Safety Science 146 (2022) 105522

Contents lists available at ScienceDirect

Safety Science

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

Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework

Zekun Yang [a], Wenping Zhang [b,*], Juan Feng [c]

[a] School of Information Resource Management, Renmin University of China, Beijing, China
[b] School of Information, Renmin University of China, Beijing, China
[c] Department of Management Science & Engineering, School of Economics & Management and Shenzhen International Graduate School, Tsinghua University, China

ARTICLE INFO

Keywords:
Deep learning
Traffic accident severity prediction
Explainable artificial intelligence
Machine learning

ABSTRACT

Predicting traffic accident severity is essential for traffic accident prevention and vulnerable road user safety. Furthermore, the explainability of the prediction is crucial for practitioners to extract relevant risk factors and implement corresponding countermeasures. Most extant research ignores the property loss severity of traffic accidents and fails to predict different levels of death and property loss severity. Moreover, while the explainability of traditional models is easy to achieve, an explainable design of deep neural network (DNN) is extremely deficient in existing research. Few attempts that incorporate neural networks suffer from the lack of multiple hidden layers and the negligence of structural information when explaining predictions. In this study, we propose a multi-task DNN framework for predicting different levels of injury, death, and property loss severity. The multi-task and deep learning design enables a comprehensive and precise analysis of traffic accident severity. Unlike many black-box DNN algorithms, our framework could identify key factors that cause the three types of traffic accident severity via layer-wise relevance propagation, which generates explanations based on the structure and weights of DNN. Based on the experiments conducted using Chinese traffic accident data, our proposed model predicts traffic accident severity risks with good accuracy and outperforms state-of-the-art methods. Furthermore, the case studies show that the key factors provided by our framework are more reasonable and informative than the explanations provided by baseline methods. Our model is the first multi-task learning model and the first DNN-based model for traffic accident severity prediction to the best of our knowledge.

1. Introduction

Traffic accident has become a deeply concerning issue around the world, as it causes significant injury, death, and property loss. The number of road traffic deaths is around 1.4 million in 2016, and the burden of road traffic injuries and deaths is disproportionately borne by vulnerable road users (VRUs) globally (World Health Organization, 2018). In fact, more than half of all road traffic deaths are among VRUs, such as pedestrians, cyclists, and motorcyclists (World Health Organization, 2018). In China, according to the National Bureau of Statistics, 62,763 people died in traffic accidents, and the direct property loss in traffic accidents is around 1.3 billion yuan in 2019 (National Bureau of Statistics of China, 2020). Traffic accident severity refers to the degree of injury, death, and property loss caused by the accident. Predicting the severity of traffic accidents is important for the transportation departments of governments to design traffic safety policies, as it could reduce the injury, death, and property loss of people and prevent traffic accidents (Tambouratzis et al., 2014; Zhu et al., 2019). In addition, discovering critical factors that influence traffic accident severity is essential for choosing safety countermeasures and strategies to mitigate the severity of traffic crashes (Arteaga et al., 2020).

Due to the significance of predicting and identifying key factors of traffic accident severity, various models have been constructed by previous research. Prior traffic accident severity prediction models are generally categorized into discrete outcome models, data mining models, and soft computing models (Mujalli and de Oña, 2013). Factors used by traffic accident severity prediction models include driver features, vehicle features, road features, and environmental features. However, extant studies fail to predict and analyze injury severity, death severity, and property loss severity comprehensively. While most extant research predicts injury/death severity, they ignore the degree of property loss severity. Furthermore, previous research is able to predict

* Corresponding author.
E-mail addresses: zekunyang@outlook.com (Z. Yang), wpzhang@ruc.edu.cn (W. Zhang), fengjuan@sem.tsinghua.edu.cn (J. Feng).

https://doi.org/10.1016/j.ssci.2021.105522
Received 14 April 2021; Received in revised form 20 June 2021; Accepted 24 September 2021
Available online 13 October 2021

whether anyone is dead or whether there is property loss in the accident. However, it fails to predict different levels of death and property loss. When analyzing causes of traffic accident severity, factors that affect property loss severity have been rarely investigated by previous research. In addition, extant traffic accident severity prediction models only use single-task designs, in which each type of traffic accident severity is predicted independently. Since the three types of traffic accident severity are closely related to each other, the prediction accuracy would decrease if information sharing is not allowed across different traffic accident prediction tasks during the training and prediction process. Moreover, the explainability of a prediction model is essential for practitioners to understand the model and extract relevant factors. Previous single-layer neural network models for traffic accident severity apply sensitivity analysis and Local Interpretable Model-Agnostic Explanations (LIME) to explain the prediction outcome. Yet these methods are not specifically designed for neural networks and do not utilize the structure and weights of neural networks to generate explanations, which might lead to deviation and error of explanations, especially when explaining neural networks with multiple layers.

To remedy the deficiencies of previous research, we propose a multi-task deep neural network (DNN) framework that predicts various levels of injury severity, death severity, and property loss severity of traffic accidents. The deep learning structure and multi-task design of our model ensure information sharing among three different DNNs during the training process, which enhances the prediction and explanation quality of the three types of traffic accident severity. To identify which factors influence the three types of traffic accident severity the most, we apply layer-wise relevance propagation (LRP), an explanation technique explicitly designed for DNNs which utilizes the structural information of DNNs. Experiments on Chinese traffic accident data demonstrate that our model achieves outstanding prediction accuracy and explainability and constantly outperforms state-of-the-art baseline models. More specifically, our contributions are as follows:

- We create the first DNN-based model and the first multi-task learning model for traffic accident severity prediction to the best of our knowledge. DNNs have proven to be superior to single-layer neural network designs. Unlike single-task learning, multi-task learning leverages domain-specific information from related tasks, which enables a comprehensive and precise analysis of traffic accident severity.
- Our proposed framework identifies key factors that cause traffic accident severities via LRP, which generates explanations based on the structure and weight of DNN. Based on these key factors, public policies could be implemented to promote VRUs safety.
- Our multi-task DNN framework generates novel prediction outcomes and key factors of traffic accident severity. Our framework outperforms all state-of-the-art baseline methods in terms of prediction accuracy. The results of case studies demonstrate that the key factors provided by our framework are more reasonable and informative than the explanations of the baseline methods.

The rest of the paper is organized as follows. Section 2 reviews the literature on the factors influencing traffic accident severity, traffic accident severity prediction models, and the explainability of traffic accident severity prediction. The structure and explainability of our proposed multi-task deep learning model as well as the traffic accident dataset used in this research are illustrated in Section 3. Section 4 demonstrates the results of the quantitative experiment on prediction quality and the case studies on the explainability of our proposed model. Finally, Section 5 draws the conclusion of this study and discusses future research directions.

2. Literature review

In this section, we first summarize the literature on factors related to traffic accident severity. The key factors help models generate accurate traffic accident severity predictions and assist the design process of public policies. Next, we review extant designs of traffic accident severity prediction models. Finally, we discuss existing explanation methods for identifying key factors that contribute the most to the model's prediction outcome.

2.1. Factors related to traffic accident severity

Identifying the related factors is a precondition for the prediction of traffic accident severity, as these factors serve as the input of traffic accident severity prediction models. Meanwhile, discovering factors that affect traffic accident severity could help practitioners to design and implement public policies, which ensure traffic safety and reduce the injury, death, and property loss of VRUs. There are four general categories of factors related to traffic accident severity: driver factor, vehicle factor, road factor, and environmental factor.

Many driver characteristics could affect accident severity, such as gender and age. Males are linked to a higher probability of fatal injury compared with females (Kim et al., 2013; Monárrez-Espino et al., 2006; Zhang et al., 2013). Age also influences accident severity. Previous research shows that there is an increase in fatality likelihoods for older males, and the possibility of fatality for young and middle-aged male drivers could go up (Islam and Mannering, 2006). Drinking alcohol before and during driving could adversely affect the degree of accident injury (Kasantikul et al., 2005; Soderstrom et al., 2005). Furthermore, drug abuse such as taking benzodiazepines and cocaine also worsens the injury severity (Beirness et al., 2006; Soderstrom et al., 2005). The experience of the driver is a significant factor when analyzing traffic accident severity. Generally, the more inexperienced the driver is, the more likely he/she will encounter severe traffic accidents (Rosenberg and Martinez, 1996). In extreme case, a driver without a driving license has a higher likelihood to encounter a fatal injury (Kraus et al., 1991). Lastly, disregarding the rules of safe riding would increase the probability of fatal injury (Rutter and Quine, 1996). For example, safety belts have proven to be effective in preventing fatalities (Evans, 1986). Therefore, not using a safety belt while driving would increase the likelihood of fatal injury.

