Article

An Injury-Severity-Prediction-Driven Accident Prevention System

Gulsum Alicioglu [1] , Bo Sun [2],* and Shen Shyang Ho [2]

1   Department of Electrical and Computer Engineering, Rowan University, Glassboro, NJ 08028, USA;
    alicio87@rowan.edu
2   Department of Computer Science, Rowan University, Glassboro, NJ 08028, USA; hos@rowan.edu
*   Correspondence: sunb@rowan.edu

Abstract: Traffic accidents are inevitable events that occur unexpectedly and unintentionally. Therefore, analyzing traffic data is essential to prevent fatal accidents. Traffic data analysis provided insights into significant factors and driver behavioral patterns causing accidents. Combining these patterns and the prediction model into an accident prevention system can assist in reducing and preventing traffic accidents. This study applied various machine learning models, including neural network, ordinal regression, decision tree, support vector machines, and logistic regression to have a robust prediction model in injury severity. The trained model provides timely and accurate predictions on accident occurrence and injury severity using real-world traffic accident datasets. We proposed an informative negative data generator using feature weights derived from multinomial logit regression to balance the non-fatal accident data. Our aim is to resolve the bias that happens in the favor of the majority class as well as performance improvement. We evaluated the overall and class-level performance of the machine learning models based on accuracy and mean squared error scores. Three hidden layered neural networks outperformed the other models with 0.254 ± 0.038 and 0.173 ± 0.016 MSE scores for two different datasets. A neural network, which provides more accurate and reliable results, should be integrated into the accident prevention system.

Keywords: ordinal regression; neural network; transportation safety; injury severity prediction; sustainable transportation

1. Introduction

Complex traffic environments with unpredictable events threaten the safety of pedestrians, passengers, and drivers. With the increase in population and vehicles, traffic accidents have become a major concern for transportation safety. Traffic accidents increase visible and hidden costs including physical and psychological health issues, and insurance, and impact the economy [1]. A prediction system of potential accidents and injuries helps to improve transportation safety, and reduces costs. The automotive industry focuses on developing and improving sensor-based data-driven intelligent technologies in vehicles to maintain a safe environment in traffic. These intelligent vehicle technologies include functionalities such as determining following distance and perceiving on-road objects [2]. Developing advanced transportation safety systems with the timely prediction of potential traffic accidents and possible injury severity in an intelligent vehicle would ensure road, driver, and passenger safety [2,3]. Designing safe road roundabouts contribute to addressing traffic congestion and increasing pedestrian and road safety [4]. Roundabout intersections cause a small number of collision points due to their geometry [5,6], which reduces the severity of injury levels in accidents. In addition, these intersections ensure the flow of traffic by reducing the time loss at inlets [5] and preventing accidents caused by congestion. However, designing safe roundabout intersections alone is not enough to create sustainable transportation. Therefore, we propose an accident prevention and alerting system supported by a robust prediction model selected after exhaustive experiments. The system predicts traffic accident occurrence and injury severity based on driving status and

environmental conditions to establish advanced transportation safety. However, injury severity prediction is a challenging problem due to imbalanced data, mainly consisting of accident records. Traditional machine learning (ML) algorithms focus on maximizing the overall accuracy of the whole dataset and tend to show poor performance on imbalanced data due to a lack of information on negative or positive samples [7]. Since ML models have a bias in favor of the majority class, achieving a good prediction model in imbalanced learning is crucial in advanced transportation safety systems. Traditional sampling techniques require to have minority classes to balance data. Since traffic data do not include negative, i.e., non-accident data, traditional sampling techniques are impractical for this domain. To overcome these challenges, we propose a negative data generation scheme based on feature weights derived from multinomial logistic regression using positive instances, i.e., accident data.

