Research Article

An Improved Deep Learning Model for Traffic Crash Prediction

Chunjiao Dong [1,2,3] **Chunfu Shao,**[1,2] **Juan Li,**[1] **and Zhihua Xiong** [1]

[1] *MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China*
[2] *Key Laboratory of Industry of Big Data Application Technology for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing 100044, China*
[3] *Center for Transportation Research, Tickle College of Engineering, University of Tennessee, 600 Henley Street, Knoxville, 37996, Tennessee, USA*

Correspondence should be addressed to Chunjiao Dong; cjdong@bjtu.edu.cn

Machine-learning technology powers many aspects of modern society. Compared to the conventional machine learning techniques that were limited in processing natural data in the raw form, deep learning allows computational models to learn representations of data with multiple levels of abstraction. In this study, an improved deep learning model is proposed to explore the complex interactions among roadways, traffic, environmental elements, and traffic crashes. The proposed model includes two modules, an unsupervised feature learning module to identify functional network between the explanatory variables and the feature representations and a supervised fine tuning module to perform traffic crash prediction. To address the unobserved heterogeneity issues in the traffic crash prediction, a multivariate negative binomial (MVNB) model is embedding into the supervised fine tuning module as a regression layer. The proposed model was applied to the dataset that was collected from Knox County in Tennessee to validate the performances. The results indicate that the feature learning module identifies relational information between the explanatory variables and the feature representations, which reduces the dimensionality of the input and preserves the original information. The proposed model that includes the MVNB regression layer in the supervised fine tuning module can better account for differential distribution patterns in traffic crashes across injury severities and provides superior traffic crash predictions. The findings suggest that the proposed model is a superior alternative for traffic crash predictions and the average accuracy of the prediction that was measured by RMSD can be improved by 84.58% and 158.27% compared to the deep learning model without the regression layer and the SVM model, respectively.

1. Introduction

Road traffic injuries are a leading cause of preventable death, especially, for the young people. In the United States, traffic crashes were the number one cause of death among people from 16 to 24 years old for each year from 2012 to 2014 [1]. In 2015, the nation lost 35,092 people in traffic crashes, a 7.2-percent increase from 32,744 in 2014, which is the largest percentage increase in nearly 50 years [2]. This is an average of approximately 96 people being killed on the nation's roadways every day of the year, and an average of more than four people per hour. In other words, one person dies on roadways every 15 minutes.

To understand the relationship between the influence factors and traffic crash outcomes, with the extracted data from police reports and state highway-asset-management databases, the analyses of traffic safety estimate and predicate the likelihood of a traffic crash. The number of crashes occurring on a defined spatial entity over a specific time period (for example, the number of crashes per year occurring at a roadway intersection, over a specified roadway segment, or in a region) would be considered as the dependent variables and some of the many factors affecting the likelihood of a traffic crash are analyzed and examined (see [3–5] for a comprehensive review). Though more and more factors that are relevant to the traffic crashes have been incorporated and

the proposed models became more and more sophisticated, there are still some factors that are not available to the researchers and the models result in bias estimations and erroneous predictions. In this study, we proposed an innovative approach for traffic crash prediction, which incorporates a multivariate regression layer into a dynamic deep learning model that contains an unsupervised feature learning module and a supervised fine tuning module governing the state dynamics to improve the performances of prediction.

2. Literature Review

The statistical methodologies, such as the Poisson, negative binomial (NB), and their variants in univariate and multivariate regression frameworks, have been successfully applied in crash count analyses [4, 6–8], which attempt to deal with the data and methodological issues associated with traffic crash estimations and predictions, and enhance our understanding on the relationship between the influence factors and traffic crash outcomes. However, current research in traffic safety indicates that the applied statistical modeling fails when dealing with complex and highly nonlinear data [9], which could suggest that the relationship between the influence factors and traffic crash outcomes is more complicated than can be captured by a single statistical approach. In addition, most of the statistical methods are based on some strong assumptions, such as specifying a priori and the error distribution. Moreover, a problematic issue is multicollinearity, i.e., the high degree of correlation between two or more independent variables. Furthermore, statistical models have difficulty when dealing with outliers, missing or noisy data [10].

To deal with the limitations of statistical methodologies, the machine learning methods, including Artificial Neural Network (ANN), Support Vector Machine (SVM) models, and deep learning models, have been applied to various traffic safety problems and used as data analytic methods because of their ability to work with massive amounts of multidimensional data. In addition, because of the modeling flexibility, learning and generalization ability, and good predictive ability, the machine learning has been considered as generic, accurate, and convenient mathematical models in the field of traffic safety.

Because the commonly used Poisson or NB regression models assume the predefined underlying relationship between dependent and independent variables and the violation of the assumption would lead to erroneous estimation, ANN and Bayesian neural network (BNN) models have been employed to analyze the traffic safety problems for many years. Although both ANN and BNN models have similar multilevel network structures, they are different in predicting the outcome variables. For ANN, the weights are assumed to fix. However, the weights of BNN follow a probability distribution and the prediction needs to be integrated over all the probability weights. Basically, the ANN can be characterized by three features: network architecture, model of a neuron, and learning algorithms. Chang [11] compared the performances of NB regression model and ANN in crash frequency analyses. The results showed that

ANN is a consistent alternative method for analyzing crash frequency. Abdelwahab and Abdel-Aty [12] employed two well-known ANN paradigms [the multilayer perceptron and radial basis functions (RBF) neural networks] to analyze the traffic safety of toll plazas and evaluate the impacts of electronic toll collection (ETC) systems on highway safety. The performance of ANN was compared with calibrated logit models. Modeling results showed that the RBF neural network was the best model for analyzing driver injury severity. Xie, Lord, and Zhang [13] evaluated the application of BNN models for predicting traffic crashes by using data collected on rural frontage roads in Texas. The results showed that back-propagation neural network (BPNN) and BNN models perform better than the NB regression model in terms of traffic crash prediction. The results also showed that BNNs could be used to address other issues in highway safety, such as the development of crash modification factors, and enhance the prediction capabilities for evaluating different highway design alternatives. Kunt, Aghayan, and Noii [14] employed a genetic algorithm (GA), pattern search, and ANN models to predict the severity of freeway traffic crashes. The results showed that the ANN provided the best predictions. Jadaan, Al-Fayyad, and Gammoh [15] developed a traffic crash prediction model using the ANN simulation with the purpose of identifying its suitability for predicting traffic crashes under Jordanian conditions. The results demonstrated that the estimated traffic crashes are close to actual traffic crashes. Akin and Akbas [16] proposed an ANN model to predict intersection crashes in Macomb County of the State of Michigan. The predictive capability of the ANN model was determined by classifying the crashes into these types: fatal, injury, and property damage only (PDO) crashes. The results were very promising and showed that ANN model is capable of providing an accurate prediction (90.9%) of the crash types. In summary, though ANN and BNN models show better linear/nonlinear approximation properties than traditional statistical approaches, these models often cannot be generalized to other data sets [3].