Besides driver factors, vehicle factors also contribute to accident severity. Malfunctions of the vehicle, such as tire-related malfunctions, would increase the likelihood of severe injuries in crashes (Islam, 2015). The age of the vehicle is also associated with the severity of the accident, with older vehicles demonstrating higher risks of fatal or severe injuries (Yau et al., 2006). Moreover, the size of the vehicle matters. Larger vehicles have a higher chance to result in a deadly crash (Zajac and Ivan, 2003).

Road conditions influence traffic accident severity. A road with more facilities is usually less likely to have severe accidents. Traffic accidents happening at road intersections are more likely to be severe or fatal than accidents happening in other sections of the road (Moore et al., 2011). Furthermore, the presence of traffic signals reduces the probability of fatal injuries, as it reduces the incidences of side-impact crashes (Rifaat et al., 2011). The width of the road is also an essential factor. Wider roads have less occurrence of severe and fatal accidents (Zajac and Ivan, 2003). Roads equipped with speed camera networks are less likely to have severe accidents than roads that do not have speed camera networks (Christie et al., 2003).

Environmental factors contribute to traffic accident severity as well. Daylight is a significant factor for injury severity. Crashes happening in the day are less injurious than crashes occurring at night, as the visibility of the driver would decrease when there is no daylight (Behnood and Al-Bdairi, 2020; Lee and Abdel-Aty, 2005). Weather conditions also strongly affect injury severity. Precipitation such as rain and snow would adversely affect injury severity (Edwards, 1998). Accidents happening on foggy days have remarkably higher injury and fatality rates (Al-Ghamdi, 2007).

Despite previous research has discovered many factors that influence traffic accident severity, these factors are only proven to be related to injury and death severity. Factors that affect the value of property loss in a traffic accident have been little studied. Since injury, death, and property loss occur simultaneously in a traffic accident, it is necessary to analyze factors related to all three types of traffic accident severity. Only in this way can we conduct a more comprehensive and precise analysis of traffic accident severity. In addition, since the risk factors of injury, death, and property loss are interrelated, identifying the risk factors of each severity independently would ignore the correlation between the three traffic accident severities and neglect some shared factors between them.

2.2. Traffic accident severity prediction models

Traffic accident severity prediction models aim at accurately predicting the severity level of traffic accidents and discovering factors that affect traffic accident severity outcomes (Wahab and Jiang, 2019). Traffic accident severity prediction models help to prevent traffic accidents, reduce accident rates, and mitigate injury, death, and property loss of VRUs (Tambouratzis et al., 2014; Zhu et al., 2019). Extant research on predicting traffic accident severity mainly applies three groups of techniques: discrete outcome models, data mining techniques, and soft computing techniques (Mujalli and de Oña, 2013).

Discrete-outcome models represent probabilities of outcomes based on certain factors or characteristics (Mujalli and de Oña, 2013). The most widely used discrete-outcome model is the logit model (Kononen et al., 2011; Kwon et al., 2015; Wang and Kockelman, 2005). Logit model utilizes a logistic function to model a binary dependent variable, and the model could be extended to the multinomial case where the dependent variable has three or more outcomes. Kononen et al. (2011) develop a multivariate logistic regression model based upon National Automotive Sampling System Crashworthiness Data System for predicting the probability that a crash-involved vehicle will contain one or more occupants with serious or incapacitating injuries. Another popular discrete-outcome model is the probit model (Xie et al., 2009; Zhu and Srinivasan, 2011). The major difference between logit models and probit models is that the residual of the logit model follows a logistics distribution, but the residual of the probit model follows a normal distribution. Zhu and Srinivasan (2011) use the ordered probit model to predict four categories of injury/death severity of large-truck crashes. The factors used by the model include characteristics of the crash, the vehicle, and the driver.

Data mining is the process of discovering patterns in large and complex data (Hand et al., 2001). Decision tree is a data mining model that uses a tree structure to partition the data into different categories recursively (Abellán et al., 2013; Chang and Chien, 2013; Kwon et al., 2015; López et al., 2012; Moral-García et al., 2019; Tambouratzis et al., 2014; Wahab and Jiang, 2019). Moral-García et al. (2019) construct a decision tree ensemble model called information root node variation for predicting whether accidents that involve inexperienced drivers in urban areas are fatal. Another data mining technique is the Bayesian network (de Oña et al., 2011; Halbersberg and Lerner, 2019; Kwon et al., 2015; Mujalli et al., 2016; van Wyk et al., 2019; Zhu et al., 2019). van Wyk et al. (2019) create a Bayesian network of the driver and autonomous networks for predicting two categories of consequences for accidents: property damage and injury/fatality.

Soft computing techniques exploit the tolerance for imprecision and uncertainty to achieve manageability, robustness, and solutions at a low cost (Marks, 1994). Neural network is a soft computing technique that consists of a series of connected units called artificial neurons, and it has become increasingly popular over the years due to its excellent performance in discovering latent patterns (Arteaga et al., 2020; Assi et al., 2020; Delen et al., 2006; Mussone et al., 2017). While discrete-outcome models are often negatively influenced by missing values and outliers in the dataset, neural networks are non-parametric tools that are good at handling outliers and missing values (Wahab and Jiang, 2019). Mussone et al. (2017) apply a single-layer back-propagation neural network for predicting injury/death severity levels. Arteaga et al. (2020) design a single-layer perceptron neural network to classify injury/death severity of traffic accidents.

Although previous models generate nice predictions and explanation outcomes for traffic accident severity, some deficiencies exist. First, a multifaceted risk profiling of traffic accident severity is missing. While three types of traffic accident severity need to be predicted, most research considers only injury/death severity when modeling traffic accident severity and does not take the value of property loss into account. Reducing property loss is as important as reducing the number of injuries and deaths in traffic accidents, and VRUs are also concerned about property loss. Therefore, in order to demonstrate multi-aspect risk sources of traffic accident severity comprehensively, it is important to cover property loss in traffic accident severity prediction models. Furthermore, previous research fails to predict different levels of death severity and property loss severity. Injury severity has been classified into different levels such as no injury, complaint of pain, other visible injury, severe injury, and fatal injury (Kwon et al., 2015). However, previous research only predicts death severity and property loss as binary variables, i.e., with or without death or property loss. While death severity and property loss severity could be measured by the number of deaths and the value of property loss respectively, these different levels are not studied by previous traffic accident severity models. Second, extant traffic accident severity prediction models only use single-task designs, in which each type of traffic accident severity is predicted independently. However, the three types of traffic accident severity are closely associated with each other. For example, the value of property loss in a traffic accident often depends on the number of injured and deaths in the accident, and the number of injured is closely related to the number of deaths in the accident. Without sharing the information of other severity prediction tasks during the training and prediction process of a traffic accident severity model, the prediction accuracy would be reduced. Third, a deep design of neural network model for traffic accident severity prediction is missing. DNNs with more than three layers have proven to be superior in many tasks to simple neural networks with one or two layers (Sze et al., 2017). While previous research constructs neural networks for predicting traffic accident severity, all of them are single-layer neural networks (Arteaga et al., 2020; Assi et al., 2020; Delen et al., 2006; Mussone et al., 2017).

2.3. Explainability of traffic accident severity prediction

The explainability of traffic accident severity prediction models is essential for practitioners to understand the decision process of the model, extract key features that influence traffic accident severity, and design policies to facilitate VRU safety.

For discrete-outcome models and data mining techniques, explainability is relatively easy to achieve as most of these models are interpretable by nature. As examples, the logit model and the probit model are directly explainable, because the coefficients of the features demonstrate how much the feature contributes to the prediction. For decision trees like the random forest, feature importance could be calculated from the model, and decision rules could be extracted from the model to discover factors related to the severity of accidents (Moral-García et al., 2019). As for Bayesian network, probability inference analyses could be performed on the network to reveal variables that affect traffic accident severity.