Another challenge in injury severity problems is the ordinality of classes. Many studies have been conducted to determine accident risks using ML algorithms [8–10]. However, these studies assume that injury severity levels are nominal and none of them used ordinal regression (OR) algorithms, which demonstrate better results than conventional ML algorithms in a classification problem where the class order is essential [11]. Since the injury severity level of an accident is usually ordinal, i.e., from non-fatal injury to fatal injury level, we use ordinal regression algorithms to have a robust accident prevention system with a reliable prediction model for intelligent vehicles to develop an advanced transportation safety system. A prevention and alerting system will detect accident-prone situations and dangerous human driving behaviors to decrease the likelihood of traffic accidents and potential injury severity. In particular, the warnings by this system allow the intelligent vehicles and drivers to take timely precautions by applying safety maneuvers such as decreasing vehicle speed, automatic braking, automatic lane keeping/control, precise maneuvering, etc. in complex traffic environments [2,3,12].

Deep learning (DL) models have achieved impressive performance in various domains such as autonomous vehicle systems with advances in computing power and technologies [13,14]. Since neural networks (NN) have become a powerful technique in finding complex patterns in high dimensional datasets and providing high prediction accuracy, they can provide robust and reliable predictions in ordinal datasets, as well [15,16]. In this study, we compared the performance of NNs, OR models, decision tree (DT), support vector machines (SVM), and logistic regression (LR) to have a robust prediction model in the accident prevention system. We also examined the effect of different hyperparameters and architectures on injury prediction performance.

The contributions of this paper are as follows:

- A new framework to generate non-accident data based on the accident instances using the most contributing factors of traffic accidents. This will ensure a more balanced dataset and improve the predictive model in accident prevention systems for intelligent vehicles,
- A robust and more accurate NN prediction model to estimate injury severity compared to ordinal regression and other methods. With NN, we overcome the disadvantages of ordinal regression models (i.e., low robustness, not dealing with multicollinearity).

The rest of the paper is organized as follows: Section 2 covers the literature review of accident risks and injury severity prediction models with commonly used methods. Section 3 presents the methodology including an overview of the proposed accident prevention system, used methods, and data generation process. Section 4 describes the experimental detail and discusses the experimental results and comparison. Section 5 concludes the paper with future work.

2. Literature Review

Accident data commonly include weather conditions, road conditions, temporal factors, and driving behaviors. ML models have been extensively applied [8–10,12,17–22] in assessing injury severity and determination of critical factors for motor vehicle accidents

of their ability to solve non-linear relationships as seen in traffic data. Yuan et al. [9] applied various machine learning algorithms, including support vector machine, decision tree, neural network, and random forest (RF), to classify accident and non-accident classes. Additionally, they used informative negative sampling to balance the binary classification problem. Zhu et al. [20] proposed a machine learning-based framework to detect driver injury patterns using NN and RF. Pradhan et al. [21] modeled traffic accident severity using NN and SVM methods based on actually reported causes with seven explanatory features. Results indicated that linear SVM has the highest accuracy value. In Delen's research [22], the contributing factors of injury severity are examined using NN, SVM, DT, and LR by modeling the problem as a binary classification. Liao et al. [23] studied injury severity prediction in autonomous vehicles for emergency decision making. They used SVM with three types of kernels and compared their results with ordered-logit and NN algorithms. Table 1 lists all the classes used for accident classification and proposed ML algorithms. The most common ML algorithms (DT, SVM, k-Nearest Neighbor, NN, RF, etc.) are frequently applied in assessing injury severity in accidents [9,10,17–22]. The studies with new and/or modified approaches assessing injury severity in accidents mainly compared with NN because of its learning power [24]. Among these studies, only Yuan et al. [9] focus on resolving the imbalanced data problem by using informative data sampling to create non-accident data. Other studies use only accident data to predict injury severity levels using various ML models. However, none of these studies investigate the effect of different NN architectures on the prediction performance for accident and injury severity.

Table 1. Summary of research related to accident risk assessment.