The SVM models have recently been introduced for traffic safety analyses [17, 18], which are a new class of models that are based on statistical learning theory and structural risk minimization [19]. These models are supposed to approximate any multivariate function to any desired degree of accuracy with a set of related supervised learning methods. Li et al. [17] evaluated the application of SVM models for predicting motor vehicle crashes. The results showed that SVM models predict crash data more effectively and accurately than traditional NB models. In addition, the findings indicated that the SVM models provide better (or at least comparable) performance than BPNN models and do not over-fit the data. To identify the relationship between severe crashes and the explanatory variables and enhance model goodness-of-fit, Yu and Abdel-Aty [20] developed three models to analyze crash injury severity, which include a fixed parameter logit model, a SVM model, and a random parameter logit model. The results showed that the SVM models and the random parameter models provide superior model fits compared to the fixed parameter logit model. Findings also demonstrate

that it is important to consider possible nonlinearity and individual heterogeneity when analyzing traffic crashes. Chen et al. [21] employed the SVM models to investigate driver injury severity patterns in rollover crashes using two-year crash data collected in New Mexico. The results showed that the SVM models produce reasonable predictions and the polynomial kernel outperforms the Gaussian RBF kernel. Dong, Huang, and Zheng [22] proposed a SVM model to handle multidimensional spatial data in crash prediction. The results showed that the SVM models outperform the nonspatial models in terms of model fitting and predictive performance. In addition, the SVM models provide better goodness-of-fit compared with Bayesian spatial model with conditional autoregressive prior when utilizing the whole dataset as the samples. Ren and Zhou [23] proposed a novel approach that combines particle swarm optimization (PSO) and SVM for traffic safety prediction. The results showed that the predictions of PSO-SVM are better than that from BP neural network. Yu and Abdel-Aty [24] proposed the SVM models with different kernel functions to evaluate real-time crash risk. The results showed that the SVM model with RBF kernel outperformed the SVM model with linear kernel and Bayesian logistic regression model. In addition, the findings showed that smaller sample size could improve the classification accuracy of the SVM models and variable selection procedure is needed prior to the SVM model development. Overall, it has been found that the SVM models showed better or comparable results to the outcomes predicted by ANN/BNN and other statistical models [19]. However, like ANN and BNN, the SVM models often cannot be generalized to other data sets and they all tend to behave as black-boxes, which cannot provide the interpretable parameters as statistical models do.

Other than the ANN/BNN and SVM models, other machine learning methods have been introduced in traffic safety analyses. Abdel-Aty and Haleem [25] introduced a recently developed machine learning technique—multivariate adaptive regression splines (MARS) to predict vehicle angle crashes using extensive data collected on unsignalized intersections in Florida. The results showed that MARS outperformed the NB models. The proposed MARS models showed promising results after screening the covariates using random forest. The findings suggested that MARS is an efficient technique for predicting crashes at unsignalized intersections.

Deep learning is a recently developed branch of machine learning method and has been successfully applied in speech recognition, visual object recognition, object detection, and many other domains such as drug discovery and genomics [26, 27]. Compared to the conventional machine learning techniques that were limited in their ability to process natural data in their raw form, deep learning constructs computational models aiming to extract inherent features in data from the lowest level to the highest level. Though the deep learning methods have shown outstanding performances in many applications [26], the applications of deep learning in the field of transportation are relatively few and only focusing on the topic of traffic flow prediction [28–30]. In this study, we proposed an improved deep learning model

for traffic crash prediction. The proposed model includes two modules: an unsupervised feature learning module and a supervised fine tuning module. To discover nonlinear relationship between the investigated variables and identify the impacts of influence factors on traffic crashes for roadway network, a DAE model is proposed in the unsupervised feature learning module to learn the features of explanatory variables. In addition, a multivariate negative binomial (MVNB) regression is embedding into the supervised fine tuning module to address the heterogeneity issues. The proposed model performances are evaluated by comparing to the deep learning model without the MVNB layer and SVM models by using five-year data from Knox County in Tennessee.

3. The Modeling Framework Formulation

A novel model is proposed for the traffic crash prediction and Figure 1 illustrates the modeling framework. The proposed model includes two modules. One is the unsupervised feature learning module and another is the supervised fine tuning module. The obtained encoded feature representations from the unsupervised feature module are used as the input for the supervised fine tuning module.

3.1. Unsupervised Feature Learning Module. Compared to the commonly used deep learning architectures including deep belief network [31], stacked autoencoder [32], and convolutional neural networks [27], the symmetrical neural networks in an unsupervised manner have shown better performances, which can automatically learn an appropriate sparse feature representation from the raw data [33]. The unsupervised feature learning module includes a denoising autoencoder (DAE) model to learn the underlying structure of the dynamic pattern among the characteristics of roadway, traffic, and environment. With the explanatory variables, such as roadway geometric design features, traffic factors, pavement factors, and environmental characteristics as the input, the designed DAE model can identify the nonlinear relationship between the investigated variables in an unsupervised and hierarchical manner and the robust feature representations can be obtained. In addition, the designed DAE model can encode the explanatory variables into an embedding low-dimensional space. The proposed DAE model contains a visible layer, a hidden layer, an output layer, and a reconstruction layer. Unlike the conventional DAE model with K hidden layers [34], the proposed model uses a reconstruction error optimizing the output layer and the noisy input layer to generate higher level representations.

