Accident Analysis and Prevention 154 (2021) 106090

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A deep learning based traffic crash severity prediction framework

Md Adilur Rahim *, Hany M. Hassan

Department of Civil and Environmental Engineering, Louisiana State University, Patrick Taylor Hall, Baton Rouge, LA, 70803, USA

ARTICLE INFO

Keywords:
Traffic collision/accident severity
Deep learning
Transfer learning
Numeric to image transformation
Customized loss function

ABSTRACT

Highway work zones are most vulnerable roadway segments for congestion and traffic collisions. Hence, providing accurate and timely prediction of the severity of traffic collisions at work zones is vital to reduce the response time for emergency units (e.g., medical aid), accordingly improve traffic safety and reduce congestion. In predicting the severity of traffic collisions, previous studies used different statistical and machine learning models with accuracy as the main evaluating factor. However, the performance of these models was generally not good, especially on fatal and injury crashes. Also, looking into the prediction accuracy only is misleading.

This paper aims to propose a novel deep learning-based approach with a customized f1-loss function to predict the severity of traffic crashes. Underlying this objective is to compare the results of deep learning models with machine learning model considering two performance indicators, namely precision, and recall. The data used in the analysis include a sample of traffic crashes that occurred at work zones in Louisiana from 2014 to 2018. This dataset includes valuable information (features) related to road, vehicle, and human factors affecting the occurrence and severity of those crashes.

The proposed methodology is based on transforming these features/variables into images. Image transformation is conducted using a nonlinear dimensionality reduction technique t-SNE and convex hull algorithm. A CNN based deep learning algorithm with a customized loss function was used to directly optimize the model for precision and recall. The results showed improved performance in predicting the crash severity of fatal and injury crashes using the deep learning approach, which can help to improve traffic safety as well as traffic congestion at work zones and possibly other roadways segments.

1. Introduction

In the transportation system, work zone is an area of a roadway with construction, maintenance, or utility-related work activities. Driving through work zones area is challenging for drivers due to uneven pavement, lane closures, slow-moving heavy equipment, concrete barriers, etc. (Texas Department of Transportation, 2020; Turner, 1999). As a result, work zones have significant impacts on traffic operation and roadway safety. Work zones also exacerbate the congestion problem in peak hours and account for an estimated 10 percent of overall congestion and 24 percent of unexpected freeway delays. In addition, fatal crashes in work zones increased by 3 percent between 2016–2017 in the US, where fatal crashes outside of work zones decreased by 1.5 percent (Federal Highway Administration, 2019, March 25). So, accurate and timely prediction of traffic accident severity in work zones is a critical step in traffic collisions management.

To measure the performance of crash severity model, many previous studies used accuracy as the evaluation metric or used a loss function (e.g., Cross-entropy loss) that is optimized for accuracy (Abdelwahab and Abdel-Aty, 2001; Chang and Wang, 2006; Chen et al., 2016; Z. Li et al., 2012; Yu and Abdel-Aty, 2013; Zheng et al., 2019). Since traffic crash data is highly imbalanced, looking into only the prediction accuracy component is misleading (Bekkar et al., 2013; Guo et al., 2008; Tharwat, 2018). In an imbalanced dataset, accuracy places more weight on the common classes than on rare classes. Therefore, the model's performance gets worse in rare classes (e.g., fatal crashes). To overcome this problem, several other metrics can be used to better evaluate the model's performance efficiently. A lot of recent papers used different metrics to handle the crash data imbalance problem (Elamrani Abou Elassad et al., 2020a, b; Fiorentini and Losa, 2020; Jeong et al., 2018; Peng et al., 2020; Yahaya et al., 2019). These metrics include precision and recall that penalizes a model for ignoring the minority classes (Ting, 2010). From the literature, it is well established that, the precision of a class represents the trustworthiness of the model to answer if a point

* Corresponding author.
E-mail addresses: mrahim6@lsu.edu (M.A. Rahim), hassan1@lsu.edu (H.M. Hassan).

https://doi.org/10.1016/j.aap.2021.106090
Received 21 September 2020; Received in revised form 9 February 2021; Accepted 10 March 2021
Available online 16 March 2021
0001-4575/

belongs to that class, and the recall of a class represents the model's detection ability of that class (Rocca, 2019, Jan 27). A high recall and precision value mean the model handled the class appropriately, and a low recall and precision value mean the model handled the class poorly.

Several previous studies used different statistical and machine learning algorithms to predict the severity of traffic accidents, but very few had used a CNN based deep learning algorithm for the prediction task (Abdelwahab and Abdel-Aty, 2001; Chang and Wang, 2006; Chen et al., 2016; Z. Li et al., 2012; Yu and Abdel-Aty, 2013; Zheng et al., 2019; Yu et al., 2020). Most of the prior crash severity prediction models still requires improvements in terms of balanced precision and recall outputs. For example, improvement in precision does not always signify the applicability of models in real-time as the model may end up with lower recall value, which means lower predictability in fatal crashes. In this paper we focused on the recall value of the fatal crashes more because we wanted to avoid our model wrongly predicting fatal crashes, as injury crashes or property damage only (PDO) crashes. In a situation if our model has lower precision, our model wrongly predicting PDO or injury crashes as fatal crashes, it will not develop a life-threatening scenario. Therefore, this approach is more optimum for crash severity prediction.

In this paper we used deep learning, which is an emerging field that has been applied lately in many research fields with very promising results (Goodfellow et al., 2016; LeCun et al., 2015; Minar and Naher, 2018). In several classification applications, deep learning models surpassed human-level accuracy (Gibbons et al., 2017; He et al., 2015; Russakovsky et al., 2015). A deep learning model that is used for classification has two parts; a feature extractor that extracts useful information from the data using CNN layers automatically, and a classifier that consists of a fully connected layer classifies the data based on the information collected from the feature extractor. Although CNN based deep learning model shows excellent performance, it requires image-based data to perform the convolutional operation (Al-Saffar et al., 2017; Jamro and Wiatr, 2001). Since the crash data is numeric, we have utilized a generalized numeric to image transformation technique developed by Sharma et al. (2019). This procedure showed good performance on different tasks, such as classifying genomics, vowels, and text dataset. Another popular strategy that we have adopted in our research is transfer learning, which is to utilize knowledge from a well-established model (Pan and Yang, 2010; Quattoni et al., 2008). Many research studies showed that transfer learning could significantly improve a model's performance (Benjamin et al., 2016; Maqsood et al., 2019).