Studies	Class Descriptions	Algorithms	
[8]	Slight Injured Killed or Seriously Injured	Bayesian Networks	
[9]	Accident No Accident	SVM DT	RF NN
[10]	Fatal Injury Incapacitating Injury Non-Incapacitating Injury Possible Injury No Injury	Logistic Regression (LR) Gradient Boosting Model	NN DT Naïve Bayes
[12]	Non-Fatal Injury Fatal Injury	k-NN Naïve Bayes NN	DT SVM LR
[17]	No Injury Possible Injury Non-Incapacitating Injury Incapacitating Injury Fatal Injury	DT SVM Hybrid DT-Artificial NN	
[18]	Property Damage Only Possible Injury Visible Injury Fatal Injury	Multinomial Logit k-NN SVM	RF k-Means
[20]	No Injury Possible Injury Evident Injury Fatal Injury	RF NN	

Ordinal regression models are used when the classes represent levels of an inherent order [3,11,25–28]. Some examples of the applications include evaluating disease severity in plants [26], healthcare applications [11], and assessing credit-rating agencies [25]. This

study provides a comprehensive comparison with commonly used ML algorithms in accident risk assessment and ordinal regression models [3].

We conducted exhaustive experiments to determine the best NN architecture and hyperparameter configuration for injury severity prediction. We also compared our classification results with four different ordinal regression models [3], and three commonly used methods, namely DT, Linear SVM, and logistic regression, for two different real-world fatal accident datasets. Since most accident datasets only have positive instances, we also proposed a new negative data generation process to overcome the challenges in imbalanced learning and traditional sampling techniques.

3. Methodology

3.1. Overview of Accident Prevention and Alert System

This section provides an overview of the prevention and alerting system framework and details of the prediction model [3]. The details of the prevention system and prediction model framework are shown in Figure 1. The prevention system takes inputs including driver information, GPS data, weather and road situation, and historical accident records. The inputs are used to create predictions and accident risks, then it receives the risks and provides alert messages to warn drivers to take precautions such as reducing speed, keeping a safe following distance, etc. Integrating such a reliable prediction model into intelligent vehicles helps decrease the likelihood of accidents and injury severities, and improve the safety of vulnerable road users, drivers, and passengers. Towards this end, the prediction model is critical to the overall system. The study mainly focuses on developing a robust prediction model to determine and integrate the best classification method into the system.
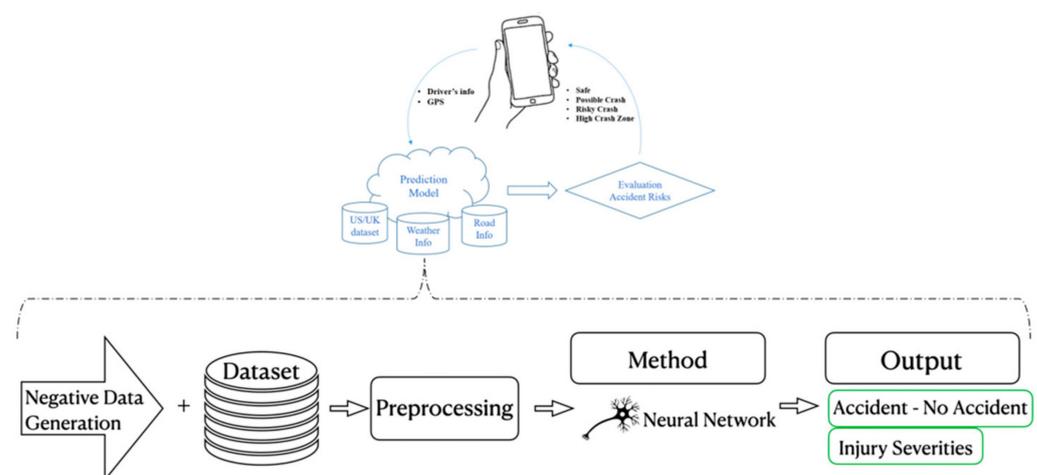

Figure 1. The detailed framework of the prevention system and prediction model.