Assume $\Phi = \{[\mathbf{v}_i, \mathbf{u}_i]\}_{i=1}^{n}$ is a training set that contains n roadway entities, where $\mathbf{v}_i \in \mathfrak{R}^D$ is the explanatory variable vector with dimension D and $\mathbf{u}_i$ is the multivariate traffic safety outcomes for roadway entity i. Given the training set, the proposed DAE model is trained to develop a robust feature representation by reconstructing the
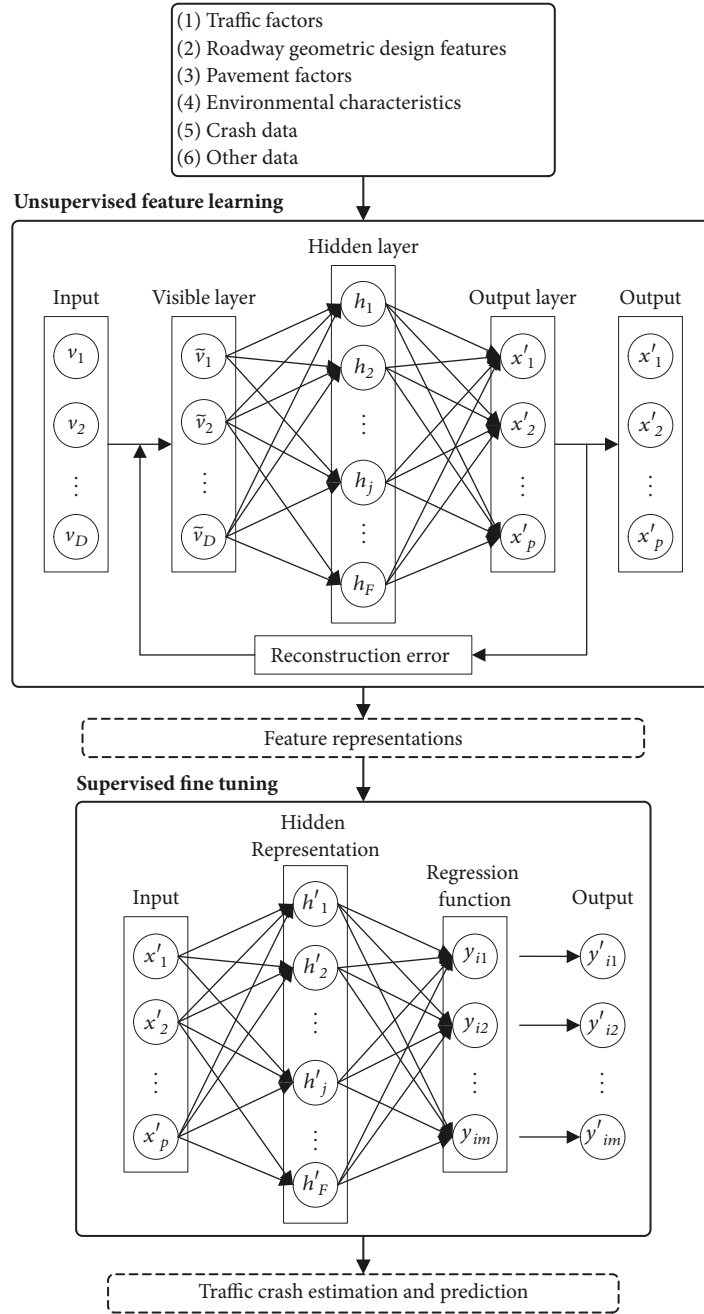
FIGURE 1: Flowchart of the proposed model for traffic crash prediction.

input $\mathbf{v}_i$ from its noisy corrupted version $\widetilde{v}_i$, as shown in Figure 1.

There are D units in the visible layer and F units in the hidden layer and the proposed model can be defined by a parameter set $\Theta = \{\mathbf{W}, \mathbf{a}, \mathbf{b}, \mathbf{c}\}$, where $\mathbf{W} = [W_{ij}] \in \mathfrak{R}^{D \times F}$ is the interlayer connection weights, $\mathbf{a} = [a_i] \in \mathfrak{R}^D$ is the visible self-interactions or biases, $\mathbf{b} = [b_j] \in \mathfrak{R}^F$ is the hidden biases, and $\mathbf{c} = [c_k] \in \mathfrak{R}^K$ is the reconstruction error. The joint probability distribution between the noisy input variables and

hidden variables is defined as

$$P(\mathbf{v}, \mathbf{h}; \Theta) = \frac{1}{Z(\Theta)} \exp\left[-\varepsilon(\mathbf{v}, \mathbf{h}; \Theta)\right] \qquad (1)$$

where $\varepsilon(\mathbf{v}, \mathbf{h}; \Theta)$ is an energy function defined by symmetric interactions between the noisy input variable, hidden variables, and a set of interaction parameters Θ; $Z(\Theta)$ is a normalized factor.

For a binary variable, a Bernoulli-Bernoulli energy function [34, 35] and the conditional distribution of a single stochastic hidden variable are given by

$$\varepsilon\left(\mathbf{v}, \mathbf{h}; \Theta\right) = \mathbf{v}^T \mathbf{W} \mathbf{h} - \mathbf{a}^T \mathbf{v} - \mathbf{b}^T \mathbf{h}$$

$$P\left(h_j = 1 \mid \mathbf{v}; \Theta\right) = \frac{1}{1 + \exp\left[-\left(b_j + \sum_{i=1}^{D} W_{ij} v_i\right)\right]} \qquad (2)$$

For a continuous variable, a Gaussian-Bernoulli energy function [31] and the conditional distribution of a single stochastic hidden variable are given by

$$\varepsilon\left(\mathbf{v}, \mathbf{h}; \Theta\right) = -\mathbf{v}'^T \mathbf{W} \mathbf{h} + \mathbf{a}'^T \mathbf{a}' - \mathbf{b}^T \mathbf{h}$$

$$P\left(h_j = 1 \mid \mathbf{v}; \Theta\right) = \frac{1}{1 + \exp\left[-\left(b_j + \sum_{i=1}^{D} \left(W_{ij} v_i / \sigma_i\right)\right)\right]} \qquad (3)$$

where $\mathbf{v}' = (v_1/\sigma_1, v_2/\sigma_2, \cdots, v_i/\sigma_i, \cdots v_D/\sigma_D)^T$, $\mathbf{a}' = ((v_1 - a_1)/\sigma_1, (v_2 - a_2)/\sigma_2, \cdots, (v_i - a_i)/\sigma_i, \cdots, (v_D - a_D)/\sigma_D)$, and σ_i is the standard deviation of the i-th visible variable v_i centered on the bias a_i.