This paper aims to contribute to the literature and minimizes the shortcomings of prior crash severity prediction models by developing a novel deep learning approach. To simultaneously optimize the model on precision and recall, we used a customized f1-loss function, which was originally written by Haltuf (Haltuf, 2018). A transfer learning approach was applied to replace the feature extractor of the deep learning model with a pre-trained model. To combine the features of crash data and prepare the input for deep learning model, a generalized numeric to image transformation technique was selected (Sharma et al., 2019). This technique can take all the crash variables into account while transforming the data into images. It first applies a dimensionality reduction technique to project the features into a 2D plane and then converts the Cartesian coordinates to pixel locations. Using these pixel locations of the crash features, every crash can be converted into an image that is unique for that crash. For comparison purposes, the results of the deep learning model were compared with the results of a machine learning model (support vector machine).

The main contributions of this paper are summarized below:

1 Utilize a generalized numeric to image transformation technique that transforms the crash data into images.
2 Develop a novel deep learning-based approach with a customized f1-loss function to predict the severity of traffic crashes, especially predicting the most important classes of crash severity correctly (e.g., fatal crashes).
3 Compare the results of deep learning models with support vector machine models using two performance indicators, namely precision and recall.

The remainder of this paper is organized as follows: Section 2 discusses the literature review. Section 3 describes the proposed deep learning model and loss function, the method for numeric to image conversion, and deep learning inference procedure. Section 4 presents the results of the analysis. Finally, section 5 completes this paper with conclusions and recommendations.

2. Background

Prediction models of traffic collisions severity can be classified into two main categories: Statistical Learning & Machine Learning. Many researchers have used statistical learning models widely. Li et al. (2012) used the Support Vector Machine (SVM) and Ordered Probit (OP) model for crash injury severity analysis and found that the SVM model outperformed the OP model in terms of accuracy. Yu and Abdel-Aty (2013) utilized a Classification and Regression Tree (CART) model to select the most important explanatory variables and then used the result to compare Bayesian logistic regression models and SVM models with different kernel functions. ROC curve (AUC) metric was used for evaluation, and the SVM model with Radial-basis kernel function outperformed logistic regression models. An important outcome of the result was the use of a variable selection procedure prior to the model estimation. Chen et al. (2016) investigated driver injury severity patterns in rollover crash using SVM models. The CART model was utilized to identify significant variables, and the result showed that SVM models have reasonable prediction performance with the polynomial kernel, which outperformed the Gaussian radial basis kernel. Alkheder et al. (2017) used an artificial neural network (ANN) algorithm to predict traffic accident severity and compared its performance with an ordered probit model. A k-means algorithm was used to cluster the dataset into three clusters to improve the ANN model's performance. Their result showed an improved accuracy of 74.6 % achieved by the ANN model compared to the accuracy of 59.5 % achieved by the Ordered Probit model. AlMamlook et al. (2019) compared different machine learning algorithms performance on predicting traffic accident severity and found that the Random Forest algorithm (75.5 % accuracy) outperformed Logistic Regression (74.5 % accuracy), Naïve Bayes (73.1 % accuracy), and AdaBoost (74.5 % accuracy) algorithm.

In recent years, deep learning models have dramatically improved state-of-the-art visual object recognition, object detection, speech recognition, and many other domains (LeCun et al., 2015). Deep learning models have brought revolutionary advances to computer vision applications. It has been successfully implemented in different image classification applications and competitions with high accuracy. Al-Zuhairi and Pradhan (2017) used a Long-Short Term Memory (LSTM) model to predict the severity of traffic accidents. The result of the LSTM model was compared with the Multilayer Perceptron (MLP) and Bayesian Logistic Regression (BLR) model. It showed that the LSTM model (71.77 % accuracy) achieved better accuracy than MLP (65.48 % accuracy) and BLR models (58.30 % accuracy). Zheng et al. (2019) were among the first few studies to use a deep learning approach (Convolutional Neural Network) to predict traffic collision severity. A feature matrix to gray image (FM2GI) algorithm was proposed in that study to convert traffic accident data into gray images. The images were then used as input in the image classification model, and cross-entropy loss was used with Adam optimizer to optimize the model. To deal with the imbalanced dataset, a synthetic minority oversampling technique (SMOTE) was used. The performance of the novel deep learning model was compared with nine different models from literature, and the results showed that this novel model outperformed other models on the precision value of the fatal class. Bao et al. (2019) investigated a deep learning approach to predict short-term crash risk. A spatiotemporal convolutional long short-term memory network (STCL-Net) was proposed. The result of the deep learning model was compared with four econometric models and four machine learning models. The comparative analysis showed that the STCL-Net model outperformed the benchmark models in MSE, MAE, and MAPE metrics. Huang et al. (2020) utilized deep learning models in solving traffic safety problems by detecting and estimating crash risks. Convolutional neural networks (CNN) with dropout operation performed better in detecting crashes than shallow models. The deep learning models with simple structures performed better than a complex model. Li et al. (2020) proposed a long short-term memory convolutional network (LSTM-CNN) to predict real-time crash risk on arterials. The model outperformed other methods such as LSTM, CNN, XGBoost, Bayesian Logistic Regression, etc., in terms of sensitivity and false alarm rate. The model also showed a higher AUC value of 0.93. Formosa et al. (2020) presented a new deep learning model for predicting traffic conflicts. The video data was processed by an image processing technique aided by R-CNN. A deep neural network (DNN) was developed that can predict traffic conflicts in real-time. The model showed high prediction accuracy, sensitivity, and low false alarm rate. The best developed DNN model showed 94 % accuracy. Jiang et al. (2020) proposed a long short-term memory (LSTM) based framework to detect crashes. Six LSTM models for different temporal resolutions (DTR) were developed. The LSTMDTR model developed for one freeway can be used to predict crashes on a similar freeway. The model's performance was compared with different machine learning models and showed better prediction performance. The number of neurons used in the model affected the performance and computation time, with prediction accuracy ranging from 63.48% to 70.43%. Wu and Hsu (2021) developed a practical approach to predict at-fault crash driver frequency by the fusion of deep learning approaches. The authors combined a convolutional neural network and gated recurrent units (GRU) and compared its performance with one econometric model, two machine learning models, and one deep learning model. The fusion model outperformed other models in terms of accuracy.

One limitation of the previous models is that its performance on the rare class of crash dataset (e.g., fatal crashes) is generally not good. To improve the performance of model on rare class, this paper also introduced a customized f1-loss function that will force the deep learning model to balance between the common and the rare classes of the crash database (e.g., fatal, injury and PDO crashes). This function will not optimize precision only; rather, it will use precision and recall value of the predicted output to balance between these two.

Zheng et al. (2019) used a few crash-related features to estimate the feature weights and used the weights to transform the data into images. In the crash dataset, there are many features related to driver, crash, environment, roadway, vehicles, etc. To consider all the features, a more generalized image transformation technique is necessary. To address this gap, this paper employed a generalized image transformation technique that can be used with any data of any size. The deep learning model can then extract the relevant information from the image while classifying. A detailed flow diagram of the image transformation technique used in the current paper with crash data is presented in the methodology (section 4.3).

