	0.981	0.678	0.718
F_1	0.954	0.851	0.900	0.964	0.891	0.913	0.735	0.749	0.740
Accuracy	0.919			0.937			0.870		

intersection No. 4. The ground-truth annual crash rate at this intersection is 16 and therefore it's been labeled as high-risk intersection. As shown in the annotated output probability values in Fig. 8, the predicted probabilities for high (a), medium (b) and low (c) risk levels are 0.843, 0.154 and 0.004 respectively. The probability of high-risk level is the largest, so it becomes the predicted risk level for this intersection, which is consistent with the ground-truth. Fig. 8 also shows which features the model are relied upon to make such decisions (i.e. probability of each risk level). Each feature is assigned an importance value (SHAP value), as shown by the length of the color bar, to represent the contribution to the decision. The longer the bar, the larger the feature contributes to the predicted probability. The color represents the contribution manner with red for positive while blue for negative. The value under each bar represents the input value of the feature. In addition, each risk level starts with a base value (initial probability, i.e., 33 %) and then aggregate all feature SHAP values to get a final predicted probability value. As shown in Fig. 8(a), the important features such as signalized intersection, traffic flow, vehicle waiting time and number of major arterials increase the predicted probability of high-risk level. While in Fig. 8(b, c), these features decrease the predicted probability of medium and low risk levels, causing the probability of high-risk is far greater than that of medium and low risk. The rest of the features also contribute to the final prediction, but to a lesser importance. Consequently, the final prediction result is high-risk level.

In addition to facilitate the interpretation of model prediction at a particular intersection, SHAP is also capable of identifying the impact of each feature on overall model predicted risk at intersections (i.e. the positive and negative relationships of the features with the probability of different risk levels). Fig. 9 shows the SHAP values of all test datasets and average SHAP value of each feature for predicting the probability of risk levels. In Fig. 9(a)–(c), the features listed in the left are sorted by the average of SHAP value for all test samples with the magnitude listed in Fig. 9(d). In the row of each feature, the points indicate the corresponding SHAP values with the magnitude measured by horizontal ordinate and the color represents the input value of that individual (red high, blue low). For example, "signalized" is a categorized feature that includes 0 and 1. Therefore, the signalized intersections are colored as red while the non-signalized intersections are colored as blue. Similarly, "AADT" is a continuous feature with magnitude ranges from 301 to 44,718 and coded by color. Intersections with larger AADT are indicated in red, while these with smaller AADT are indicated in blue. Meanwhile, the points on the left side of midline make the predicted probability lower, while on the right side make the predicted probability higher. The farther away from the midline, the greater the impact on the predictions. It is noted that in Fig. 9(a), signalized intersection, AADT, number of major arterials, standard deviation of positive angular velocity, etc. have a positive impact on predicted probability of high-risk level. For the predicted probability of medium-risk level, as shown in Fig. 9(b), signalized intersection tends to have a negative impact on prediction, while other features have both positive