The conditional distribution over hidden units can be factorized and computed by

$$P\left(\mathbf{h} \mid \mathbf{v}; \Theta\right) = \prod_{j=1}^{F} P\left(h_j \mid \mathbf{v}; \Theta\right) \qquad (4)$$

To estimate the parameters, the method proposed by Hjelm et al. [35], which maximizes the log-likelihood of the marginal distribution of the hidden units to find the gradient of the log-likelihood, is employed in this research. To simplify the estimation process, the free energy in terms of the probability at a data point $\mathbf{v}_n$ can be used to replace the energy function and the gradient has the following form:

$$\frac{\partial}{\partial \Theta} \sum_{\mathbf{v} \in D} \log p\left(\mathbf{v}\right) = -\sum_{\mathbf{v} \in D} \left\langle \frac{\partial}{\partial \Theta} \varphi\left(\mathbf{v}, \mathbf{h}; \Theta\right) \right\rangle_{p(\mathbf{h}|\mathbf{v};\Theta)}$$
$$+ \sum_{\mathbf{v} \in D} \left\langle \frac{\partial}{\partial \Theta} \varphi\left(\mathbf{v}, \mathbf{h}; \Theta\right) \right\rangle_{p(\mathbf{v},\mathbf{h}|\Theta)} \qquad (5)$$

where $\varphi(\mathbf{v}, \mathbf{h}; \Theta) = -\log \sum_h \exp[-\varepsilon(\mathbf{v} = \mathbf{v}_n, \mathbf{h}; \Theta)]$, and the conditional distribution over hidden units should be replaced by a loss function.

$$\varphi\left(\mathbf{h} \mid \mathbf{v}_n\right) = \begin{cases} -\mathbf{a}^T \mathbf{v}_n - \sum_{j=1}^{F} \log\left[1 + \exp\left(\mathbf{v}_n^T W_j + b_j\right)\right] & \mathbf{v}_n \text{ is binary} \\ -\left\| \mathbf{a} - \mathbf{v}_n \right\|^2 - \sum_{j=1}^{F} \log\left[1 + \exp\left(\mathbf{v}_n^T W_j + b_j\right)\right] & \mathbf{v}_n \text{ is continuous} \end{cases} \qquad (6)$$

Considering the hidden layer as the input layer for the outcome layer and outcome layer as the hidden layer, the joint probability distribution between the hidden variables and outcome variables, energy function, the conditional distributions of the outcome variables and units, and parameter estimation process can be obtained as those for hidden layer. The model will stop training when c_k satisfies the reconstruction error requirements or the dimension of the feature representation achieves the designed goals.

3.2. Supervised Fine Tuning Module. The supervised fine tuning module is a supervised fine tuning procedure that includes a regression layer on the top of the resulting hidden representation layers to estimate the likelihood of the crash occurrences, as shown in Figure 1. The obtained encoded feature representations from the unsupervised feature module are used as the input for the supervised fine tuning module. To jointly estimate the occurrence likelihood for more than one type of crashes simultaneously and address the potential heterogeneity issues in the interdependent crash data, a multivariate negative binomial (MVNB) model is used in the supervised fine tuning module to estimate and predict the traffic crashes across injury severities. Assume $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})'$ is a vector of crash occurrence likelihood for roadway entities i, which includes m types of crashes. The

particular NB regression model employed in this study has the following form:

$$p\left(y_{ij}\right)$$
$$= \frac{\Gamma\left(\lambda_{ij}/\sigma + y_{ij}\right)}{\Gamma\left(\lambda_{ij}/\sigma\right) \Gamma\left(y_{ij} + 1\right)} \left(\frac{1}{1 + \sigma}\right)^{\lambda_{ij}/\sigma} \left(\frac{\sigma}{1 + \sigma}\right)^{y_{ij}} \qquad (7)$$

where $\Gamma(\cdot)$ is the gamma function, y_{ij} is the crash number of crash type j for roadway segment i and $E[y_{ij}] = \lambda_{ij} = \exp(\boldsymbol{\beta}_j \mathbf{X}_{ij} + \varepsilon_{ij})$. $\exp(\varepsilon ij)$ is a multivariate gamma-distributed error term with mean 1 and variance α^{-1}.

As described in Shi and Valdez [36] and Anastasopoulos et al. [37], with the expected number of crashes λ_{ij}, the MVNB model has a joint probability function:

$$p\left(y_{1j}, y_{2j}, \cdots, y_{ij}\right)$$
$$= \sum_{k=0}^{\min(y_{1j}, y_{2j}, \cdots, y_{ij})} p\left(k\right) \prod_{i=1}^{N} p\left(y_{ij} - k\right) \qquad (8)$$

The model parameters can be estimated by maximizing the log-likelihood function:

$$L = \sum_{i,j} \left\{ \begin{array}{c} \log\left[\Gamma\left(\sigma^{-1} + y_{ij}\right)\right] - \log\left[\Gamma\left(\sigma^{-1}\right)\right] - \log\left(y_{ij}!\right) \\ -\sigma^{-1}\ln\left[1 + \sigma\exp\left(\boldsymbol{\beta}_j \mathbf{X}_{ij}\right)\right] - y_{ij}\log\left[1 + \sigma^{-1}\exp\left(\boldsymbol{\beta}_j \mathbf{X}_{ij}\right)\right] \end{array} \right\} \tag{9}$$

The MVNB regression layer is added on the top of the resulting hidden representation layers to perform traffic crash prediction. This yields a deep learning model tailored to a task-specific supervised learning. Then we fine tune the module 2 using backpropagation by minimizing the following cost function:

$$f(\boldsymbol{\theta})$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} s\left(y'_{ij} = y_{ij}\right)\log\left[\frac{\exp\left(h_\theta\left(x_i\right)\right)}{\sum_{l=1}^{F}\exp\left(h_\theta\left(x_i\right)\right)}\right] \tag{10}$$

$$+ \frac{\alpha}{2p}\left[\left\|W_{regression}\right\|_F^2 + \sum_{l=1}^{F}\left\|W_l\right\|_F^2\right]$$

where s(·) is an indicator function, if y_{ij}'=y_{ij}, then s(·)=1; otherwise, s(·)=0, α is a regularization parameter, and $\|\|_F$ is the Frobenius norm. The first term refers to the cross entropy loss for the regression layer, the second term is the weight decay penalty, and $h_\theta(x_i)$ is the output of deep learning for an input x_i'.

The cost function is minimized with a min-batch gradient descent algorithm [27]. The parameters in module 2 $\boldsymbol{\theta}$ = $\{\mathbf{W}_l, \mathbf{b}, \mathbf{W}_{regression}\}$ are estimated by initializing the weights $\mathbf{W}_{regression}$ of the regression layer to small random values and the weights of the F hidden layers are initialized by the encoding weights obtained in the unsupervised feature learning module.