3. Data

The crash data used in this study was obtained from the Louisiana Department of Transportation and Development (LDOTD). The crash data consisted of 10,048 crashes that occurred between 2014 to 2018. The crash severity variable was categorized into three levels: PDO, Injuries, and Fatalities. Among all the crashes, only 42 crashes were fatal crashes, 2699 were injury crashes, and 7307 were PDO crashes. The dataset contained 98 variables, including information about unique crash id, year, date, and time of the crash, and many other features such as crash-related features, environmental features, roadway features, vehicle features, etc.

4. Methodology

4.1. Data cleaning & processing

The crash dataset contained a lot of variables with null values. Besides, some of the variable value was not realistic. First, all the null values from the pavement width column were removed. It removed a significant number of crashes (2610 crashes) with null values. Then a more precise approach was conducted for data cleaning by going through every variable and check for null and unrealistic values. There were an additional 260 crashes in the data with null values from different variables that were removed. Sixty-three crashes were removed as they occurred at roads with 0 lanes. There was an inconsistency between the number of lanes and pavement width for some crashes. A maximum and minimum pavement width for a different number of lanes were defined, and the inconsistent crashes were removed from the data. The posted speed variable also got some values below 15 and above 75. The 265 crashes with posted speed below 15 were removed from the data. When the number of vehicles was more than 1, the driver and vehicle condition for the second driver and vehicle was null for 94 crashes. Those were also removed from the data. The median width value of greater than 64 removed an additional 242 crashes. After the rigorous cleaning of the crash dataset, the final cleaned data included 6336 crashes and 54 variables. Among the crashes, 33 crashes were from fatal crashes, 1806 were injury crashes, and 4497 were PDO crashes.

Data processing was done by converting nominal and ordinal level data to dummy variables. So for each variable, if the variable can take on n values, that variable was converted to n-1 dummy variables. For interval and ratio level data, the variables were kept as is. The data process resulted in 435 total variables in the crash data. For testing purposes, the data was divided into 80–20 split (80 % for training and 20 % for testing). A max-min normalizer was fitted on the training data, then the train and test data were transformed using the fitted normalizer.

4.2. Variable selection

Features (variables) selection is an important part of minimizing the redundant explanatory variables to significant ones (Heinze et al., 2018). Prior to the classification task, feature selection is an important procedure to achieve an improved statistical model accuracy by reducing noise introduced in the data from insignificant factors. For feature selection, the Classification and Regression Tree (CART) model and Multivariate Adaptive Regression Spline (MARS) was used in previous studies (Breiman et al., 1984; Friedman, 1991; Mohamed Amin et al., 2020; Sauvé and Tuleau-Malot, 2011). Kuhnert et al. used logistic regression, CART and MARS to analyze injury of motor vehicles and suggested CART as a precursor of the logistic regression model (Kuhnert et al., 2000). Prior to classifying crash severity using the support vector machine, this paper used the CART and MARS model to select the variables that have significant associations with our dependent variable (crash severity).

4.3. Numeric to image conversion

The numeric crash dataset was converted to an image to be used as an input to the deep learning model. This paper used a generalized technique mentioned at Sharma et al. to convert numeric crash data into image (Sharma et al., 2019). Previously this image conversion technique was used to classify genomics, vowels, and text data.

The main purpose was to convert the feature vectors to a feature matrix, where the feature vectors were the variables of crash data, and the feature matrix had the positions of the variables. Once the feature matrix is finalized, we can populate the matrix with variable value for each crash. So, for each crash, we will have one feature matrix, and the position of variables in that matrix will be populated with that specific crash variable values.

Fig. 1 illustrates the overall procedures to convert the crash data from numeric to an image. The processed crash data was in a form where the crashes were in the row, and the variables were in the columns (C x V). Then the dataset was transposed such that the crashes were now in the column, and variables were the rows (V x C). Then the dimensionality reduction technique, tSNE was applied to this transposed data to project the rows/variables to a 2-D plane (van der Maaten and Hinton, 2008). Each variable now had a Cartesian coordinate associated with it. Once the variables were projected, the convex hull algorithm was applied to find the minimum rectangle covering all the points/variables.

The rectangle and the variable points were rotated to map the rectangle and the points to a horizontal position. Then the coordinates of the points were converted to pixel locations by normalizing the Cartesian points and multiplying with a predefined image height (H). So for each variable, Cartesian coordinates were converted to pixel locations shown as x and y in Fig. 1. The pixel location of each variable is now fixed and will be used to convert crashes into images. Then, for each crash, first an all-ones matrix of size H x H was created. Then, the pixel location of every variables in all-ones matrix was populated with the variables value. This will produce a unique matrix for the specific crash. Then the procedure is repeated for all the crashes and the final images were generated of size H×H×C, where C is the number of crashes.

When converting the numeric dataset into images, we applied a dimensionality reduction technique (t-SNE) to map the variables to 2-D projections. The t-SNE model tries to group the similar variables together. Therefore, in mapping the dataset, some of the variables overlap with each other when the Cartesian coordinates are converted to pixel locations. For this type of overlapping variables, an averaged value was calculated by averaging the values of the overlapping variables and the pixel location was populated using this averaged value. This image conversion process is an effective element arrangement method, which places the similar elements together and dissimilar elements further.

4.4. Evaluation metrics

For imbalanced data, accuracy is not a proper evaluation metric as the model may predict the common classes more accurately while performing poorly on rare classes (Bekkar et al., 2013; Guo et al., 2008; Tharwat, 2018). For this reason, this paper used precision and recall as evaluation metrics. Precision and recall can be defined as follows (Ting, 2010):

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where TP (true positive) represents the cases where the model predicted the positive classes correctly. FP (false positive) represents the cases where the model predicted the positive classes incorrectly. FN (false negative) represents the cases where the model predicted the negative classes incorrectly. So, if a model does well in predicting common classes but fails in rare classes, the precision and recall value for the rare class will be lower than the precision and recall value for the common class.