Although neural networks have recently become popular for traffic accident prediction, they are not explainable by nature, and external explanation methods need to be applied to explain the prediction outcome neural networks. Mussone et al. (2017) use sensitivity analysis to explain their single-layer back-propagation neural network. Sensitivity analysis discovers how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model's

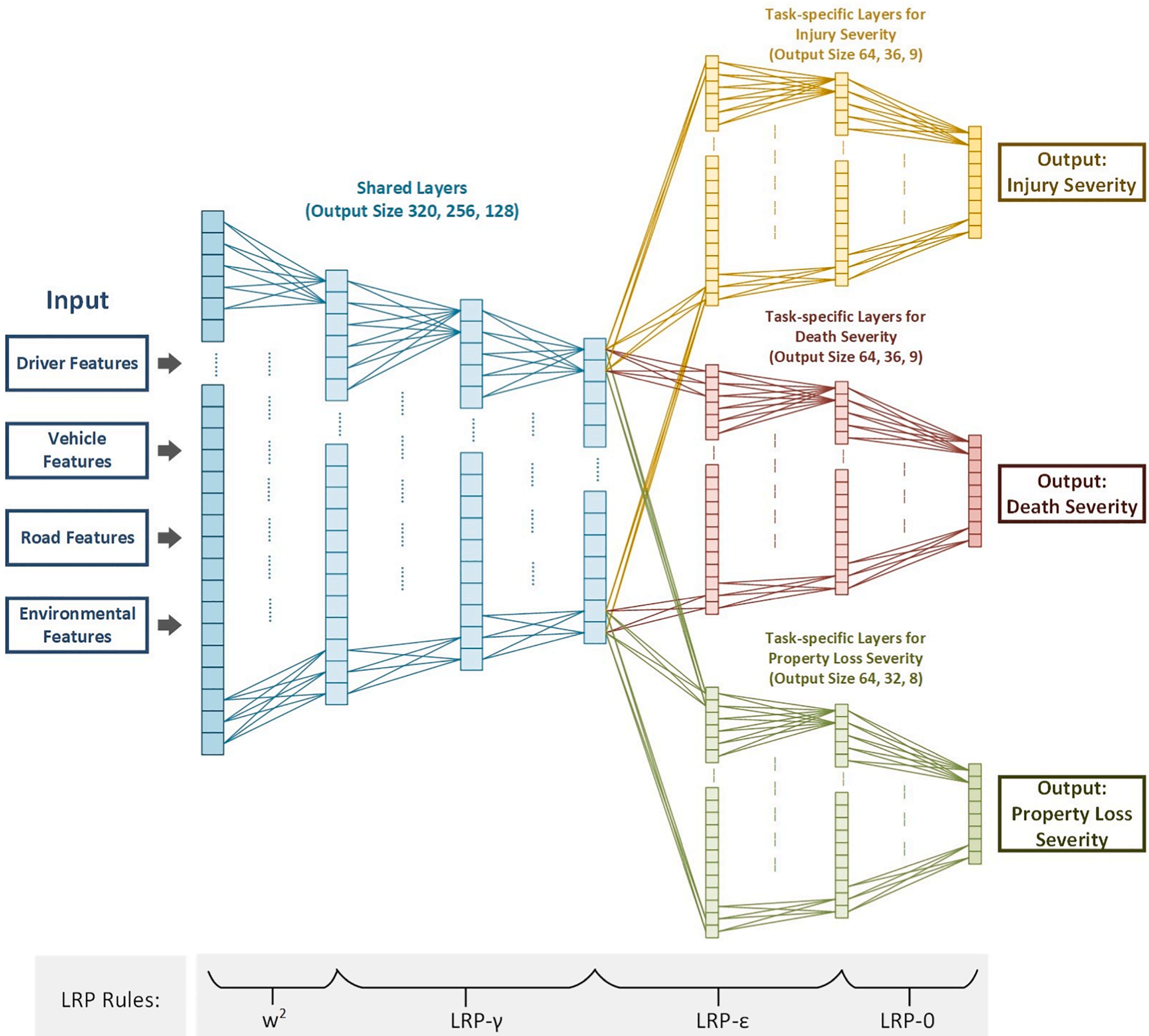

Fig. 1. Explainable multi-task deep learning framework for predicting multiple traffic accident severity.

input features (Saltelli, 2002). Almost all data mining methods and soft computing methods could be explained by sensitivity analysis, as only a set of model inputs and corresponding outputs is needed. Furthermore, Arteaga et al. (2020) propose Global Cross-Validation Local Interpretable Model-Agnostic Explanations (GCV-LIME) to discover likely causality factors for injury/death severities (Ribeiro et al., 2016). GCV-LIME has proven to be effective in explaining single-layer neural networks, as well as random forest, support vector machine, and logistic regression (Arteaga et al., 2020).

While previous research applies sensitivity analysis and LIME to explain the traffic accident severity predictions generated by neural networks, these methods are not specifically designed for neural networks and do not utilize the structure and weights of neural networks to generate explanations. Therefore, these methods might generate less accurate and reliable explanations than methods that utilize structural information to explain neural networks explicitly. Furthermore, previous explanation methods for neural networks are only tested on single-layer neural networks, and they might not be suitable for explaining DNNs with more than three layers.

3. Explainable multi-task deep learning framework for predicting multiple traffic accident severity

To better predict multiple types of traffic accident severity, we propose an explainable multi-task framework based on DNN. To identify key factors that affect traffic accidents in a reasonable and informative way, we apply LRP to explain the prediction outcome of our proposed model. This framework is the first multi-task learning model and the first DNN-based model for traffic accident severity prediction to the best of our knowledge.

3.1. Model structure

Since it is necessary to predict all three types of traffic accident severity, we construct a multi-task deep learning framework. Deep learning has proven to be effective in processing huge amount of data and tackling a wide variety of prediction tasks (Jan et al., 2019; Jeon et al., 2020). Multi-task learning is an inductive transfer mechanism that improves generalization by leveraging domain-specific information

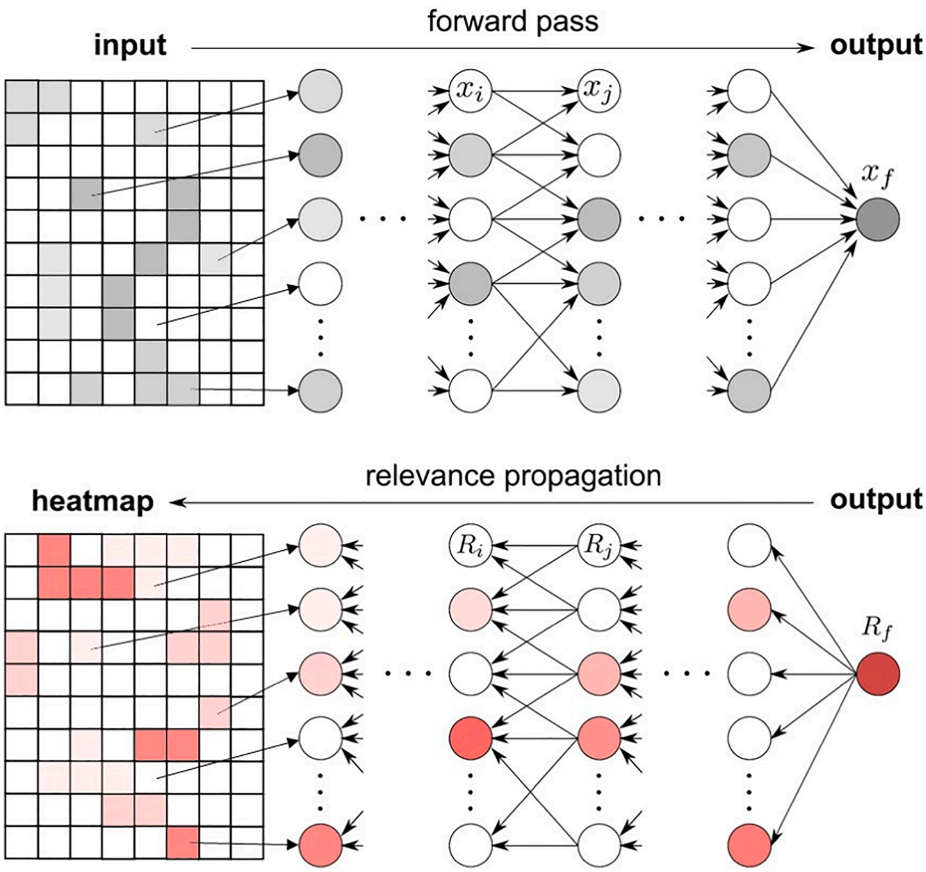

Fig. 2. The computational flow of LRP in a DNN (adapted from Montavon et al. (2017)). The input of the DNN is a matrix of different values (represented by the grey squares). By forwardly propagating the input through the DNN, an output is obtained. Then, to generate the explanation of the output, the output is backwardly propagated through the DNN. At each DNN layer, an LRP rule is applied to calculate the relevance score of each layer. When this backward propagation reaches the input layer, a heatmap is generated to highlight which parts in the input (represented by the red squares) contribute the most to the output. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from related tasks (Radwan et al., 2018). Comparing to traditional single-task learning, multi-task learning enhances the prediction quality of the three types of traffic accidents. In single-task learning, each task is modeled using a distinct dataset and training process. However, in the case of predicting the injury severity, death severity, and property loss severity of traffic accidents, the three types of traffic accident severity are closely associated with each other. For example, the value of property loss in a traffic accident often depends on the number of injured and deaths in the accident. Therefore, utilizing the information of the predicted injury/death severity in the training and prediction process of property loss severity would increase its prediction accuracy, and vice versa. In this study, we apply multi-task learning techniques to share common structure and data among the three DNN models for predicting the three types of traffic accident severities together, and train the three DNN models jointly. Multi-task learning has a better performance than single-task learning in modeling tasks that share common features or attributes (Hajiramezanali et al., 2018). Since the injury severity, death severity, and property loss severity of traffic accidents are closely associated with each other, we believe that modeling the three types of severity together using multi-task learning and deep learning will enhance the prediction accuracy of all three types of severity.