4. Data

The data are obtained from the Tennessee Roadway Information Management System (TRIMS) and the Pavement Management System (PMS), which are maintained by the Tennessee Department of Transportation (TDOT). The dataset includes crash data, traffic factors, geometric design features, pavement factors, and environmental characteristics. The traffic, geometric, pavement, and environmental characteristics are linked to the crash data through the common variable *id_number*. An extensive and comprehensive data screening that includes cleaning, consistency, and accuracy checks is processed and performed to ensure the data are useable, reliable, and valid for the analyses. After the initial data screening, in total 635 roadway segments in Knox County are chosen for the analyses. For the selected roadway segments, each of them has a completed dataset that links to the crash data. In other words, the dataset contains detailed information on traffic factors, geometric design features, pavement characteristics, and environment factors.

In TRIMS, the crash data have been classified into five categories, fatal crashes, incapacitating injury crashes, nonincapacitating injury crashes, possible-injury crashes, and PDO crashes. Because the category of fatal crashes has only a few observations, the categories of fatal crashes and incapacitating injury crashes have been combined and referred to as major injury crashes. The possible-injury crashes and PDO crashes have been combined and referred to as no-injury crashes. The nonincapacitating injury crashes are referred to as minor injury crashes. A few of pervious literature [38–40] has used a similar classification for injury outcomes. For those selected 635 roadway segments, from 2010 to 2014, a total of 5365 traffic crashes were reported by the police officers, which include 135 (2.51%) major injury crashes, 1312 (24.46%) minor injury crashes, and 3917 (73.02%) PDO crashes. Individual roadway segment experienced from 0 to 23 crashes per year with a mean of 1.54 and a standard deviation of 1.89. As expected, a significant amount of zeros is observed. The dependent variables and their descriptive statistics are shown in Table 1. The descriptive statistics of continuous independent variables and categorical independent variables are shown in Tables 1 and 2, respectively.

The considered traffic factors include the logarithm of annual average daily traffic (AADT) per lane, truck traffic percentage, and posted speed limits. Roadway segment AADT per lane from 2010 to 2014 varies from 851 to 32,359 vehicles with a mean of 3,388.44 and a standard deviation of 5495.41. Other than the TRIMS dataset maintained by the DDOT, traffic flow information can be obtained from https://www.tdot.tn.gov/APPLICATIONS/traffichistory, which is an AADT map providing traffic volumes based on a 24-hour, two-directional count at a given location. The website also provides the traffic history of any specific count station. The variable of posted speed limit has a mean of 38.65 and a standard deviation of 6.69 with a minimum value of 30 and a maximum value of 70. The truck traffic percentage varies from 1 to 33 with a mean of 6.71 and a standard deviation of 4.98.

Important measurements of geometric design features considered in this study include segment length, degree of horizontal curvature, median widths, outsider shoulder widths, number of through lanes, lane widths, number of left-turn lanes, median types, and shoulder type. Among them, the segment length, degree of horizontal curvature, median widths, and outsider shoulder widths are considered as the continuous variables and the others are considered as the categorical variables. Other than the traffic factors and geometric design features, the impacts of pavement surface characteristics are considered to better address traffic safety issues for roadway design and maintenance. The considered pavement surface characteristics include international roughness index (IRI) and rut depth (RD). The analyzed IRI varies from 25.45 to 182.58 with a mean of 65.85 and a standard deviation of 27.75, which is calculated

TABLE 1: Summary statistics of analyzed continuous variables.

Variable	Mean	Std. Dev.	Min.	Max.
Independent variable				
The number of major injury crashes per year per roadway segment	0.04	0.32	0	3
The number of minor injury crashes per year per roadway segment	0.41	0.54	0	9
The number of no injury crashes per year per roadway segment	1.23	1.74	0	22
Traffic factors				
The logarithm of AADT per lane	3.53	3.74	2.93	4.51
Truck traffic percentage	6.71	4.98	1	33
Posted speed limits	38.65	6.69	30	70
Geometric design features				
Segment length (miles)	0.81	1.03	0.02	12.31
Degree of horizontal curvature	1.51	3.23	0	14.00
Median widths	1.12	2.02	0	12
Outside shoulder widths	3.06	1.88	3.52	8
Pavement factors				
International roughness index	65.85	27.75	25.45	182.58
Rut depth (in.)	0.13	0.05	0.06	0.41

TABLE 2: Summary statistics of analyzed categorical variables.

Variable	Category	Frequency	Percent
Geometric design features			
Number of through lanes	3	585	18.43
	2	1660	52.28
	1	930	29.29
Lane widths (ft)	12	915	28.82
	11	1650	51.97
	10	610	19.21
Number of left-turn lanes	2	920	28.98
	1	650	20.47
	0	1605	50.55
Median type	1 for non-traversable median	1185	37.32
	0 for traversable median	1990	62.68
Shoulder type	2 for pavement	760	23.94
	1 for gravel	1595	50.24
	0 for dirt	820	25.83
Environmental factors			
Terrain type	1 for mountainous	1206	37.97
	0 for rolling	1969	62.03
Land use type	2 for residential	1622	51.10
	1 for commercial	775	24.41
	0 for rural	778	24.49
Indicator for lighting	1 for lighting exists on the roadway segments	1304	41.06
	0 for others	1871	58.94

using a quarter-car vehicle math model and the response is accumulated to yield a roughness index with units of slope (in/mi). Another pavement condition indicator is the RD, which is measured at roadway speeds with a laser/inertial profilograph. The analyzed RD varies from 0.06 to 0.41 with a mean of 0.13 and a standard deviation of 0.05.

The environmental factors, including terrain types, lighting condition, and land use type, are considered. Two terrain types are examined, which include rolling terrace (62.03%) and mountainous terrace (37.97%). Lighting condition was considered as a category variable, which indicated whether lighting devises are provided at the roadway segments.

Three types of land use are considered, including commercial (24.41%), rural (24.49%), and residential (51.10). These variables are considered because they might have potential significant effects on traffic safety.