4.5. Deep learning model

In this paper, we used the transfer learning method to utilize a model's knowledge in a different domain. Recently, the EfficientNet model outperformed all other models in ImageNet competition (Russakovsky et al., 2015; Tan and Le, 2019). EfficientNet uses a novel scaling method to scale the network dimensions, so it can predict with more accuracy while maintaining efficiency. It also transferred well and achieved top accuracy on CIFAR-100, Flowers, and three other transfer learning dataset (Tan and Le, 2019). A detailed architecture of this

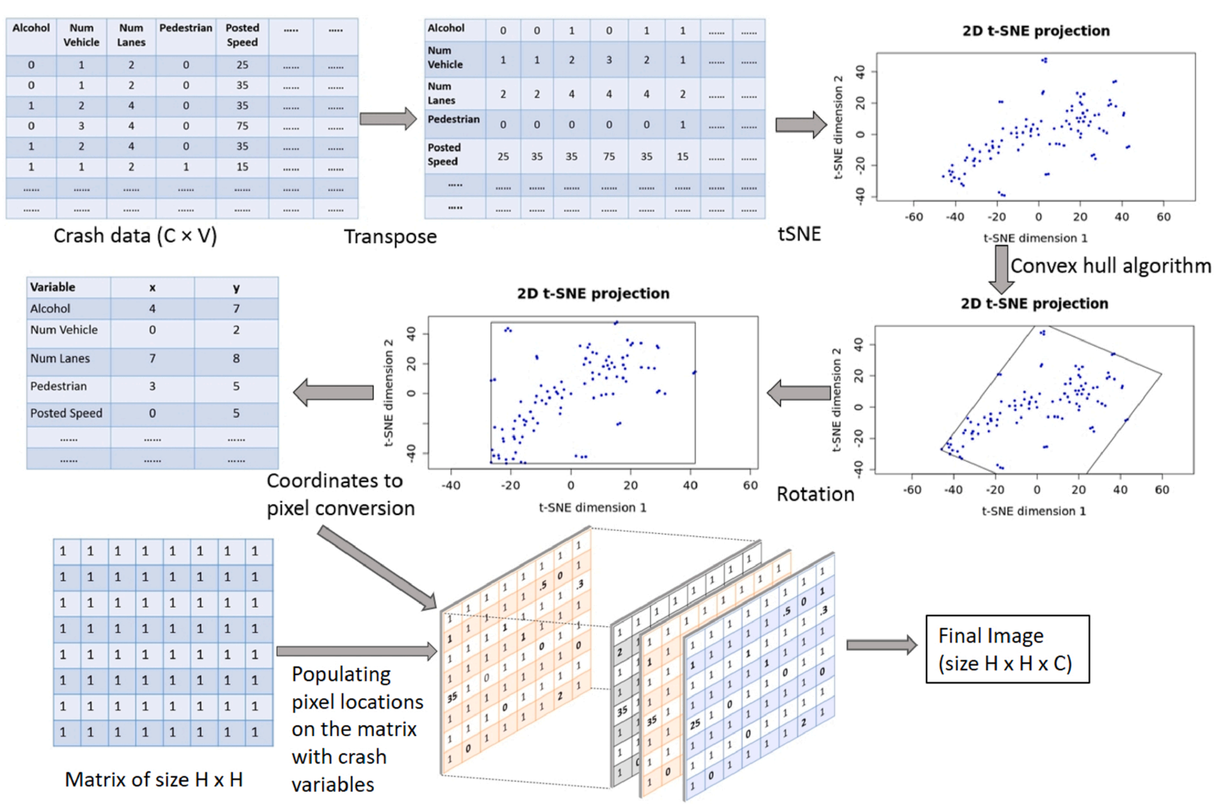

Fig. 1. Flow chart of Numeric to Image conversion.

model can be found at (Agarwal, 2020, May 23). This paper used EfficientNet-B7, pre-trained on ImageNet, as a feature extractor in the classification task. As for the classifier, we changed the last fully connected layer of the EfficientNet model and used a 2-layer neural network with batch normalization, dropout, and ReLU activation function (Quattoni et al., 2008). A threshold needs to be set for each label to get the binary predictions of the model generated probability values (Maiza, 2019). This study used the default value of 0.5. This proposed deep learning model is presented in Fig. 2.

For the deep learning model, a differentiable loss function should be defined (Goodfellow et al., 2016). According to the literature, evaluating the model using accuracy, the cross-entropy loss function is very popular (X. Li et al., 2019; Zhang and Sabuncu, 2018). The cross-entropy loss function tries to minimize the error between the predicted output and the ground truth. For a balanced dataset, this loss function can achieve high accuracy, but its performance deteriorates when the data becomes imbalanced. So, to work with an imbalanced dataset, this paper used a customized f1 loss function, originally written by Michal Haltuf on Kaggle (Haltuf, 2018). This function is differentiable and directly optimize for f1-score. Park reproduced the function in PyTorch, and we modified it to incorporate the f-beta score (Hand and Christen, 2018; Park, 2019).

$$\mathbf{f-beta\ score} = 2\mathbf{x}\frac{(1+\boldsymbol{beta}^2)*\boldsymbol{precision}*\ \boldsymbol{recall}}{\boldsymbol{beta}^2*\boldsymbol{precision}+\boldsymbol{recall}}$$

If the β value is 1, the f-beta score will give equal weights to precision and recall. If the β score is less than 1, it will give more importance to precision, and if the β value is greater than 1, it will give more importance to recall.

4.6. Inference

The f-beta loss function can be used with a binary level target variable only. The target variable of crash data had three-level: no injury (PDO), injury, and fatal. So, we needed two deep learning models to classify the crash data. We prepared and trained a series of two models, one for fatal vs. others, and another for injury vs. PDO. This proposed inference procedure is shown in Fig. 3. To prepare the input, the test crash data was transformed to image data. The first deep learning model takes the image as input and predicts an output. If the prediction is fatal, then the crash is fatal. If the first deep learning model predicts others, then the image will be used as an input to the second deep learning model. The second deep learning model's prediction will be the final severity result of the crash.

5. Analysis and result

5.1. Variable selection result

This paper used both CART and MARS models to select significant variables before the crash severity prediction task. The variable selection procedure was implemented using the caret and earth library in R. The CART model resulted in 10 significant variables that have a significant association with the severity of crashes, while the MARS model resulted in 12 significant variables. Both the models have some common variables like pedestrians involvement, drivers airbag deployed or not, number of vehicles, annual daily traffic, violations type "careless operation", and most harm event of a rollover. The results of variable selection using the CART and MARS model are shown in Tables 1 & 2 , respectively.

As shown in Table 1, the results of CART showed the variables that have a significant association with crash severity are: number of vehicles, driver airbag non deployed, collision manner of sideswipe same direction, collision manner of rear end, most harmful event of rollover, pedestrian's involvement, and violations type of careless operation.

While, as shown in Table 2, the results of MARS showed that the variables that have a significant association with crash severity are: driver airbag non deployed, number of vehicles, collision manner of sideswipe same direction, pedestrian's involvement, alcohol involved, most harmful event of rollover, violations type of careless operation, vehicle type of passenger car, annual daily traffic, and driver's age. Vehicle type of semi-trailer, location type of intersection, and annual daily traffic showed minimal score on CART result. On the other hand, MARS result showed minimal score for normal and inattentive driver's condition.