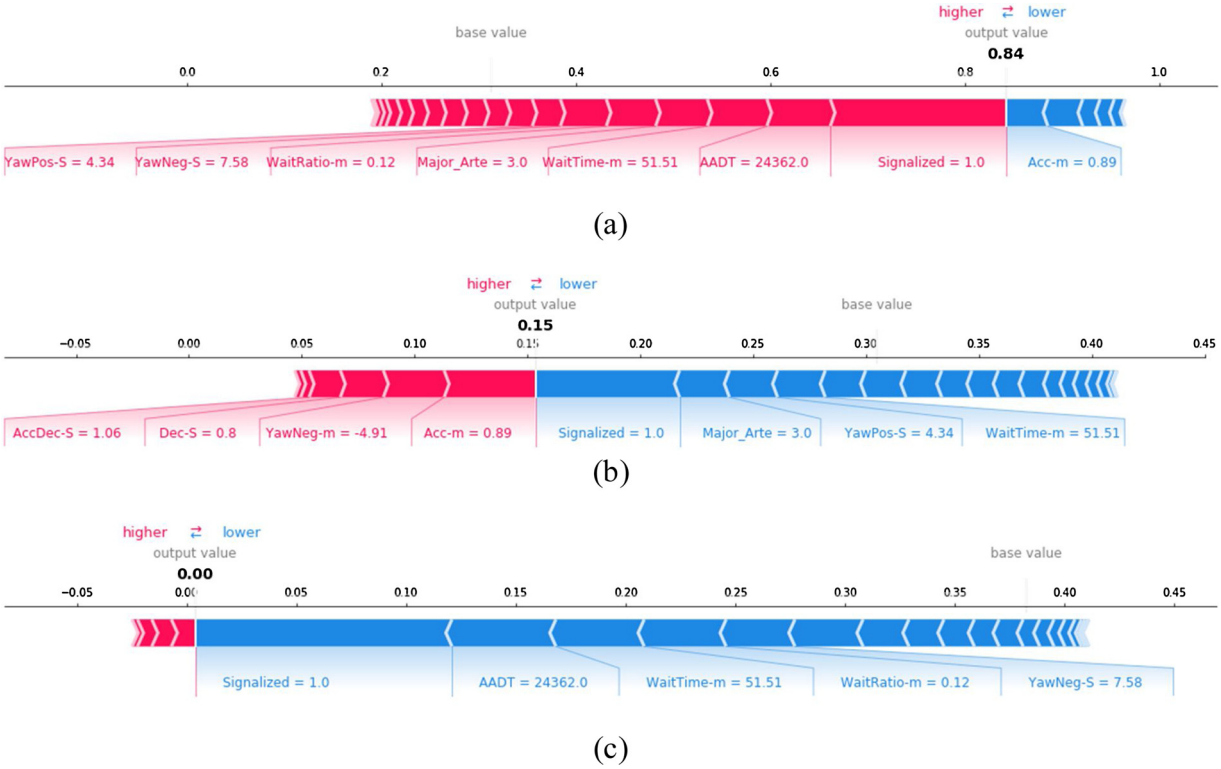

Fig. 8. SHAP values contributing to the predicted probability of different risk levels at intersection No. 4. (a) high-risk level (b) medium-risk level (c) low-risk level.