5. Modelling Results

The MATLAB was employed for model development. Four-year data, from 2010 to 2013, were used as the training set and one-year data, the year of 2014, were used as the testing set. In order to obtain the model with superior performance, module 1 was developed using 9 Gaussian visible units, 13 binary visible units, and a number of hyperbolic tangent hidden units ranging from 32 to 128 in steps of 2. The number of hidden units was setting based on two rules, greater than the input data dimensionality and the powers of two. The parameters for learning rate and weight decay were selected to optimize reduction of reconstruction error over training. Module 1 was trained with a sample size of 2540 (four-year crash data) to allow for full convergence of the parameters. The input data were processed by using module 1 to capture the relationship between traffic factors, geometric design features, pavement conditions, and environmental characteristics. Module 2 was developed using 4 Gaussian visible units and a number of hyperbolic tangent hidden units. The initial number of hyperbolic tangent hidden units ranging from 8 to 32 was tested and examined. The learning procedure will stop when the number of feature representations is achieved to four or the reconstruction error is less than 0.01.

5.1. Results of Unsupervised Feature Learning Module. For the final model, the reconstruction error is less than 0.01 and four hidden layers are included. The weights between the input layer and output layer can be calculated as $W_{32\times16}W_{16\times8}W_{8\times4}=W_{32\times4}$. The results are shown in Table 3. The negative sign represents a crash-prone condition and a positive sign represents a safe-prone condition. The valued number indicates an evaluation score. Because the feature learning module identifies relational information between input variables and output feature representations, the connection weights between visible units and hidden units can be interpreted as the functional networks [35]. The results show that each of the output units is significant associated with traffic factors, geometric factors, pavement factors, and environmental factors.

Since the proposed feature learning module has symmetric connection between visible and hidden layers and the units in both layers have the probabilistic characteristics, the proposed feature learning module also can be called as an auto-encoder. Thus, the values of output units can be interpreted as a feature representation and the results are shown in Table 4. The number of output units is smaller than the number of visible units, which indicates the feature representation with the values of output units has reduced the dimensionality of the input, but still preserving the original information. Hence the four output units are defined as traffic feature representation, geometric feature representation,

pavement feature representation, and environmental feature representation.

The results show that the signs of the means of four feature representations are negative, which indicate that the feature representations are associated with crash-prone condition. In other words, the current traffic, geometric, pavement, and environmental features are the main factors that attribute to the risk of crash occurrences. The findings indicate that the traffic, geometric, pavement, and environmental factors have a direct influence on traffic safety and need to be improved.

The traffic feature representations have a wide range with a minimum value of -4.420 and a maximum value of 6.650, which indicate the traffic factors have a significant impact on traffic safety. The pavement factors have comparative impacts on traffic safety with a minimum value of -12.076 and a maximum value of 6.753. The ranges of geometric and environmental feature representations are from -1.386 to 1.280 and from -1.805 to 0.847, respectively, which indicate that the geometric and environmental feature representations have comparative effects on traffic safety. Compared to those crash modeling techniques that use only geometric design features as the input factors, the current research reveals new insights that would benefit the development of updated guidelines.

5.2. Results of the Supervised Fine Tuning Module. For the supervised fine tuning module, the visible units of the input layer use the feature representations as the input and the crash counts across injury severities are used as the training target. The aggregate weights between the input layer and output layer are shown in Table 5. The results show that the traffic and geometric feature representations have positive effects and pavement and environmental feature representations have negative effects on traffic crashes across injury severities. The findings indicate that decreasing the values of traffic and geometric feature representations will increase the likelihood of crashes and increasing the values of geometric and environmental feature representations will increase the likelihood of crashes. The comparison results show that traffic feature representation and geometric feature representation have significant impact on PDO crashes. The pavement feature representation and environmental feature representation have significant impact on minor injury crashes. Among four feature representations, the geometric feature representation, pavement feature representation, and traffic feature representation have most direct impacts on major injury, minor injury, and PDO crashes, respectively.

Considering the data in the year of 2014 as the input variables, the developed deep learning model is used to predict the crash counts for the year of 2014. To validate the superiority of the proposed models, the predicted results are compared to the observed values. In addition, a deep learning model without the regression layer and a support vector machine (SVM) model are also developed to predict crash counts across injury severities. There are two key issues related to the development of SVR models, kernel selection, and parameters optimization. To address the nonlinear relationship between the outcomes and attributes, the commonly

TABLE 3: The functional network between the input variables and the feature representations.

Variable	Traffic feature representation	Geometric feature representation	Pavement feature representation	Environmental feature representation
Traffic factor				
The logarithm of AADT per lane	-0.529	-0.007	-0.016	-0.034
Truck traffic percentage	-0.659	-0.002	-0.024	-0.010
Posted speed limits	-0.030	-0.010	0.175	-0.047
Geometric factors				
Segment length (miles)	0.007	-0.580	-0.007	-0.003
Degree of horizontal curvature	0.002	-0.918	-0.006	-0.009
Median widths	0.002	0.826	0.068	0.015
Outside shoulder widths	0.005	0.419	0.016	0.064
Number of through lanes				
3 lanes	-0.021	-0.891	-0.006	-0.026
2 lanes	-0.015	-0.864	-0.022	-0.060
Lane widths				
12 ft	0.004	0.776	0.005	0.001
11 ft	0.009	0.592	0.076	0.004
Number of left-turn lanes				
2 left-turn lanes	-0.003	-0.212	-0.019	-0.017
1 left-turn lane	0.002	-0.783	0.028	0.016
Median type				
Non-traversable median	0.013	0.247	0.006	0.007
Shoulder type				
Pavement	-0.035	-0.144	-0.027	-0.016
Gravel	0.005	-0.743	0.011	0.060
Pavement factor				
International roughness index (in./mile)	0.101	0.007	-0.183	-0.001
Rut depth (in.)	-0.004	-0.004	-0.341	-0.038
Environmental factor				
Terrain type				
Mountainous	-0.014	-0.011	-0.010	-0.954
Land use type				
Residential	0.042	0.015	0.011	0.847
Commercial	0.058	0.001	0.006	-0.915
Indicator for lighting				
Lighting exists on roadway	0.018	0.008	0.027	0.781

Table 4: Summary statistics of feature representations.

Statistics	Mean	Std. Dev.	Min.	Max.
Traffic feature representation	-0.697	1.464	-4.420	6.650
Geometric feature representation	-1.620	0.521	-1.386	1.280
Pavement feature representation	-5.388	2.441	-12.076	6.753
Environmental feature representation	-2.167	0.442	-1.805	0.847

Table 5: The functional network between the feature representation and different crash types.