5.2. SVM models

The selected variables form CART and MARS models were used as inputs for the SVM model. The model was fitted on the training dataset. To assess the performance of the model, the evaluation metrics were evaluated on the test dataset. A grid search algorithm was run to find out the optimal C, gamma, and kernel function of the model. The result showed that a C value of 100, a gamma value of 1, and the RBF kernel function was optimum for the dataset. A SMOTE algorithm was also run to oversample the training dataset (Chawla et al., 2002). The SMOTE was implemented using imblearn library in python. A total of 4-datasets were prepared; data prepared using CART variables, data prepared using MARS variables, smote applied to CART variable data, and smote applied to MARS variable data. These datasets were fitted through SVM classifiers. The performance of the SVM models on the test dataset is

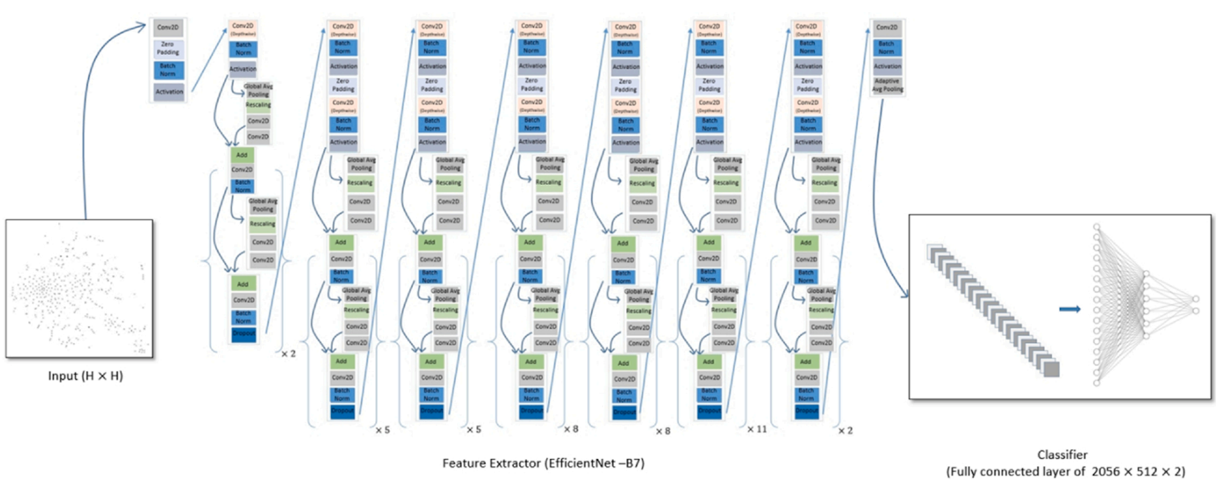

Fig. 2. Proposed Deep Learning Model.

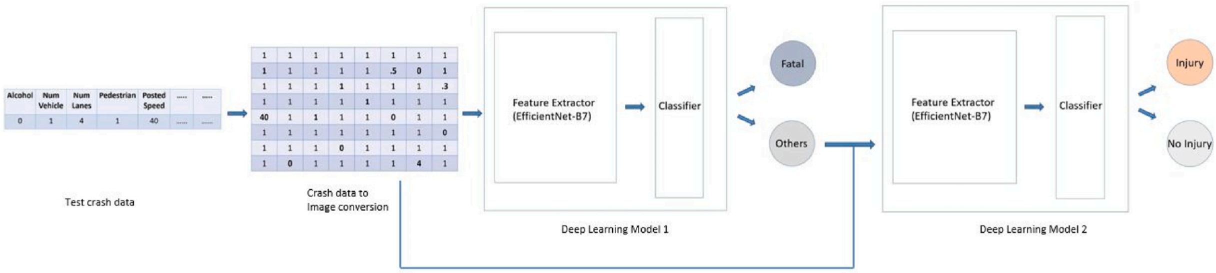

Fig. 3. Flow chart of Inference.

Table 1
CART variable importance result.

Variable	Score
Number of Vehicle	100
Driver airbag non-deployed	75.344
Collision manner (sideswipe same direction)	59.353
Collision manner (rear-end)	20.359
Most Harm Event (rollover)	14.482
Pedestrian involved	9.433
Violations type (careless operation)	6.176
Vehicle type (semi-trailer)	2.044
Location type (intersection)	1.176
Annual daily traffic	0.676

Table 2
MARS variable importance result.

Variable	Score
Driver airbag non-deployed	100
Number of Vehicle	77.541
Collision manner (sideswipe same direction)	62.005
Pedestrian involved	53.768
Alcohol involved	45.386
Most Harm Event (rollover)	37.814
Violations type (careless operation)	23.378
Vehicle type (passenger car)	20.096
Annual daily traffic	17.516
Driver's age	15.304
Driver condition (normal)	7.341
Driver condition (inattentive)	4.986

presented in Table 3.

Form Table 3, SVM models fitted on the MARS variable data, and the smote applied MARS variable data could not classify the fatal crashes. So the precision and recall value is 0. The SVM model fitted on CART variable data did a good job of classifying the fatal crashes with a precision and recall value of 0.09 and 0.17, respectively. The model fitted on the smote applied CART variable data achieved the best result for fatal crashes with a recall value of 0.33. The same model achieved recall value of 0.76 and 0.44 on PDO and injury crashes, respectively.

Table 3
Results of SVM models.

Crash Severity	Data	Precision	Recall	F1-score
PDO	***CART**	0.77	**0.97**	0.86
	CART (smote applied)	0.81	0.76	0.78
	MARS	0.78	0.92	0.84
	MARS (smote applied)	0.8	0.72	0.76
Injury	CART	0.67	0.18	0.28
	CART (smote applied)	0.41	0.44	0.42
	MARS	0.54	0.27	0.36
	MARS (smote applied)	0.4	0.51	0.45
Fatal	CART	0.09	0.17	0.12
	CART (smote applied)	0.06	0.33	0.1
	MARS	0	0	0
	MARS (smote applied)	0	0	0

5.3. Deep learning models

The deep learning model was implemented in the Google Cloud Platform using python and PyTorch (Arif, 2020). To train the model with image data, a TESLA P100 GPU was used. The batch size was selected to be 16 to manage the space on the GPU. A stochastic gradient descent optimizer was used with a learning rate scheduler that reduces the learning rate after every five epochs (Goodfellow et al., 2016). The training dataset was further divided into 80 % training set and 20 % validation set. The validation set was used to track the value of the loss and save the model's weights. The test dataset used in evaluating the SVM classifier was also used to evaluate the deep learning models.