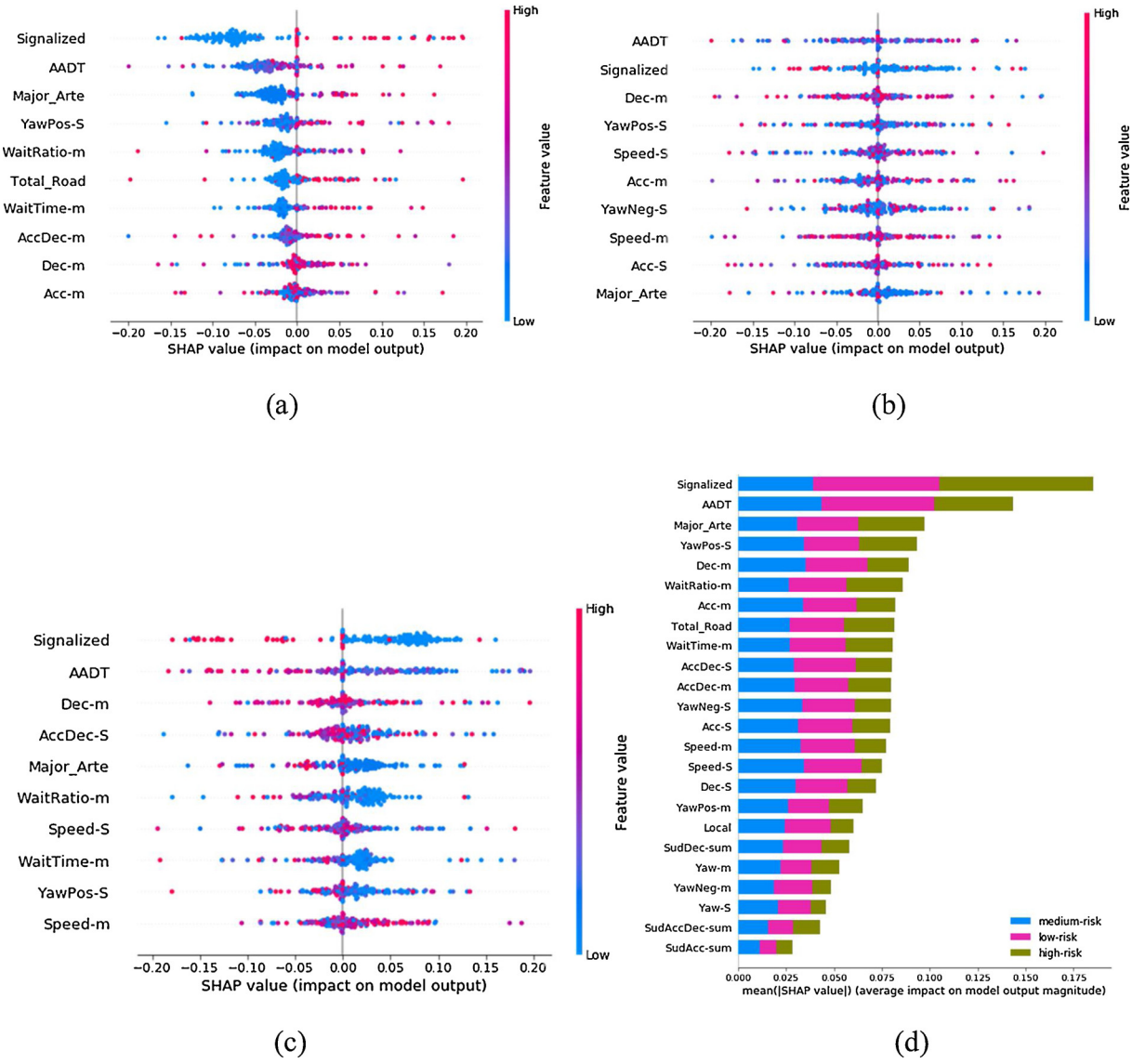

Fig. 9. Density scatter plot of SHAP values of the top-10 important features for the prediction of probability of (a) high-risk level (b) medium-risk level and (c) low-risk level. (d) Bar plot of the mean absolute value of the SHAP values for each feature.

and negative impact on predictions. That's also explained why some medium-risk intersections are misclassified as low-risk or high-risk intersections. In Fig. 9(c), signalized intersection, AADT, number of major arterials, standard deviation of acceleration/deceleration, etc. have a negative impact on predicted probability of low-risk level. Overall, the traffic signal deployment, larger traffic flow, larger number of major arterials, frequent vehicle turns and acceleration/deceleration tend to increase the risk level of intersections. Also in Fig. 9(d), it is noted that features such as the average AADT, signalized intersection, number of local roads and major arterials, vehicle acceleration changes and average speed play an important role in predicting probability of each risk level. The findings are reasonable and consistent with the statistical results of crashes (Miaou and Lord, 2003; Kamrani and Arvin, 2018) and the common observations, i.e., crashes tend to occur at the intersection with multiple major traffic arterials and the frequent acceleration or deceleration of vehicles are the signs of high-risk intersections. These heuristic observations verify that the proposed CNN model properly identifies the important factors contributing to crashes at intersections.

The SHAP values can also provide guidance for countermeasures to mitigate risks. Based on the importance of features, the countermeasures can be taken from two aspects: road design (signal, branches, etc.) and driver behaviors (driving speed, car following distance, etc.). As an example of a high-risk intersection, if the SHAP values of AADT and major arterial number are dominant, strategies could be implemented to reduce peak traffic volume; if the SHAP values of vehicle waiting time/ratio for traffic light are dominant, optimization of traffic signal might be efficient in crash risk mitigation; if the SHAP value of the vehicle angular velocity change is dominant, setting stop signs and left-turn protection should be more effective. If the SHAP values of vehicle speed and mean and standard deviation of acceleration/deceleration are dominant, setting proper speed limits and dynamic warning signs should be more effective in crash risk mitigation.

7. Conclusion

This paper aims to use connected vehicles and deep-learning technology to identify crash risky intersections and guide deployment of safety measures. With CVs technology, sufficient amount of data can be collected over a short period of time even under a low market penetration rate of CVs (this study is based on one-month data by CVs which accounts for between 3 %–12 % overall trips at the intersections). It is

achievable in the transportation network in the near future and will significantly accelerate the time for data collection of intersection safety study. The CVs data from USDOT SPDM program and historical traffic data from Michigan DOT are used to train and evaluate MLP and CNN models to classify the risk levels at the intersections. Bayesian optimization algorithm is used to optimize the model hyperparameters. The results show that the deep learning models, which are black-box model, achieved higher performance in modeling the inherent complexity of the dataset and provide higher accuracy compared with the traditional white-box type of machine learning model such as the decision tree model. Both deep learning models achieve the prediction accuracy of over 90 % with CNN model achieves slightly better performance (93.8 %). The interpretability of the CNN model is studied by SHAP method indicating that CNN can properly identify the important features contributing to risk level decisions such as signal light, traffic flow and vehicle start/brake frequency. The findings are reasonable and consistent with the statistical results of crashes, which verifies the reliability of the model.