Variable	Major injury crashes	Minor injury crashes	PDO crashes
Traffic feature representation	-0.109	-0.421	-1.535
Geometric feature representation	0.129	0.261	0.539
Pavement feature representation	-0.033	-0.872	-0.315
Environmental feature representation	0.113	0.298	0.146

used RBF kernel is chosen to develop the SVM model. To precisely reflect the performance on regressing unknown data and prevent the overfitting problem, the k-fold cross-validation approach is employed for optimizing the two parameters in RBF kernels—C and ε [41, 42]. The optimized parameters are (42.30, 37.09) and the optimized SVM model yields to a training mean absolute error (MAE) less than 0.01 and a training root-mean-squared deviation (RMSD) less than 5%.

To evaluate the accuracy and reliability of the prediction results, the predicted means for each injury severity level by the proposed model, the deep learning model without the regression layer, and the SVM model are compared to the observed means and two evaluation measures are used, which include MAE and RMSD. The comparison and validation results are shown in Table 6. Results in Table 6 indicate that the predicted means from the proposed model (0.110, 0.573, 1.335, and 2.019 for major injury, minor injury, no-injury, and all crashes, respectively) are very close to the observed means (0.096, 0.556, 1.306, and 1.986). We believe that a very importation feature of the proposed model is that the included regression layer can provide a good estimate of the chance that the roadway segment is in the crash-free state or some crash-prone propensity states.

For all the observed samples, the proposed model results in a MAE of 0.030, 0.080, 0.071, and 0.150 and a RMSD of 17.298%, 29.961%, 27.206%, and 43.652% for major injury, minor injury, PDO, and all crashes, respectively. The deep learning mode without the regression layer results in a MAE of 0.043, 0.202, 0.405, and 0.520 and a RMSD of 20.620%, 51.741%, 65.086%, and 82.862% for major injury, minor injury, PDO, and all crashes, respectively. The SVM model results in a MAE of 0.055, 0.257, 0.471, and 0.660 and a RMSD of 26.022%, 61.350%, 72.416%, and 96.636% for major injury, minor injury, PDO, and all crashes, respectively. The results suggest that the proposed model predicts better than the deep learning model without the regression layer and the SVM model. In summary, compared to the deep learning model without the regression layer and the SVM model, the

proposed model results in the smallest prediction MAE and RMSD, no matter for the crash types across injury severities. We hypothesize that the proposed model better addresses the issue of heterogeneity and allows for excess zero counts in correlated data.

The findings indicate that the predictions from the proposed model have significant improvements over all comparison models, in both accuracy and robustness. The best-performing result of the proposed model for major injury crashes has a MAE of 0.030, which indicates an 42.105% and 84.211% improvement from the deep learning model without the regression layer and the SVM model, respectively. The proposed models perform worse for the minor injury crashes, with a 0.080 RMSD for all observed samples. However, it is still better than the evaluation measurements from the deep learning model without the regression layer and the SVM model, which are 0.202 and 0.257, respectively. This represents a MAE improvement of 150.980% and 219.608% for minor injury crash prediction. For the PDO crashes, the MAE improvements of the proposed models over the comparison models range from 471.111% to 564.444%. For all the observed samples, the MAE improvements range from 247.368% to 341.053%. Clearly, the improvement is significant for the traffic crash predictions. Therefore, the proposed model seems to be a better alternative for crash count predictions.

The proposed model has better performances in terms of small error variances than the comparison models, since the regression model is imbedding into the proposed model. The overall performances of the proposed model for all crashes show an 89.824% RMSD improvement over the deep learning mode without regression layer and an 121.378% RMSD improvement over the SVM model. It is clear that the predictions obtained from the proposed models are superior to those obtained from the comparison models. The greatest difference is demonstrated for the PDO crashes where the proposed model yields a RMSD of 27.206% compared to a 65.086% RMSD value from the deep learning model without the regression layer and a 72.416% RMSD value from the

TABLE 6: Results of roadway segment crash count prediction.

	Major injury	Minor injury	No injury	Total
Observation mean	0.096	0.556	1.306	1.986
Proposed model				
Estimated mean	0.110	0.573	1.335	2.019
MAE	0.030	0.080	0.071	0.150
RMSD (%)	17.298	29.961	27.206	43.652
Deep learning model without the regression layer				
Estimated mean	0.101	0.556	1.332	1.989
MAE	0.043	0.202	0.405	0.520
RMSD (%)	20.620	51.741	65.086	82.862
SVM model				
Estimated mean	0.101	0.513	1.329	1.943
MAE	0.055	0.257	0.471	0.660
RMSD (%)	26.022	61.350	72.416	96.636

SVM model. The differences in the proposed model and the SVM model for the minor injury crashes are also significant (29.961% versus 61.350%).

6. Conclusions

Because traffic crashes are a big concern of the public, agencies, and policy makers and result in countless fatalities and injuries, there is a need to perform a comprehensive analysis that aims to understand the relationship between the influence factors and traffic crash outcomes. In this study, we presented an innovative approach for traffic crash prediction. Methodologically, we demonstrated a novel deep learning technique embedded within a multivariate regression model can be used to identify relationship between the examined variables and the traffic crashes. Future applications of this approach have the potential to provide insights into basic questions regarding roadway spatial and temporal dynamic function and practical questions regarding countermeasures. The investigation results provide sufficient evidence for the following conclusions:

(1) The results show that the feature learning module identifies relational information between input variables and output feature representations. The findings indicate that the feature representations have reduced the dimensionality of the input, but still preserving the original information.

(2) The proposed model that includes a MVNB regression layer in the supervised fine tuning module can better account for different patterns in crash data across injury severities and provide superior traffic crash predictions. In addition, the proposed model can perform multivariate estimations simultaneously with a superior fit.

(3) The proposed model has superior performances in terms of prediction power compared to the deep learning model without a regression layer and the SVM model. The overall performances of the proposed model for all crashes show an 89.824% RMSD improvement over the deep learning model without a regression layer and an 121.378% RMSD improvement over the SVM model.

(4) The findings suggest that the proposed model is a superior alternative for traffic crash predictions. The proposed model can better account for heterogeneity issues in traffic crash prediction.

The proposed models can perform traffic crash prediction for a given facility. The proposed methodology could be applied to other roadway networks if appropriate attribute variables are available. Traffic and transportation engineering agencies can employ the proposed models with relative cases and develop them to their needs to obtain traffic crash predictions for various time periods. Further investigation of the proposed models includes the predictions of spatial-temporal dynamic pattern in crash data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] NHTSA, *Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2012-2014*, Department of Transportation, National Highway Traffic Safety Administration, Traffic Safety Facts, DOT HS 812 297, 2016.