The general structure of the proposed multi-task DNN model, which shares common model structures across the three traffic accident severity prediction tasks, is presented in Fig. 1. Our model takes a vector as input, which contains four types of traffic accident features including driver features, vehicle features, road features, and environmental features. The traffic accident features first pass through three shared neural network layers with an output size of 320, 256, and 128, respectively. The three shared layers enable information and data sharing among the three traffic accident severity prediction tasks. By utilizing the shared information of other tasks, the prediction performance of each of the three types of traffic accident severity is enhanced. Each layer in our multi-task DNN model transforms the data into different dimensions and learns traffic accident patterns from the data using the following equation:

$$\mathbf{x}_i = \mathbf{W}_i\mathbf{x}_{i-1} + \mathbf{b}_i, \tag{1}$$

where $\boldsymbol{x}_i$ denotes the output of the i-th layer, $\boldsymbol{W}_i$ is the weight matrix of the i-th layer, and $\boldsymbol{b}_i$ is a bias vector. Given that $\boldsymbol{x}_{i-1} \in \mathrm{R}^m$ and $\boldsymbol{x}_i \in \mathrm{R}^n$, a $n \times m$ weight matrix $\boldsymbol{W}_i$ could transform the input from dimension m to dimension n. In order to avoid the vanishing gradient issue while learning high dimensional patterns from traffic accident data, we use the rectified linear unit (ReLU) as our activation function for all three shared layers. The ReLU function is defined as:

$$\mathrm{ReLU}(\mathbf{x}) = \max(\mathbf{x}, 0), \tag{2}$$

and the modified layer function of our neural network is:

$$\mathbf{x}_i = \mathrm{ReLU}(\mathbf{W}_i\mathbf{x}_{i-1} + \mathbf{b}_i) \tag{3}$$

After the traffic accident features pass through the three shared layers, it is fed into three task-specific layers that are specially constructed for each of the three traffic accident severity prediction tasks. The task-specific layers aim at enhancing the individual prediction quality of each type of traffic accident severity by utilizing the global patterns identified by the shared layers. The sizes of task-specific layers are specifically designed to fit the dimensions of tasks' outputs. For predicting injury severity and death severity, we use three layers with an output size of 64, 36, and 9. For modeling property loss severity, we use three layers with an output size of 64, 32, and 8. We use the ReLU activated neural network design (Equation (3)) for the first two task-specific layers of each task to avoid gradient vanishing. For the last task-specific layer, we use the softmax activation function to obtain the probability of each level of traffic accident severity:

Table 1
Categories of traffic accident severity variables.

Dependent Variable	Data Type	Number of Categories	Description
Injury severity	Ordinal	9	0: $n = 0$ 1: $n = 1$ 2: $n = 2$ 3: $n = 3$ 4: $n = 4$ 5: $5 \leq n \leq 9$ 6: $10 \leq n \leq 14$ 7: $15 \leq n \leq 19$ 8: $n \geq 20$ (n = Number of injured)
Death severity	Ordinal	9	0: $n = 0$ 1: $n = 1$ 2: $n = 2$ 3: $n = 3$ 4: $n = 4$ 5: $5 \leq n \leq 9$ 6: $10 \leq n \leq 14$ 7: $15 \leq n \leq 19$ 8: $n \geq 20$ (n = Number of deaths)
Property loss severity	Ordinal	8	0: $n = 0$ 1: $1 \leq n \leq 9$ 2: $10 \leq n \leq 99$ 3: $100 \leq n \leq 999$ 4: $1000 \leq n \leq 9999$ 5: $10000 \leq n \leq 99999$ 6: $100000 \leq n \leq 999999$ 7: $n \geq 1000000$ (n = Value of property loss)

$$\text{Softmax}(\mathbf{x}_i) = \frac{exp(\mathbf{x}_i)}{\sum_j exp(\mathbf{x}_j)}, \tag{4}$$

and the neural network function of the last task-specific layer is:

$$\mathbf{x}_i = \text{Softmax}(\mathbf{W}_i\mathbf{x}_{i-1} + \mathbf{b}_i) \tag{5}$$

The softmax activation function transforms the components in the input vector to positive values that sum up to 1, which could be interpreted as a probability. The outputs of our model are three vectors of sizes 9, 9, and 8, which represent the likelihood that each of the injury, death, and property loss severity levels would occur. Our multi-task DNN is trained by minimizing the following mean squared error loss function:

$$L(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{N} (\mathbf{y}_k - h_{\mathbf{W}}(\mathbf{x}_k))^2, \tag{6}$$

where $\boldsymbol{y}$ is the true probability that a traffic accident severity level occurs, and $h_{\boldsymbol{W}}(\boldsymbol{x}_k)$ is the predicted probability based on the input features $\boldsymbol{x}$ and the learned weights $\boldsymbol{W}$.

3.2. Model explainability

The explanations of traffic accident severity predictions of our proposed framework are generated by LRP, an explainable artificial intelligence method that highlights the contribution of each input element to the output probability by propagating the output probability back through the neural network using local propagation rules (Bach et al., 2015; Montavon et al., 2017). LRP is a technique explicitly designed for explaining neural network models such as highly complex DNNs and convolutional neural networks, and it calculates explanations based on the weights and inputs of each neural network layer (Montavon et al., 2019). Thus, LRP is different from sensitivity analysis and LIME, which could provide explanations for almost any machine-learning model and does not use the structure and weights of the explained neural network (Arteaga et al., 2020; Ribeiro et al., 2016). Fig. 2 shows how a DNN is explained by LRP. Specifically, LRP propagates the output R_f backward through the neural network and generates a heatmap, which is a vector of weights representing the contribution of each element in the input vector. Utilizing the structural information of neural networks, LRP could offer effective explanations for DNNs containing the ReLU activation function (Kohlbrenner et al., 2020; Montavon et al., 2017). Thus, LRP is suitable for explaining the traffic accident severity predictions provided by our proposed multi-task DNN with ReLU activation functions.

Since each neural network layer in our proposed multi-task DNN has different functionalities, we design an LRP procedure with different LRP rules for each layer. The bottom part of Fig. 1 specifies the relevance score propagation rules used by each layer of our proposed model. We describe our proposed LRP procedure backwardly from the last layer to the first layer based on Montavon et al. (2017) and Montavon et al. (2019). To propagate the output probability of the traffic accident severity backward to the last task-specific layer in our multi-task DNN, we apply the fundamental rule of LRP (LRP-0):

$$R_i = \sum_j \frac{x_i w_{ij}}{\sum_i x_i w_{ij}} R_j, \tag{7}$$

where i and j are two neurons in consecutive layers, x_i is the input of the neuron, w_{ij} is the weight of the network, and R_i is the relevance score of the neuron. For the final layer in the neural network, R_j is the traffic accident severity prediction generated by the model that is expected to be explained. The LRP-0 rule is suitable for propagating relevance scores of the last layer in a neural network, because it is insensitive to entanglements.

Next, as the second and first task-specific layer contains more noise than the last task-specific layer, we utilize a more robust rule called the epsilon rule (LRP-ε) to mitigate false relevance. The LRP-ε adds a small positive term ε to the denominator of the LRP-0 rule to regulate the inference of relevance scores in two consecutive layers:

$$R_i = \sum_j \frac{x_i w_{ij}}{\varepsilon + \sum_i x_i w_{ij}} R_j. \tag{8}$$

Then, for the shared layers which contain global patterns shared among all tasks, we wish to highlight the positive contributions of these patterns to the traffic accident severity prediction outcome. Thus, we use the gamma rule (LRP-γ):

$$R_i = \sum_j \frac{x_i \left(w_{ij} + \gamma w_{ij}^{+}\right)}{\sum_i x_i \left(w_{ij} + \gamma w_{ij}^{+}\right)} R_j, \tag{9}$$

in which γ adds extra weight to positive contributions.

For the first shared layer in our proposed model, we use a special rule called the w^2-rule for propagating the relevance score from the first layer in the DNN to the input. The w^2-rule utilizes the square of the weights:

$$R_i = \sum_j \frac{w_{ij}^2}{\sum_i w_{ij}^2} R_j. \tag{10}$$

This propagation rule aims at redistributing relevance according to the square magnitude of the weights and pooling relevance across all neurons in the first shared layer.

When the relevance score is propagated back to the input, a vector of weights is obtained, which indicates the contribution of each corresponding feature in the input vectors to the predicted probability of each traffic accident severity level. For numerical and dummy variables, the relevance score is directly obtained from the vector. For ordinal and

Table 2
Independent variables related to traffic accident severity.