The method proposed by this study can be used to deploy effective safety measures at intersections. The ability of rapid dynamic mapping high-risk intersections on the whole traffic network with CVs is advantageous over the current practice of safety studies in terms of the speed of safety assessment and the complete coverage of the whole road network. Network level mapping of the risk at intersections is important for traffic safety practitioners to prioritize intersections in deploying countermeasures. With such information, proper short term or long term countermeasures can be deployed to mitigate the risks of crashes at the intersections. Meanwhile, based on the importance of features contributing to crashing risk at a specific intersection, countermeasures can be adapted to the characteristics of a particular intersection. This will lead to the most cost-effective countermeasures to effectively mitigate crashes. Such measures might include adjusting the traffic signals, setting proper speed limits, or measures to reduce peak traffic volume, etc. In addition, with the projected rapid growth of CVs penetration into the transportation system, the information of high-risk intersections can be shared with the surrounding vehicles via DSRC or 5 G technology to assist the drivers (human operated or autonomous vehicles alike) to institute preventative driving operations to avoid potential crashes. Therefore, the proposed technology can potentially significantly improve the transportation safety.

The future work mainly includes two parts. Firstly, integrate more crash-related features into the current prediction model, such as AADT of minor road, pavement condition and weather condition to further improve performance. Secondly, expand the prediction of risk level to other types of road sections such as straight road, curve, on-ramp, off-ramp.

CRediT authorship contribution statement

Jiajie Hu: Conceptualization, Data curation, Formal analysis, Writing - original draft. **Ming-Chun Huang:** Writing - review & editing, Supervision. **Xiong Yu:** Investigation, Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is partially support by the US National Science Foundation (Grant No. 1638320). We also appreciate anonymous reviewers for their constructive comments which helped to improve this manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2020.105665.

References

Archer, J., 2005. Indicators for Traffic Safety Assessment and Prediction and their Application in Micro-Simulation Modelling: a Study of Urban and Suburban Intersections. KTH.
Bao, J., Liu, P., et al., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accid. Anal. Prev. 122, 239–254.
Brochu, E., Cora, V.M., et al., 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint arXiv:1012.2599.
Chen, J., Wu, Z., et al., 2019. Driving safety risk prediction using cost-sensitive with nonnegativity-constrained autoencoders based on imbalanced naturalistic driving data. IEEE Trans. Intell. Transp. Syst.
Collobert, R., Weston, J., 2008. A unified architecture for natural language processing: deep neural networks with multitask learning. Proceedings of the 25th International Conference on Machine Learning, ACM.
Collobert, R., Weston, J., et al., 2011. Natural language processing (almost) from scratch. J. Mach. Learn. Res. 12 (August), 2493–2537.
Committee, D., 2009. Dedicated short range communications (DSRC) message set dictionary. SAE Standard J 2735, 2015.
Cooper, D.F., Ferguson, N., 1976. Traffic studies at T-Junctions. 2. A conflict simulation Record. Traffic Engineering & Control 17 (Analytic).
Cunto, F., Saccomanno, F.F., 2008. Calibration and validation of simulated vehicle safety performance at signalized intersections. Accid. Anal. Prev. 40 (3), 1171–1179.
Davis, G.A., Hourdos, J., et al., 2011. Outline for a causal model of traffic conflicts and crashes. Accid. Anal. Prev. 43 (6), 1907–1919.
Gao, J., Ozbay, K., et al., 2018. A Life-Cycle Cost-analysis Approach for Emerging Intelligent Transportation Systems with Connected and Autonomous Vehicles.
Gardner, M.W., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32 (14-15), 2627–2636.
Goodfellow, I., Bengio, Y., et al., 2016. Deep Learning. MIT press, Cambridge.
Graves, A., Mohamed, A.-r., et al., 2013. Speech recognition with deep recurrent neural networks. Acoustics, speech and signal processing (icassp). 2013 IEEE International Conference on, IEEE.
Guido, G., Saccomanno, F., et al., 2010. Comparing safety performance measures obtained from video capture data. J. Transp. Eng. 137 (7), 481–491.
Han, I., Yang, K., 2009. Characteristic analysis for cognition of dangerous driving using automobile black boxes. Int. J. Automot. Technol. 10 (5), 597–605.
Hinton, G., Deng, L., et al., 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Mag. 29 (6), 82–97.
Hogema, J., Janssen, W., 1996. Effects of Intelligent Cruise Control on Driving Behaviour: a Simulator Study.
Horst, R., 1991. Time-to-collision as a cue for decision-making in braking. Vision in Vehicles–III.
Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36.
Kamrani, M., Arvin, R., et al., 2018. Extracting Useful Information From Connected Vehicle Data: an Empirical Study of Driving Volatility Measures and Crash Frequency at Intersections.
Kim, Y., 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.
Kim, E., Choi, E., 2013. Estimates of critical values of aggressive acceleration from a viewpoint of fuel consumption and emissions. 2013 Transportation Research Board Annual Meeting.
Krizhevsky, A., Sutskever, I., et al., 2012. Imagenet Classification with Deep Convolutional Neural networksAdvances in Neural Information Processing Systems.
Kuang, Y., Qu, X., et al., 2015. A tree-structured crash surrogate measure for freeways. Accid. Anal. Prev. 77, 137–148.
Laureshyn, A., Johnsson, C., et al., 2016. Review of Current Study Methods for VRU Safety. Appendix 6–Scoping Review: Surrogate Measures of Safety in Site-based Road Traffic Observations: Deliverable 2.1–part 4.
LeCun, Y., Bengio, Y., et al., 2015. Deep learning. Nature 521 (7553), 436.
Lee, S., Choeh, J.Y., 2014. Predicting the helpfulness of online reviews using multilayer perceptron neural networks. Expert Syst. Appl. 41 (6), 3041–3046.
Lu, N., Cheng, N., et al., 2014. Connected vehicles: solutions and challenges. Ieee Internet Things J. 1 (4), 289–299.
Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems.
Lv, Y., Duan, Y., et al., 2015. Traffic flow prediction with big data: a deep learning approach. IEEE trans. Intell. Transp. Syst. 16 (2), 865–873.
Ma, X., Yu, H., et al., 2015. Large-scale transportation network congestion evolution prediction using deep learning theory. PLoS One 10 (3), e0119044.
Manning, C., Surdeanu, M., et al., 2014. The Stanford CoreNLP natural language processing toolkit. Proceedings of 52nd Annual Meeting of the Association for