[2] NHTSA, *2015 Motor Vehicle Crashes: Overview*, U.S. Department of Transportation, National Highway Traffic Safety Administration, Traffic Safety Facts, DOT HS 812 318, 2016.

[3] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.

[4] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: methodological frontier and future directions," *Analytic Methods in Accident Research*, vol. 1, pp. 1–22, 2014.

[5] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Analytic Methods in Accident Research*, vol. 11, pp. 1–16, 2016.

[6] J. C. Milton, V. N. Shankar, and F. L. Mannering, "Highway accident severities and the mixed logit model: an exploratory empirical analysis," *Accident Analysis & Prevention*, vol. 40, no. 1, pp. 260–266, 2008.

[7] C. Dong, D. B. Clarke, X. Yan, A. Khattak, and B. Huang, "Multivariate random-parameters zero-inflated negative binomial regression model: an application to estimate crash frequencies at intersections," *Accident Analysis & Prevention*, vol. 70, pp. 320–329, 2014a.

[8] C. Dong, D. B. Clarke, S. S. Nambisan, and B. Huang, "Analyzing injury crashes using random-parameter bivariate regression models," *Transportmetrica A: Transport Science*, vol. 12, no. 9, pp. 794–810, 2016.

[9] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: differences, similarities and some insights," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 3, pp. 387–399, 2011.

[10] J. C. Principe, N. R. Euliano, and W. C. Lefebvre, *Neural and Adaptive Systems: Fundamentals through Simulations*, vol. 672, Wiley, New York, NY, USA, 2000.

[11] L.-Y. Chang, "Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network," *Safety Science*, vol. 43, no. 8, pp. 541–557, 2005.

[12] H. T. Abdelwahab and M. A. Abdel-Aty, "Artificial neural networks and logit models for traffic safety analysis of toll plazas," *Transportation Research Record*, no. 1784, pp. 115–125, 2002.

[13] Y. Xie, D. Lord, and Y. Zhang, "Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis," *Accident Analysis & Prevention*, vol. 39, no. 5, pp. 922–933, 2007.

[14] M. M. Kunt, I. Aghayan, and N. Noii, "Prediction for traffic accident severity: Comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods," *Transport*, vol. 26, no. 4, pp. 353–366, 2011.

[15] K. S. Jadaan, M. Al-Fayyad, and H. F. Gammoh, "Prediction of Road Traffic Accidents in Jordan using Artificial Neural Network (ANN)," *Journal of Traffic and Logistics Engineering*, vol. 2, no. 2, pp. 92–94, 2014.

[16] D. Akin and B. Akbaş, "A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics," *Scientific Research and Essays*, vol. 5, no. 19, pp. 2837–2847, 2010.

[17] X. Li, D. Lord, Y. Zhang, and Y. Xie, "Predicting motor vehicle crashes using Support Vector Machine models," *Accident Analysis & Prevention*, vol. 40, no. 4, pp. 1611–1618, 2008.

[18] Y. Zhang and Y. Xie, "Forecasting of short-term freeway volume with v-support vector machines," *Transportation Research Record*, vol. 2024, no. 1, pp. 92–99, 2007.

[19] V. Kecman, "Support Vector Machines – An Introduction," in *Support Vector Machines: Theory and Applications*, vol. 177 of *Studies in Fuzziness and Soft Computing*, pp. 1–47, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[20] R. Yu and M. Abdel-Aty, "Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data," *Safety Science*, vol. 63, pp. 50–56, 2014.

[21] C. Chen, G. Zhang, Z. Qian, R. A. Tarefder, and Z. Tian, "Investigating driver injury severity patterns in rollover crashes using support vector machine models," *Accident Analysis & Prevention*, vol. 90, pp. 128–139, 2016.

[22] N. Dong, H. Huang, and L. Zheng, "Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects," *Accident Analysis & Prevention*, vol. 82, pp. 192–198, 2015.

[23] G. Ren and Z. Zhou, "Traffic safety forecasting method by particle swarm optimization and support vector machine," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10420–10424, 2011.

[24] R. Yu and M. Abdel-Aty, "Utilizing support vector machine in real-time crash risk evaluation," *Accident Analysis & Prevention*, vol. 51, pp. 252–259, 2013.

[25] M. Abdel-Aty and K. Haleem, "Analyzing angle crashes at unsignalized intersections using machine learning techniques," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 461–470, 2011.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[27] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[28] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.

[29] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic Flow Prediction with Big Data: A Deep Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2015.

[30] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.

[31] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[32] P. Swietojanski, A. Ghoshal, and S. Renals, "Convolutional neural networks for distant speech recognition," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1120–1124, 2014.

[33] M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[34] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," *NeuroImage*, vol. 129, pp. 292–307, 2016.

[35] R. D. Hjelm, V. D. Calhoun, R. Salakhutdinov, E. A. Allen, T. Adali, and S. M. Plis, "Restricted Boltzmann machines for neuroimaging: An application in identifying intrinsic networks," *NeuroImage*, vol. 96, pp. 245–260, 2014.

[36] P. Shi and E. A. Valdez, "Multivariate negative binomial models for insurance claim counts," *Insurance: Mathematics & Economics*, vol. 55, pp. 18–29, 2014.

[37] P. C. Anastasopoulos, V. N. Shankar, J. E. Haddock, and F. L. Mannering, "A multivariate tobit analysis of highway accident-injury-severity rates," *Accident Analysis & Prevention*, vol. 45, pp. 110–119, 2012.

[38] L. Chang and J. Chien, "Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model," *Safety Science*, vol. 51, no. 1, pp. 17–22, 2013.

[39] J. Pahukula, S. Hernandez, and A. Unnikrishnan, "A time of day analysis of crashes involving large trucks in urban areas," *Accident Analysis & Prevention*, vol. 75, pp. 155–163, 2015.

[40] Q. Wu, F. Chen, and G. H. Zhang, "Mixed logit model-based driver injury severity investigations in single- and multi-vehicle crashes on rural two-lane highways," *Accident Analysis & Prevention*, vol. 72, pp. 105–115, 2014.

[41] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge,UK, 2000.

[42] C. Campbell, "Kernel methods: A survey of current techniques," *Neurocomputing*, vol. 48, pp. 63–84, 2002.