Feature Type	Independent Variable	Data Type	Number of Categories
Driver	Gender	Dummy	2
Driver	Age	Numerical	–
Driver	Driving experience	Numerical	–
Driver	Drunk	Dummy	2
Vehicle	Overloaded	Dummy	2
Vehicle	Vehicle usage	Nominal	19
Vehicle	Vehicle type	Nominal	92
Vehicle	Passenger	Nominal	15
Vehicle	Registration plate type	Nominal	16
Vehicle	Brand	Nominal	203
Road	Road condition	Nominal	6
Road	Road surface condition	Nominal	7
Road	Central isolation facility	Nominal	7
Road	Road type	Nominal	12
Road	Highway administrative level	Nominal	5
Road	Intersection	Nominal	14
Road	Road cross-section position	Nominal	7
Road	Road alignment	Nominal	10
Road	Roadside protection facility	Nominal	10
Environment	Month	Numerical	–
Environment	Weekday	Numerical	–
Environment	Hour	Numerical	–
Environment	Weather	Nominal	8
Environment	Visibility	Ordinal	4
Environment	Light	Nominal	5

Table 3
Accuracy of traffic accident severity prediction models.

	Injury Severity	Death Severity	Property Loss Severity
Multi-Task DNN	0.4658	0.6607	0.5497
SVM-FCM	0.4189	0.5756	0.4747
Single-Task DNN	0.4480	0.6555	0.5216
Random Forest	0.4181	0.5714	0.4450
Logistic Regression	0.4185	0.5918	0.4733

Table 4
AUC of traffic accident severity prediction models.

	Injury Severity	Death Severity	Property Loss Severity
Multi-Task DNN	0.6995	0.8091	0.7427
SVM-FCM	0.6731	0.7613	0.6998
Single-Task DNN	0.6895	0.8062	0.7266
Random Forest	0.6727	0.7589	0.6829
Logistic Regression	0.6729	0.7704	0.6990

nominal variables that are represented as a one-hot vector, we average all relevance scores in the vector that corresponds to the elements of the one-hot vector to get the relevance score of the variable.

3.3. Dataset Description

To test our proposed model, we use traffic accident data from three provinces of China, which are Guizhou, Jilin, and Shanxi. The dataset is provided by the Key Laboratory for Urban Transportation Complex Systems Theory and Technology of Ministry of Education, Beijing Jiaotong University. Containing 50,540 traffic accidents that happened between 2005 and 2014, the dataset includes accidents that happened on highways, urban areas, and rural areas. The variables in the dataset include information about the driver, vehicle, road, and environment of the accident. In this dataset, some variables are extracted from traffic sensors. For example, the original record of the cause of the accident, visibility, light, and road condition are collected from tachographs, traffic surveillance cameras, and visibility sensors. Therefore, the dataset experiences a complex data preprocessing process.

The dependent variables of our study are injury severity, death severity, and property loss severity in traffic accidents. Injury severity is represented by the number of people injured in the accident. Death severity refers to the number of people who died in the accident. Property loss severity is measured by the total value of property loss in Chinese yuan in a traffic accident. Without loss of generality, we convert the three numerical variables into ordinal variables by grouping them into 9, 9, and 8 groups respectively based on the frequency of each number. Table 1 shows the categories of the three traffic accident severity variables and the values each group represents.

The independent variables of our research are 25 variables of the driver, vehicle, road, and environmental features of the traffic accident. Table 2 describes all independent variables in our dataset.

For driver features, our dataset contains the gender and age of the driver. The number of years of driving experience of the driver is also documented. Furthermore, we use data mining techniques to construct the dummy variable drunk, which represents whether the driver is drunk when the accident happened, from the original record of the cause of the accident. Specifically, we check whether the word “alcohol” appears in the cause of the accident. If the word “alcohol” occurs at least once, the value of the variable drunk is set to one.

As for vehicle features, the usage, type, and registration plate type of the vehicle are included in our dataset. In addition, the dataset contains the job occupation of the passengers of the vehicle, e.g., whether the passengers are students, workers, farmers, or foreigners. A dummy variable representing whether the vehicle is overloaded in the accident is also used. Furthermore, the dataset documents the brand of the vehicle. After data cleaning of combining the same brands, 203 vehicle brands are identified.

For road features, we employ nine variables, including the road condition, road surface condition, central isolation facility, road type, highway administrative level, intersection, road cross-section position, road alignment, and roadside protection facility. All nine variables are nominal.

As for environmental features, the weather situation, visibility, and light condition of the accident are contained in our dataset. While weather and light condition are nominal variables, visibility is an ordinal variable, which is classified as less than 50 m, 50 to 100 m, 100 to 200 m, and over 200 m of visibility. In addition, we extract the month, weekday, and hour of the accident from the time of the accident originally documented in the dataset.

We prepossess the dependent and independent variables by filling the missing entries with zero. Then, we code all ordinal and nominal independent variables into a one-hot vector. Finally, we concatenate all independent variables together to form a vector of length 448 to represent the independent variables of each traffic accident.

4. Experimental results

In this section, we verify the effectiveness and explainability of our proposed multi-task deep learning model for traffic accident severity prediction via prediction quality experiments and case studies.

4.1. Experiment setting

We use the dataset described in Subsection 3.3 to conduct the experiments. The dataset is randomly split into a train set, a validation set, and a test set of 40,432, 5054, and 5054 data respectively. We train all models on the train set, tune the models on the validation set, and evaluate the models on the test set. We apply the Adam optimizer (Kingma and Ba, 2014), a learning rate of 10^{-4}, and a batch size of 500

Table 5
The first case — top five features generated by different traffic accident severity prediction methods.

	Multi-Task DNN	SVM-FCM	Single-Task DNN	Random Forest	Logistic Regression
Injury Severity					
Prediction	1 person ✓	1 person ✓	1 person ✓	1 person ✓	1 person ✓
1st feature	Intersection (Ordinary section)	Road cross-section position (Motorway)	Drunk (No)	Driving experience (None)	Gender (Male)
2nd feature	Central isolation facility (None)	Brand (None)	Weather (Sunny)	Age (35)	Driving experience (None)
3rd feature	Road cross-section position (Motorway)	Age (35)	Registration plate type (None)	Gender (Male)	Light (Daylight)
4th feature	Overloaded (No)	Month (9)	Brand (None)	Hour (6)	Month (9)
5th feature	Road type (First grade highway)	Weekday (Wednesday)	Roadside protection facility (No protection)	Month (9)	Vehicle usage (None)
Death Severity					
Prediction	1 person ✓	0 person ×	0 person ×	0 person ×	0 person ×
1st feature	Roadside protection facility (No protection)	Age (35)	Drunk (No)	Month (9)	Month (9)
2nd feature	Road alignment (Straight)	Highway administrative level (National highway)	Road alignment (Straight)	Hour (6)	Driving experience (None)
3rd feature	Drunk (No)	Central Isolation facility (None)	Road condition (Pavement intact)	Light (Daylight)	Light (Daylight)
4th feature	Road condition (Pavement intact)	Intersection (Ordinary section)	Road type (First grade highway)	Driving experience (None)	Passenger (Farmer)
5th feature	Intersection (Ordinary section)	Road surface condition (Dry)	Overloaded (No)	Roadside protection facility (No protection)	Gender (Male)
Property Loss Severity					
Prediction	10,000 to 99,999 yuan ✓	1000 to 9999 yuan ×	10,000 to 99,999 yuan ×	1000 to 9999 yuan ×	1000 to 9999 yuan ×
1st feature	Drunk (No)	Driving experience (None)	Roadside protection facility (No protection)	Highway administrative level (National highway)	Gender (Male)
2nd feature	Roadside protection facility (No protection)	Age (35)	Drunk (No)	Hour (6)	Highway administrative level (National highway)
3rd feature	Road alignment (Straight)	Visibility (over 200 m)	Road surface condition (Dry)	Road type (First grade highway)	Roadside protection facility (No protection)
4th feature	Highway administrative level (National highway)	Roadside protection facility (No protection)	Highway administrative level (National highway)	Roadside protection facility (No protection)	Hour (6)
5th feature	Passenger (Farmer)	Highway administrative level (National highway)	Intersection (Ordinary section)	Light (Daylight)	Month (9)

to train our proposed model. Furthermore, early stopping is employed, which refers to stopping the training process when the validation loss has not decreased for 20 epochs.

We select the state-of-the-art model fuzzy c-means based support vector machine (SVM-FCM), a soft computing technique single-task DNN, a data mining technique random forest, and a discrete outcome model logistic regression as the four baseline models of this study. SVM-FCM (Assi et al., 2020) first clusters the data into several clusters using FCM and then develops one SVM model for each cluster. We use the model settings specified by the authors and set other settings to default. The single-task DNN has the same six layers as our proposed multi-task model, except that there is no shared layer and all three tasks are trained separately. The single-task model is trained using the same optimizer, learning rate, and batch size as the multi-task model along with early stopping. The second baseline model is the random forest, which ensembles multiple decision trees using a randomly selected subset of training samples and variables (Belgiu and Drăguţ, 2016; Ho, 1995). The random forest model contains 100 decision trees with a depth of 6, and all other parameters are set to default. The logistic regression model is trained using cross-entropy loss with all other parameters set to default as well. We train three distinct random forests and logistic regression models for the three traffic accident severity tasks. The SVM-FCM model is developed based on Python packages scikit-fuzzy and scikit-learn. The multi-task and the single-task models are built using the Python package PyTorch, while the random forest model and the logistic regression model are constructed using the Python package scikit-learn.