Computational Linguistics: System Demonstrations.
Miaou, S.-P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Trans. Res. Record (1840), 31–40.
Minderhoud, M.M., Bovy, P.H., 2001. Extended time-to-collision measures for road traffic safety assessment. Accid. Anal. Prev. 33 (1), 89–97.
Officials, T., 2011. A Policy on Geometric Design of Highways and Streets, 2011. AASHTO.
Ozbay, K., Yang, H., et al., 2008. Derivation and validation of new simulation-based surrogate safety measure. Trans. Res. Record (2083), 105–113.
Pearson, R.K., Neuvo, Y., et al., 2016. Generalized hampel filters. EURASIP J. Adv. Signal Process. 2016 (1), 87.
Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. Transp. Res. Part C Emerg. Technol. 79, 1–17.
Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Machine Int. 1 (5), 206–215.
Snoek, J., Larochelle, H., et al., 2012. Practical bayesian optimization of machine learning algorithms. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2. Lake Tahoe, Nevada, Curran Associates Inc. 2951–2959.
Szegedy, C., Vanhoucke, V., et al., 2016. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
Theofilatos, A., Chen, C., et al., 2019. Comparing machine learning and deep learning methods for real-time crash prediction. Transp. Res. Rec. 2673 (8), 169–178.
van der Horst, R., Hogema, J., 1993. Time-to-Collision and Collision Avoidance Systems. na.. .
Varshney, K.R., Alemzadeh, H., 2017. On the safety of machine learning: cyber-physical systems, decision sciences, and data products. Big Data 5 (3), 246–255.
Wright, J., Garrett, J.K., et al., 2014. National Connected Vehicle Field Infrastructure Footprint Analysis, United States. Department of Transportation. Intelligent Transportation ….
Xie, K., Yang, D., et al., 2018. Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. Accid. Anal. Prev.
Yang, H., 2012. Simulation-based Evaluation of Traffic Safety Performance Using Surrogate Safety Measures. Rutgers University-Graduate School-New Brunswick.
Yuan, J., Abdel-Aty, M., et al., 2019. Real-time crash risk prediction using long short-term memory recurrent neural network. Transp. Res. Rec. 2673 (4), 314–326.
Zhang, Z., Li, M., et al., 2019. Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. Transp. Res. Part C Emerg. Technol. 105, 297–322.
Zhang, L., Yang, F., et al., 2016. Road crack detection using deep convolutional neural network. Image processing (ICIP). 2016 IEEE International Conference on, IEEE.
Zhong, S., Hu, J., et al., 2020. A deep neural network combined with molecular fingerprints (DNN-MF) to develop predictive models for hydroxyl radical rate constants of water contaminants. J. Hazard. Mater. 383, 121141.