The numeric to image transformation was applied to two different datasets: CART variable data and MARS variable data. The deep learning models were trained using these two image datasets. The image shape was selected to be 120×120 ($H = 120$) (Sharma et al., 2019). To read the image data, the OpenCV library was used. A learning rate of 0.0001 was selected after some trials. Data augmentation was used to prevent overfitting of the model. Different data augmentation technique like rotation, flip, and shear was applied to the training dataset. The data imbalance problem was handled by the customized f1-loss function. A β value of 0.75, 1.0, 1.25, and 1.5 was used in the loss function to show the corresponding changes in recall and precision values.

The results of the deep learning model on the test dataset are presented in Tables 4 & 5 . In Table 4, the result of the first deep learning model is presented. The deep learning model was able to detect fatal crashes using both datasets (CART and MARS). For CART variable data, the deep learning model achieved the best recall value of 0.25 with a β value of 1.5. The model trained on MARS variable data achieved the best recall value of 0.67 on fatal crashes of the test dataset with a β value of 1.5.

In Table 5, the result of the second deep learning model is presented. For CART variable data, the deep learning model achieved the best recall value of 0.84 and 0.64 with a β value of 1.5 for PDO and Injury crashes, respectively. This model achieved better recall value than the model trained on the MARS variable data. With the MARS dataset, the model achieved a recall value of 0.41 and 0.42 with a β value of 1.5 for PDO and Injury crashes, respectively. As the β value is decreased, the recall values are also decreased.

Table 4
Results of the first deep learning (EfficientNet) model for Fatal crashes.

Crash severity	β	Image Data	precision	recall	F1-score
Fatal	1.5	CART	0.01	0.25	0.02
		***MARS**	0.01	**0.67**	0.02
	1.25	CART	0.03	0.19	0.05
		***MARS**	0.045	**0.47**	0.08
	1.0	CART	0.05	0.16	0.075
		***MARS**	0.07	**0.34**	0.12
	0.75	CART	0.08	0.09	0.08
		MARS	0.13	0.19	0.15

Table 5
Results of the second deep learning (EfficientNet) model for PDO and Injury crashes.

Crash severity	β	Image Data	precision	recall	F1-score
PDO	1.5	CART	0.72	0.84	0.76
		MARS	0.71	0.41	0.52
	1.25	CART	0.75	0.79	0.77
		MARS	0.72	0.36	0.48
	1.0	CART	0.77	0.74	0.75
		MARS	0.76	0.27	0.4
	0.75	CART	0.78	0.67	0.72
		MARS	0.79	0.23	0.35
Injury	1.5	***CART**	0.27	**0.64**	0.38
		MARS	0.28	0.42	0.34
	1.25	***CART**	0.41	**0.53**	0.46
		MARS	0.39	0.37	0.38
	1.0	CART	0.47	0.44	0.45
		MARS	0.43	0.31	0.36
	0.75	CART	0.5	0.39	0.44
		MARS	0.46	0.28	0.35

5.4. Discussion

In the loss function, different β scores were used to put different weights on the recall value. This is because, if the recall value for fatal crashes is low, then the model may classify fatal crashes as injury or PDO crashes. So for fatal crashes, the recall value should be higher. If the precision value is lower, that means the model may predict more injury or PDO crashes as fatal crashes. This should not be a major problem, and we can allow for more errors in precision, but if the model predicts a fatal crash as a PDO or injury crash, that could be a life-threatening situation.

The result of the SVM model shown in Table 3 indicated that the performance of the CART data model outperformed the performance of the MARS data model. The smote applied CART model gave better recall value for fatal and injury crashes than the CART model without smote. The recall value of the smote applied CART model for the fatal class was 0.33, which means the model predicted 33 % of the fatal crashes correctly. The recall value for injury and PDO crashes was 0.44 and 0.76, respectively. In addition, the CART dataset showed an improved f1-score for Fatal and PDO crashes, where the smote-applied MARS dataset showed better results on Injury crashes.

The results of deep learning models are presented in Tables 4 & 5 . The fatal crash result was obtained from the first deep learning model. The result for injury and PDO crashes were obtained from the second deep learning model (as explained in Fig. 3). The results showed that the MARS data model outperformed the CART data model for fatal crashes on the recall metric. The recall value for the MARS data model was 0.67, which means that in the test dataset, the model predicted 67 % of the fatal crashes correctly. For the second model with injury and PDO crashes, the CART data model outperformed the MARS data model with a recall value of 0.64 for injury crashes and 0.84 for PDO crashes.

Comparing the results of Tables 3,4, and 5, the deep learning model outperformed the SVM model in terms of recall on both fatal and injury crashes. In fatal crash prediction, the first deep learning model, trained on the MARS data, showed improved recall values of 0.67, 0.47, and 0.34 with beta values of 1.5, 1.25, and 1.0 than the SVM model trained on the smote applied CART data (0.33). In injury crash prediction, the second deep learning model, trained on the CART data, showed a recall value of 0.64 with a beta value of 1.5, where the SVM model, trained on the smote applied MARS data, showed a recall value of 0.51. Also, for fatal and injury crashes, the deep learning model showed an improved f1-score than the SVM model.

The previous deep learning-based prediction model by Zheng et al. (2019) was optimized using a cross-entropy loss function and achieved an average precision of 0.063 and an average recall of 0.063 on fatal crashes. As the dataset used in their model is different from ours, the direct comparison in fatal crashes between these two deep learning models is not suitable. However, we achieved a significantly higher recall value (0.67), which is more feasible for real-life applications.

6. Conclusions and recommendations

Traffic crash severity prediction is an important component in traffic collisions management. Any improvement in crash severity prediction should be adopted since it involves human injury and the chance of survival. Also, an efficient and accurate crash severity prediction model is highly-demandable for real-world applications. Previous studies used different statistical and machine learning models for this purpose. The main drawback of these models was the low performance on fatal crashes. Few studies used deep learning models to improve the model's performance, but these models were optimized for accuracy only. As the traffic crash dataset is highly imbalanced, they failed to improve the performance of rare classes (e.g., fatal crashes). In this paper, the proposed deep learning approach was able to predict crash severity with improved performance.

The proposed method utilized a generalized numeric to image transformation technique to transform the crash dataset to images and train deep learning models using a customized f1-loss function. The customized f1-loss function optimized the deep learning model on the f-beta score. As correctly predicting the fatal crashes is most important for real-world application purposes, we put more emphasis on the recall value.

The performance of the deep learning model was evaluated using precision and recall, and compared with standard statistical learning models (SVM). The CART and MARS models were employed to select the significant variables from the crash dataset. To account for the data imbalance problem for the SVM model, the SMOTE algorithm was applied to oversample the dataset. In the deep learning model, the imbalance dataset was handled by the loss function.