4.2. Experiments for prediction quality comparison

We conduct a quantitative experiment on the prediction quality and

Table 6
The second case — top five features generated by different traffic accident severity prediction methods.

Injury Severity					
	Multi-Task DNN	**SVM-FCM**	**Single-Task DNN**	**Random Forest**	**Logistic Regression**
Prediction	15 to 18 people ✓	1 person ×	2 people ×	1 person ×	1 person ×
1st feature	Drunk (No)	Road cross-section position (Mixed motorized and non-motorized way)	Drunk (No)	Driving experience (None)	Gender (Male)
2nd feature	Light (Night with street light)	Brand (None)	Highway administrative level (County road)	Age (59)	Driving experience (None)
3rd feature	Road condition (Bump)	Age (59)	Central isolation facility (None)	Gender (Male)	Light (Night with street light)
4th feature	Road alignment (Straight)	Month (5)	Registration plate type (None)	Hour (19)	Month (5)
5th feature	Road surface condition (Wet)	Weekday (Wednesday)	Intersection (None)	Month (5)	Vehicle usage (None)

Death Severity					
	Multi-Task DNN	**SVM-FCM**	**Single-Task DNN**	**Random Forest**	**Logistic Regression**
Prediction	2 people ✓	0 person ×	2 people ✓	0 person ×	1 person ×
1st feature	Road condition (Bump)	Age (59)	Light (Night with street light)	Month (5)	Gender (Male)
2nd feature	Light (Night with street light)	Highway administrative level (County road)	Intersection (None)	Hour (19)	Month (5)
3rd feature	Road surface condition (Wet)	Central isolation facility (None)	Central isolation facility (None)	Light (Night with street light)	Driving experience (None)
4th feature	Road alignment (Straight)	Intersection (None)	Visibility (50 to 100 m)	Driving experience (None)	Drunk (No)
5th feature	Drunk (No)	Road surface condition (Wet)	Drunk (No)	Roadside protection facility (No protection)	Light (Night with street light)

Property Loss Severity					
	Multi-Task DNN	**SVM-FCM**	**Single-Task DNN**	**Random Forest**	**Logistic Regression**
Prediction	1000 to 9999 yuan ✓	1000 to 9999 yuan ✓	1000 to 9999 yuan ✓	1000 to 9999 yuan ✓	1000 to 9999 yuan ✓
1st feature	Drunk (No)	Driving experience (None)	Roadside protection facility (No protection)	Highway administrative level (County road)	Gender (Male)
2nd feature	Central isolation facility (None)	Age (59)	Drunk (No)	Hour (19)	Highway administrative level (County road)
3rd feature	Roadside protection facility (No protection)	Visibility (50 to 100 m)	Intersection (None)	Road type (Fourth grade highway)	Roadside protection facility (No protection)
4th feature	Road alignment (Straight)	Roadside protection facility (No protection)	Road surface condition (Wet)	Roadside protection facility (No protection)	Hour (19)
5th feature	Light (Night with street light)	Highway administrative level (County road)	Registration plate type (None)	Light (Night with street light)	Month (5)

effectiveness of our proposed framework using the Chinese traffic accident dataset described in Subsection 3.3. To measure the prediction quality of our proposed model, we use accuracy and the area under curve (AUC) of the receiver operating characteristic (ROC) following previous research (Mujalli et al., 2016). Accuracy is the proportion of correct results in the model's prediction results. The ROC curve plots the values of the true positive and false positive as the decision threshold varies (Bradley, 1997). The overall performance of the prediction model is then represented by the area under the ROC curve, which is AUC (Mujalli et al., 2016).

Table 3 and 4 show the accuracy and AUC scores of our proposed multi-task DNN framework as well as the baseline models. From the tables, it is observed that our proposed model outperforms all baseline models on accuracy and AUC on all three traffic accident severity prediction tasks. Concretely, our proposed model outperforms the state-of-the-art model SVM-FCM by 13.93%, single-task DNN by 3.38%, random forest by 16.85%, and logistic regression by 13.02% on average for accuracy. As for AUC, our proposed model achieves 5.44% improvement compared to SVM-FCM, 1.34% improvement compared to single-task DNN, 6.45% improvement compared to random forest, and 5.07% improvement compared to logistic regression on average. The experimental result proves that our model has state-of-the-art performance, and the multi-task design is indeed superior to the single-task design of the baseline models.

4.3. Case studies for explainability comparison

We conduct case studies to demonstrate the explainability of our proposed model following previous research (Arteaga et al., 2020; Buendia et al., 2015). The case studies directly present the key factors associated with these three types of traffic accident severity for researchers and practitioners, which helps them develop and implement public policies related to VRUs safety based on the identified key factors. We randomly select two cases in Subsection 4.3.1 and 4.3.2 to examine

Table 7
The first case — top five features generated by different explanation methods for neural networks.

	LRP	GCV-LIME	Sobol
Injury Severity			
1st feature	Intersection (Ordinary section)	Overloaded (No)	Age (35)
2nd feature	Central isolation facility (None)	Passenger (Farmer)	Central isolation facility (None)
3rd feature	Road cross-section position (Motorway)	Vehicle usage (None)	Brand (None)
4th feature	Overloaded (No)	Vehicle type (None)	Registration plate type (None)
5th feature	Road type (First grade highway)	Visibility (over 200 m)	Road type (First grade highway)
Death Severty			
1st feature	Roadside protection facility (No protection)	Passenger (Farmer)	Central isolation facility (None)
2nd feature	Road alignment (Straight)	Hour (6)	Road alignment (Straight)
3rd feature	Drunk (No)	Roadside protection facility (No protection)	Vehicle type (None)
4th feature	Road condition (Pavement intact)	Road cross-section position (Motorway)	Age (35)
5th feature	Intersection (Ordinary section)	Driving experience (None)	Intersection (Ordinary section)
Property Loss Severty			
1st feature	Drunk (No)	Road surface condition (Dry)	Age (35)
2nd feature	Roadside protection facility (No protection)	Road alignment (Straight)	Brand (None)
3rd feature	Road alignment (Straight)	Road type (First grade highway)	Road type (First grade highway)
4th feature	Highway administrative level (National highway)	Light (Daylight)	Vehicle type (None)
5th feature	Passenger (Farmer)	Intersection (Ordinary section)	Registration plate type (None)

the explainability of our model. Furthermore, in Subsection 4.3.3 we present a list of global key features for the three types of traffic accident severity.

4.3.1. Comparison with traditional prediction methods

Traditional explainable discrete outcome models, data mining techniques, and computing models have been widely adopted by previous traffic accident severity research. We compare the explainability of our multi-task DNN model and corresponding LRP procedure with traditional traffic accident severity prediction methods. Table 5 and Table 6 show the top five features that contributed the most to the traffic accident severity outcomes of two accidents given by our proposed multi-task DNN model and the four baseline models. The predicted severity levels are also presented in the tables, and the severity levels are categorized according to Table 1. Here, both the key factors of multi-task DNN and single-task DNN are generated by LRP. The state-of-the-art method SVM-FCM is explained by the Sobol method. The explanations of random forest are generated by calculating feature importance based on impurity, and the explanations of logistics regression are the coefficients of the features.

In both cases, our proposed multi-task DNN predicts all three types of traffic accident severities correctly, while the four baseline methods predict one or two severities wrongly. Comparing to the top five features of multi-task DNN, the explanations given by single-task DNN and logistics regression have more features with missing values. Since missing values could not predict traffic accident severities, multi-task DNN provides more reasonable explanations than single-task DNN and logistics regression. Furthermore, the explanations of SVM-FCM and random forest are invariant to different levels of traffic accident severities, and the top five features provided by SVM-FCM and random forest are always the same for all cases, as the impurity-based feature importance is calculated globally. Therefore, the explanations of SVM-FCM and random forest are less informative and case-specific than the explanations of multi-task DNN, single-task DNN, and logistics regression. To sum up, the explanations of multi-task DNN are more reasonable and informative than the explanations of the baseline methods.

4.3.2. Comparison with other explanation methods for neural networks

Recently, DNN models have become increasingly popular, but limited methods have been developed for explaining DNNs. We examine the explanation quality of our proposed LRP procedure with other explanation methods for neural networks. We choose GCV-LIME (Arteaga et al., 2020), the state-of-the-art explanation method for traffic accident severity prediction model, and the Sobol method (Sobol, 2001), a popular sensitivity analysis model, as two baseline methods for explaining multi-task DNN. We then compare the baseline methods' performance with our proposed LRP procedure. Since we are comparing individual explanations, the individual LIME explanations in GCV-LIME are applied. Here, GCV-LIME is built using the Python package lime, and the Sobol method is developed using the Python package SALib (Herman

Table 8
The second case — top five features generated by different explanation methods for neural networks.

	Injury Severity		
	LRP	GCV-LIME	Sobol
1st feature	Drunk (No)	Road condition (Bump)	Age (59)
2nd feature	Light (Night with street light)	Road type (Fourth grade highway)	Central isolation facility (None)
3rd feature	Road condition (Bump)	Road surface condition (Wet)	Brand (None)
4th feature	Road alignment (Straight)	Highway administrative level (County road)	Registration plate type (None)
5th feature	Road surface condition (Wet)	Hour (19)	Road type (Fourth grade highway)
	Death Severty		
	LRP	GCV-LIME	Sobol
1st feature	Road condition (Bump)	Passenger (None)	Central isolation facility (None)
2nd feature	Light (Night with street light)	Road alignment (Straight)	Road alignment (Straight)
3rd feature	Road surface condition (Wet)	Intersection (None)	Vehicle type (None)
4th feature	Road alignment (Straight)	Roadside protection facility (No protection)	Age (59)
5th feature	Drunk (No)	Vehicle usage (None)	Intersection (None)
	Property Loss Severity		
	LRP	GCV-LIME	Sobol
1st feature	Drunk (No)	Road type (Fourth grade highway)	Age (59)
2nd feature	Central isolation facility (None)	Roadside protection facility (No protection)	Brand (None)
3rd feature	Roadside protection facility (No protection)	Highway administrative level (County road)	Road type (Fourth grade highway)
4th feature	Road alignment (Straight)	Age (59)	Vehicle type (None)
5th feature	Light (Night with street light)	Vehicle usage (None)	Registration plate type (None)

and Usher, 2017).

Table 7 and Table 8 demonstrate the top five features of our proposed multi-task DNN model generated by the three explanation methods. Comparing to the state-of-the-art method GCV-LIME and the Sobol method, the top five features given by LRP has much fewer features with missing values. Missing values could not predict traffic accident severities as they represent no information. Therefore, the explanations of GCV-LIME and the Sobol method are less reliable and reasonable than the explanations of LRP. Thus, we could conclude that LRP is better than GCV-LIME and the Sobol method in explaining DNNs.

4.3.3. Global explanation generated by LRP

Summarizing global key factors of traffic accident severity is also essential for developing public policies to promote VRU safety. We examine the ability to generate global explanations of our proposed LRP procedure by presenting a list of global key features generated on the whole test set of our Chinese traffic accident dataset. The global features of LRP are generated by ranking key features in ascending order according to their LRP scores in each case, assigning scores to features which equal to their ranks, and adding the scores obtained by each feature in each case together. Table 9 shows the global importance of each factor to the three types of traffic accident severity predicted by our proposed multi-task DNN. We could observe that the importance ranking of global factors for the three types of traffic accident severity are roughly the same.

Table 9
Global key factors of traffic accident severity.

Rank	Injury Severity	Death Severity	Property Loss Severity
1	Roadside protection facility	Roadside protection facility	Roadside protection facility
2	Road alignment	Road alignment	Intersection
3	Intersection	Intersection	Road alignment
4	Central isolation facility	Central isolation facility	Central isolation facility
5	Drunk	Drunk	Road condition
6	Road condition	Road condition	Road surface condition
7	Road surface condition	Road surface condition	Registration plate type
8	Road cross-section position	Road cross-section position	Road cross-section position
9	Passenger	Weather	Drunk
10	Weather	Passenger	Passenger
11	Road type	Road type	Brand
12	Registration plate type	Registration plate type	Weather
13	Vehicle usage	Light	Light
14	Light	Highway administrative level	Road type
15	Highway administrative level	Vehicle usage	Vehicle usage
16	Vehicle type	Vehicle type	Vehicle type
17	Brand	Brand	Highway administrative level
18	Overloaded	Overloaded	Overloaded
19	Visibility	Visibility	Visibility
20	Weekday	Weekday	Weekday
21	Gender	Gender	Gender
22	Month	Month	Month
23	Hour	Hour	Hour
24	Driving experience	Driving experience	Driving experience
25	Age	Age	Age

5. Conclusion

Traffic accident severity prediction is crucial for preventing future accidents and mitigating different types of traffic accident severity. In this study, we design a multi-task deep learning framework that predicts different levels of injury severity, death severity, and property loss severity. Our model also provides key factors that contribute to the three traffic accident severities as explanations. Our model is the first multi-task learning model and the first DNN-based model for traffic accident severity prediction to the best of our knowledge. The major contributions of our research are as follows. First, the multi-task and deep learning design of our framework not only provides practitioners with a comprehensive traffic accident severity analysis, but also enhances the prediction accuracy. When comparing the prediction quality of various baseline models on the Chinese traffic accident data, our model outperforms state-of-the-art methods. Second, the explanations provided by our model through LRP help to identify key factors that affect traffic accident severity. A series of case studies show that our proposed model achieves the best explanation quality comparing to discrete outcome models, data mining techniques, and other soft computing models. Third, the insights obtained with the proposed multi-task deep learning framework demonstrate great potential to aid researchers and practitioners in the area of VRU safety.

The practical implications of our proposed multi-task deep learning framework include assisting the design of VRU safety policies and facilitating smart city development. Based on the precise prediction and the identified key factors of our proposed model, the transportation and urban construction departments could impose actions to mitigate the injury, death, and property loss of VRUs and reduce the rates of traffic accidents. The transportation departments of governments could utilize our model and its explanations to predict traffic accidents more accurately, discover factors that cause severe traffic accidents, choose safety countermeasures and policies, and fix issues in transportation and urban construction. For example, the case study in Subsection 4.3 reveals that road conditions affect the level of injury/death severity in traffic accidents. Thus, the urban construction departments could utilize this information and regularly inspect the road condition to avoid traffic accidents and citizen casualties. These innovative public policies, including safety requirements, actions, and plans, not only enhance the traffic safety of citizens but also promote the development of smart cities (Anisetti et al., 2018; Chehri et al., 2020).

For future research, several possible directions exist. First, we will use more enriched data if there exist corresponding data sources. For example, image data of the car and the road could be added as data sources. Second, the explainability of our model could be improved. In the future, we could develop methods for extracting decision rules from DNNs. Third, more sophisticated multi-task techniques could be added to our model, which will enable more information sharing among the three traffic accident severity prediction tasks and provide state-of-the-art safety solutions.

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.

Acknowledgements

The authors appreciate the editors and the anonymous reviewers for their detailed and constructive comments. This work was supported in part by the National Natural Science Foundation of China (Grant numbers: 71874197, 71801217, and 71871179) and the Hong Kong Research Grants Council (Grant number: 11509419).