The results showed that the proposed deep learning model outperformed the SVM model in classifying the fatal and injury crashes. Dataset prepared with MARS variables achieved a better recall value for fatal crashes, and dataset prepared from CART variables achieved a better recall value for injury and PDO crashes. The improvement of the fatal crash recall value was also very significant from the previous studies.

An inference layout was proposed to apply the deep learning framework for practical application. To get a quick assessment right after a crash, collected data can be instantly evaluated for severity through the deep learning inference procedure. It will facilitate to predict the severity of a crash and take necessary precautions on time to minimize the emergency response time and hopefully decrease the likelihood of fatalities.

The results of our deep-learning models can be used as a benchmark for future studies. We recommend to apply the proposed deep learning model and evaluate the performance using other crash datasets. The framework presented in this paper is generalizable, specifically the numeric to image transformation and the customized f1-loss function. Future studies may further tune the weight parameter (β) of the loss function and the threshold value for classifiers to get more optimized precision and recall values suitable for real-life applications.

Authors Statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *Accident Analysis & Prevention.*

Authorship contributions

Please indicate the specific contributions made by each author (list the authors' initials followed by their surnames, e.g., Y.L. Cheung). The name of each author must appear at least once in each of the three categories below.

Conception and design of study: Md Adilur Rahim, Hany Hassan, acquisition of data: Md Adilur Rahim, analysis and/or interpretation of data: Md Adilur Rahim, Hany Hassan

Drafting the manuscript: Md Adilur Rahim, revising the manuscript critically for important intellectual content: Md Adilur Rahim, Hany Hassan

Approval of the version of the manuscript to be published (the names of all authors must be listed): Md Adilur Rahim, Hany Hassan

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The authors greatly appreciate the financial support of the Transportation Consortium of South-Central States, Tran-SET, USA (Tran-SET Project No. 20ITSLSU17, Award No. AM200336).

References

Abdelwahab, H.T., Abdel-Aty, M., 2001. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transp. Res. Rec. 1746 (1), 6–13. https://doi.org/10.3141/1746-02.

Agarwal, V., 2020. Complete Architectural Details of All EfficientNet Models. May 23. Towards Data Science. https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142.

Alkheder, S., Taamneh, M., Taamneh, S., 2017. Severity prediction of traffic accident using an artificial neural network. J. Forecast. 36 (1), 100–108. https://doi.org/10.1002/for.2425.

AlMamlook, R.E., Kwayu, K.M., Alkasisbeh, M.R., Frefer, A.A., 2019. Comparison of machine learning algorithms for predicting traffic accident severity, 9-11 April 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) 272–276. https://doi.org/10.1109/JEEIT.2019.8717393.

Al-Saffar, A.A.M., Tao, H., Talab, M.A., 2017. Review of deep convolution neural network in image classification. 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET) 26–31. https://doi.org/10.1109/ICRAMET.2017.8253139.

Al-Zuhairi, M., Pradhan, B., 2017. Severity prediction of traffic accidents with recurrent neural networks. Appl. Sci. 7 https://doi.org/10.3390/app7060476.

Arif, T.M., 2020. Introduction to Deep Learning for Engineers: Using Python and Google Cloud Platform. Morgan & Claypool Publishers.

Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accid. Anal. Prev. 122, 239–254. https://doi.org/10.1016/j.aap.2018.10.015.

Bekkar, M., Djemaa, H.K., Alitouche, T.A., 2013. Evaluation measures for models assessment over imbalanced data sets. J. Inform. Eng. Appl. 3 (10), 27–38.

Benjamin, Q.H., Hui, L., Maryellen, L.G., 2016. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J. Med. Imaging 3 (3), 1–5. https://doi.org/10.1117/1.JMI.3.3.034501.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Taylor & Francis.

Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accid. Anal. Prev. 38 (5), 1019–1027. https://doi.org/10.1016/j.aap.2006.04.009.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z., 2016. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accid. Anal. Prev. 90, 128–139. https://doi.org/10.1016/j.aap.2016.02.011.

Elamrani Abou Elassad, Z., Mousannif, H., Al Moatassime, H., 2020a. Class-imbalanced crash prediction based on real-time traffic and weather data: a driving simulator study. Traffic Inj. Prev. 21 (3), 201–208. https://doi.org/10.1080/15389588.2020.1723794.

Elamrani Abou Elassad, Z., Mousannif, H., Al Moatassime, H., 2020b. A real-time crash prediction fusion framework: an imbalance-aware strategy for collision avoidance systems. Transp. Res. Part C Emerg. Technol. 118, 102708 https://doi.org/10.1016/j.trc.2020.102708.

Federal Highway Administration, 2019. FHWA Work Zone Facts and Statistics. March 25. Federal Highway Administration. https://ops.fhwa.dot.gov/wz/resources/facts_stats.htm.

Fiorentini, N., Losa, M., 2020. Handling imbalanced data in road crash severity prediction by machine learning algorithms. Infrastructures 5 (7), 61. https://doi.org/10.3390/infrastructures5070061.

Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., Yuan, J., 2020. Predicting real-time traffic conflicts using deep learning. Accid. Anal. Prev. 136, 105429 https://doi.org/10.1016/j.aap.2019.105429.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19 (1), 1–67. https://doi.org/10.1214/aos/1176347963.

Gibbons, C., Richards, S., Valderas, J.M., Campbell, J., 2017. Supervised machine learning algorithms can classify open-text feedback of doctor performance with human-level accuracy. J. Med. Internet Res. 19 (3), e65. https://doi.org/10.2196/jmir.6533.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. The MIT Press.

Guo, X., Yin, Y., Dong, C., Yang, G., Zhou, G., 2008. On the class imbalance problem. Fourth International Conference on Natural Computation 192–201. https://doi.org/10.1109/ICNC.2008.871.

Haltuf, M., 2018. Best Loss Function for F1-score Metric. Kaggle. https://www.kaggle.com/rejpalcz/best-loss-function-for-f1-score-metric.

Hand, D., Christen, P., 2018. A note on using the F-measure for evaluating record linkage algorithms. Stat. Comput. 28 (3), 539–547. https://doi.org/10.1007/s11222-017-9746-6.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. IEEE International Conference on Computer Vision (ICCV 2015) 1502. https://doi.org/10.1109/ICCV.2015.123.

Heinze, G., Wallisch, C., Dunkler, D., 2018. Variable selection - A review and recommendations for the practicing statistician. Biom. J. 60 (3), 431–449. https://doi.org/10.1002/bimj.201700067.

Huang, T., Wang, S., Sharma, A., 2020. Highway crash detection and risk estimation using deep learning. Accid. Anal. Prev. 135, 105392 https://doi.org/10.1016/j.aap.2019.105392.

Jamro, E., Wiatr, K., 2001. Convolution operation implemented in FPGA structures for real-time image processing, 19-21 June ISPA 2001. Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis. In Conjunction With 23rd International Conference on Information Technology Interfaces (IEEE Cat.) 417–422. https://doi.org/10.1109/ISPA.2001.938666.

Jeong, H., Jang, Y., Bowman, P.J., Masoud, N., 2018. Classification of motor vehicle crash injury severity: a hybrid approach for imbalanced data. Accid. Anal. Prev. 120, 250–261. https://doi.org/10.1016/j.aap.2018.08.025.

Jiang, F., Yuen, K.K.R., Lee, E.W.M., 2020. A long short-term memory-based framework for crash detection on freeways with traffic data of different temporal resolutions. Accid. Anal. Prev. 141, 105520 https://doi.org/10.1016/j.aap.2020.105520.

Kuhnert, P.M., Do, K.A., McClure, R., 2000. Combining non-parametric models with logistic regression: an application to motor vehicle injury data. Comput. Stat. Data Anal. 34 (3), 371–386. https://doi.org/10.1016/S0167-9473(99)00099-7.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Li, Z., Liu, P., Wang, W., Xu, C., 2012. Using support vector machine models for crash injury severity analysis. Accid. Anal. Prev. 45, 478–486. https://doi.org/10.1016/j.aap.2011.08.016.

Li, X., Yu, L., Chang, D., Ma, Z., Cao, J., 2019. Dual cross-entropy loss for small-sample fine-grained vehicle classification. Ieee Trans. Veh. Technol. 68 (5), 4204–4212. https://doi.org/10.1109/TVT.2019.2895651.

Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. Accid. Anal. Prev. 135, 105371 https://doi.org/10.1016/j.aap.2019.105371.

Maiza, A., 2019. The Unknown Benefits of Using a Soft-F1 Loss in Classification Systems. Towards Data Science. https://towardsdatascience.com/the-unknown-benefits-of-using-a-soft-f1-loss-in-classification-systems-753902c0105d.

Maqsood, M., Nazir, F., Khan, U., Aadil, F., Jamal, H., Mehmood, I., Song, O.-y., 2019. Transfer learning assisted classification and detection of alzheimer's disease stages using 3D MRI scans. Sensors 19 (11), 2645. https://doi.org/10.3390/s19112645.

Minar, M.R., Naher, J., 2018. Recent advances in deep learning: an overview. ArXiv abs/1807.08169.

Mohamed Amin, M., Zainal, A., Firdaus Mohd. Azmi, N., Azizah Ali, N., 2020. Feature selection using multivariate adaptive regression splines in telecommunication fraud detection. IOP Conference Series: Materials Science and Engineering 864, 012059. https://doi.org/10.1088/1757-899x/864/1/012059.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Park, S., 2019. F1 Score in PyTorch. Github Gist. https://gist.github.com/SuperShinyEyes/dcc68a08ff8b615442e3bc6a9b55a354.

Peng, Y., Li, C., Wang, K., Gao, Z., Yu, R., 2020. Examining imbalanced classification algorithms in predicting real-time traffic crash risk. Accid. Anal. Prev. 144, 105610 https://doi.org/10.1016/j.aap.2020.105610.

Quattoni, A., Collins, M., Darrell, T., 2008. Transfer learning for image classification with sparse prototype representations. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2. https://doi.org/10.1109/CVPR.2008.4587637.

Rocca, B., 2019. Handling Imbalanced Datasets in Machine Learning. Jan 27. Towards Data Science. https://towardsdatascience.com/handling-imbalanced-datasets-in-machine-learning-7a0e84220f28.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115 (3), 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sauvé, M., Tuleau-Malot, C., 2011. Variable selection through CART. arXiv, 1101.0689 [math, stat]. http://arxiv.org/abs/1101.0689.

Sharma, A., Vans, E., Shigemizu, D., Boroevich, K.A., Tsunoda, T., 2019. DeepInsight: a methodology to transform a non-image data to an image for convolution neural network architecture. Sci. Rep. 9 (1), 11399 https://doi.org/10.1038/s41598-019-47765-6.

Tan, M., Le, Q.V., 2019. EfficientNet: rethinking model scaling for convolutional neural networks. ArXiv abs/1905.11946.

Texas Department of Transportation, 2020. Work Zones. Texas Department of Transportation. https://www.txdot.gov/inside-txdot/media-center/psas/distracted-driving/work-zones.html.

Tharwat, A., 2018. Classification assessment methods. Appl. Comput. Inform. https://doi.org/10.1016/j.aci.2018.08.003.

Ting, K.M., 2010. Precision and recall. In: Sammut, C., Webb, G.I. (Eds.), Encyclopedia of Machine Learning. Springer US., Boston, MA, p. 781.

Turner, J.D., 1999. What's a work zone? May/June Public Roads 62 (6). https://www.fhwa.dot.gov/publications/publicroads/99mayjun/workzone.cfm.

van der Maaten, L., Hinton, G., 2008. Viualizing data using t-SNE. J. Mach. Learn. Res. 9 (86), 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html.

Wu, Y.W., Hsu, T.P., 2021. Mid-term prediction of at-fault crash driver frequency using fusion deep learning with city-level traffic violation data. Accid. Anal. Prev. 150, 105910 https://doi.org/10.1016/j.aap.2020.105910.

Yahaya, M., Jiang, X., Fu, C., Bashir, K., Fan, W., 2019. Enhancing crash injury severity prediction on imbalanced crash data by sampling technique with variable selection, 27-30 Oct 2019 IEEE Intelligent Transportation Systems Conference (ITSC) 363–368. https://doi.org/10.1109/ITSC.2019.8917223.

Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. Accid. Anal. Prev. 51, 252–259. https://doi.org/10.1016/j.aap.2012.11.027.

Yu, R., Wang, Y., Zou, Z., Wang, L., 2020. Convolutional neural networks with refined loss functions for the real-time crash risk analysis. Transp. Res. Part C Emerg. Technol. 119, 102740 https://doi.org/10.1016/j.trc.2020.102740.

Zhang, Z., Sabuncu, M., 2018. Generalized cross entropy loss for training deep neural networks with noisy labels. arXiv, 1805.07836. http://papers.nips.cc/paper/8094-generalized-cross-entropy-loss-for-training-deep-neural-networks-with-noisy-labels.pdf.

Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., Wang, Z., 2019. Traffic accident's severity prediction: a deep-learning approach-based CNN network. IEEE Access 7, 39897–39910. https://doi.org/10.1109/ACCESS.2019.2903319.