References

Abellán, J., López, G., de OñA, J., 2013. Analysis of traffic accident severity using decision rules via decision trees. Expert Syst. Appl. 40 (15), 6047–6054.
Al-Ghamdi, A.S., 2007. Experimental evaluation of fog warning system. Accid. Anal. Prev. 39 (6), 1065–1072.
Anisetti, M., Ardagna, C., Bellandi, V., Cremonini, M., Frati, F., Damiani, E., 2018. Privacy-aware big data analytics as a service for public health policies in smart cities. Sustain. Cities Soc. 39, 68–77.
Arteaga, C., Paz, A., Park, J., 2020. Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. Saf. Sci. 132, 104988.
Assi, K., Rahman, S.M., Mansoor, U., Ratrout, N., 2020. Predicting crash injury severity with machine learning algorithm synergized with clustering technique: A promising protocol. Int. J. Environ. Res. Public Health 17 (15), 5497.
Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., Samek, W., 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS ONE 10 (7), e0130140.
Behnood, A., Al-Bdairi, N.S.S., 2020. Determinant of injury severities in large truck crashes: A weekly instability analysis. Saf. Sci. 131, 104911.
Beirness, D.J., Simpson, H.M., Williams, A.F., 2006. Role of cannabis and benzodiazepines in motor vehicle crashes. Transport. Res. Circular 12–21.
Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.
Bradley, A.P., 1997. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recogn. 30 (7), 1145–1159.
Buendia, R., Candefjord, S., Fagerlind, H., Bálint, A., Sjöqvist, B.A., 2015. On scene injury severity prediction (osisp) algorithm for car occupants. Accid. Anal. Prev. 81, 211–217.
Chang, L.-Y., Chien, J.-T., 2013. Analysis of driver injury severity in truck-involved accidents using a nonparametric classification tree model. Saf. Sci. 51 (1), 17–22.
Chehri, A., Quadar, N., Saadane, R., 2020. Communication and localization techniques in VANET network for intelligent traffic system in smart cities: A review. In: Smart Transportation Systems 2020: Proceedings of 3rd KES-STS International Symposium, vol. 185. Springer Nature, p. 167.
Christie, S., Lyons, R.A., Dunstan, F.D., Jones, S.J., 2003. Are mobile speed cameras effective? A controlled before and after study. Injury Prevent. 9 (4), 302–306.
de Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on spanish rural highways using bayesian networks. Accid. Anal. Prev. 43 (1), 402–411.
Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accid. Anal. Prev. 38 (3), 434–444.
Edwards, J.B., 1998. The relationship between road accident severity and recorded weather. J. Saf. Res. 29 (4), 249–262.
Evans, L., 1986. The effectiveness of safety belts in preventing fatalities. Accid. Anal. Prev. 18 (3), 229–241.
Hajiramezanali, E., Dadaneh, S.Z., Karbalayghareh, A., Zhou, M., Qian, X., 2018. Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data. In *Advances in Neural Information Processing Systems* 9115–9124.
Halbersberg, D., Lerner, B., 2019. Young driver fatal motorcycle accident analysis by jointly maximizing accuracy and information. Accid. Anal. Prev. 129, 350–361.
Hand, D.J., Mannila, H., Smyth, P., 2001. Principles of Data Mining. The MIT Press.
Herman, J., Usher, W., 2017. Salib: an open-source python library for sensitivity analysis. J. Open Source Software 2 (9), 97.
Ho, T.K., 1995. Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition, vol. 1. IEEE, pp. 278–282.
Islam, M., 2015. Multi-vehicle crashes involving large trucks: A random parameter discrete outcome modeling approach. In *Journal of the Transportation Research Forum* 54, 77–104.
Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: Some additional evidence. J. Saf. Res. 37 (3), 267–276.
Jan, B., Farman, H., Khan, M., Imran, M., Islam, I.U., Ahmad, A., Ali, S., Jeon, G., 2019. Deep learning in big data analytics: a comparative study. Comput. Electr. Eng. 75, 275–287.
Jeon, G., Anisetti, M., Damiani, E., Kantarci, B., 2020. Artificial intelligence in deep learning algorithms for multimedia analysis. Multimedia Tools Appl. 79 (45), 34129–34139.
Kasantikul, V., Ouellet, J.V., Smith, T., Sirathranont, J., Panichabhongse, V., 2005. The role of alcohol in thailand motorcycle crashes. Accid. Anal. Prev. 37 (2), 357–366.
Kim, J.-K., Ulfarsson, G.F., Kim, S., Shankar, V.N., 2013. Driver-injury severity in single-vehicle crashes in california: a mixed logit analysis of heterogeneity due to age and gender. Accid. Anal. Prev. 50, 1073–1081.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Kohlbrenner, M., Bauer, A., Nakajima, S., Binder, A., Samek, W., Lapuschkin, S., 2020. Towards best practice in explaining neural network decisions with lrp. In: International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–7.
Kononen, D.W., Flannagan, C.A., Wang, S.C., 2011. Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accid. Anal. Prev. 43 (1), 112–122.
Kraus, J.F., Anderson, C., Zador, P., Williams, A., Arzemanian, S., Li, W.-C., Salatka, M., 1991. Motorcycle licensure, ownership, and injury crash involvement. Am. J. Public Health 81 (2), 172–176.
Kwon, O.H., Rhee, W., Yoon, Y., 2015. Application of classification algorithms for analysis of road safety risk factor dependencies. Accid. Anal. Prev. 75, 1–15.
Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in florida. Accid. Anal. Prev. 37 (4), 775–786.
López, G., de Oña, J., Abellán, J., 2012. Using decision trees to extract decision rules from police reports on road accidents. Proc.-Soc. Behav. Sci. 53, 106–114.
Marks, R.J., 1994. Fuzzy Logic Technology and Applications I. IEEE press.
Monárrez-Espino, J., Hasselberg, M., Laflamme, L., 2006. First year as a licensed car driver: Gender differences in crash experience. Saf. Sci. 44 (2), 75–85.
Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K.-R., 2019. Layer-wise relevance propagation: an overview. In: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer, pp. 193–209.
Montavon, G., Lapuschkin, S., Binder, A., Samek, W., Müller, K.-R., 2017. Explaining nonlinear classification decisions with deep taylor decomposition. Pattern Recogn. 65, 211–222.
Moore, D.N., Schneider IV, W.H., Savolainen, P.T., Farzaneh, M., 2011. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. Accid. Anal. Prev. 43 (3), 621–630.
Moral-García, S., Castellano, J.G., Mantas, C.J., Montella, A., Abellán, J., 2019. Decision tree ensemble method for analyzing traffic accidents of novice drivers in urban areas. Entropy 21 (4), 360.
Mujalli, R.O., de Oña, J., 2013. Injury severity models for motor vehicle accidents: a review. Proc. Inst. Civ. Eng.-Transport 166 (5), 255–270.
Mujalli, R.O., López, G., Garach, L., 2016. Bayes classifiers for imbalanced traffic accidents datasets. Accid. Anal. Prev. 88, 37–51.
Mussone, L., Bassani, M., Masci, P., 2017. Back-propagation neural networks and generalized linear mixed models to investigate vehicular flow and weather data relationships with crash severity in urban road segments. In: International Congress on Transport Infrastructure and Systems, TIS Roma 2017. Taylor & Francis Group, London, pp. 731–739.
National Bureau of Statistics of China, 2020. China Statistical Yearbook 2020.
Radwan, N., Valada, A., Burgard, W., 2018. Vlocnet++: Deep multitask learning for semantic visual localization and odometry. IEEE Rob. Autom. Lett. 3 (4), 4407–4414.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should I trust you?" explaining the predictions of any classifier. In: In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144.

Rifaat, S.M., Tay, R., De Barros, A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accid. Anal. Prev. 43 (1), 276–283.
Rosenberg, M.L., Martinez, R., 1996. Graduated licensure: A win-win proposition for teen drivers and parents. Pediatrics 98 (5), 959–960.
Rutter, D.R., Quine, L., 1996. Age and experience in motorcycling safety. Accid. Anal. Prev. 28 (1), 15–21.
Saltelli, A., 2002. Sensitivity analysis for importance assessment. Risk Anal. 22 (3), 579–590.
Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Math. Comput. Simul 55 (1–3), 271–280.
Soderstrom, C.A., Dischinger, P.C., Kufera, J.A., Ho, S.M., Shepard, A., 2005. Crash culpability relative to age and sex for injured drivers using alcohol, marijuana or cocaine. In: Annual Proceedings/Association for the Advancement of Automotive Medicine, vol. 49. Association for the Advancement of Automotive Medicine, p. 327.
Sze, V., Chen, Y.-H., Yang, T.-J., Emer, J.S., 2017. Efficient processing of deep neural networks: A tutorial and survey. Proc. IEEE 105 (12), 2295–2329.
Tambouratzis, T., Souliou, D., Chalikias, M., Gregoriades, A., 2014. Maximising accuracy and efficiency of traffic accident prediction combining information mining with computational intelligence approaches and decision trees. J. Artificial Intell. Soft Comput. Res. 4 (1), 31–42.
van Wyk, F., Khojandi, A., Masoud, N., 2019. A path towards understanding factors affecting crash severity in autonomous vehicles using current naturalistic driving data. In *Proceedings of SAI Intelligent Systems Conference* 106–120.
Wahab, L., Jiang, H., 2019. A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. PLoS ONE 14 (4), e0214966.
Wang, X., Kockelman, K.M., 2005. Occupant injury severity using a heteroscedastic ordered logit model: distinguishing the effects of vehicle weight and type. Transp. Res. Rec. 1908, 195–204.
World Health Organization, 2018. Global status report on road safety 2018. World Health Organization.
Xie, Y., Zhang, Y., Liang, F., 2009. Crash injury severity analysis using bayesian ordered probit models. J. Transp. Eng. 135 (1), 18–25.
Yau, K.K., Lo, H.-P., Fung, S.H., 2006. Multiple-vehicle traffic accidents in Hong Kong. Accid. Anal. Prev. 38 (6), 1157–1161.
Zajac, S.S., Ivan, J.N., 2003. Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural connecticut. Accid. Anal. Prev. 35 (3), 369–379.
Zhang, G., Yau, K.K., Chen, G., 2013. Risk factors associated with traffic violations and accident severity in china. Accid. Anal. Prev. 59, 18–25.
Zhu, L., Lu, L., Zhang, W., Zhao, Y., Song, M., 2019. Analysis of accident severity for curved roadways based on bayesian networks. Sustainability 11 (8), 2223.
Zhu, X., Srinivasan, S., 2011. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. Accid. Anal. Prev. 43 (1), 49–57.