3.2. Ordinal Regression Models

Ordinal regression models developed by McCullagh, use the ordinal nature of data by defining various stochastic sorting paradigms [29]. These methods resolve the requirement of assigning scores to classes instead of ordinality [29]. Ordinal regression is a supervised learning problem where the label of the classes has an inherent order [27]. Ordinal regression algorithms benefit from this order information to improve classification performance [3,24]. Ordinal regression implementations occur in areas where human-sourced data are important, and output variables cannot be measured with high sensitivity [24,25]. The accident dataset used in the paper presents such characteristics. In this paper, ordinal regression methods are divided into two main groups: Threshold-based and Regression-based methods [3]. Threshold-based methods have two different approaches based on the application of threshold: logistic all-threshold (AT) and logistic immediate-threshold (IT) [3]. Regularization parameter is taken as 50 for threshold-based methods. Regression-

based method includes ordinal ridge and least absolute deviation (LAD). In the LAD method, ε parameter is taken as 0.001, the tolerance value is taken as 0.0001 and the regularization parameter is taken as 10 in this study. For the ordinal ridge method, the regularization parameter and the tolerance values are equal to 10 and 0.0001, respectively. More information can be found in Alicioglu et al. [3]. Equation (1) shows general ordinal regression model [29], where $\gamma j = p1(x) + \ldots + pj(x)$, β is a vector of regression coefficients and $\theta j = logkj$, $kj(x)$ is the odds.

$$\text{OR} = \log \left[\frac{\gamma_j(x)}{1 - \gamma_j(x)} \right] = \theta_j - \beta^t x, \qquad (1 \leq j < k) \tag{1}$$

3.3. Neural Network

Artificial neural networks are supervised machine learning models inspired by the learning mechanism of the human brain [30]. A NN contains more than one computational layer: input layer, hidden layers, and output layer [31]. These layers transmit the information to the consecutive layers. Neurons in these layers are associated with adaptable weights and bias. Input layers forward data with randomly initialized weights to the hidden layers which perform nonlinear transformation with activation functions [32]. The hidden layer uses the output of previous layers as input and transmits its output to the next layers [15]. The last hidden layer passes the information to the output layer to create network outputs [31]. The selection of the hyperparameters of neural networks such as the number of layers, neurons, activation functions, and training algorithm depends on the structure and complexity of the tasks. These hyperparameters affect the learning performance of neural networks [33]. Considering a single-hidden layer network, the output function is obtained as follows (2):

$$y = \sigma \, (w^{\text{T}} x + b) \tag{2}$$

σ indicates activation function, x is a n-dimensional input vector, w is the weight vector, b is the bias, and y is the output of the network. Rectified linear unit function (ReLU), logistic sigmoid function, and hyperbolic tangent function (Tanh) are commonly used activation functions. Forwarding all the information from the input layer to the output layer through activation functions is called a forward pass. Then the optimization algorithms measure the error by comparing the actual prediction with the ground truth value and tracebacks to each layer to update weights and bias. This process is called backpropagation where the algorithm aims to minimize loss by updating computational units. In our experiment, we adopted Stochastic Gradient Descent (SGD) to train a neural network by selecting random examples to estimate gradients instead of calculating the gradients of each example [34]. In addition, Adam algorithm [35] was also used to train the neural network.

3.4. Negative Data Generator

Random sampling, oversampling, and under-sampling techniques are commonly used to mitigate the imbalanced class problem. However, these techniques require both majority and minority classes in the datasets. Since fatal accident datasets lack a negative class, applying traditional random sampling techniques in these fatal accident datasets is impracticable. Therefore, we proposed a naïve data generator. We created negative instances (non-accident data) for datasets to be used in training based on the information on positive instances. The weights used in the negative data generation are obtained by Multinomial Logistic Regression (MLR) [36].

The weight of a feature reflects the importance degree of the feature. We generated negative samples by creating values that were mostly outside of the value ranges of the important features. For instance, as an important feature, the intersection mostly includes ranges from 1 to 3 (i.e., no intersection and four-way intersection) for all classes. Therefore, the negative data for this feature should cover mostly outside of these ranges. For the less

important features based on weight values, existing ranges are used to create negative samples. Thus, negative samples, which most but not all of them are out of the range of positive instances determined by the feature distribution of the positive instances [3,9], are created. Figure 2 describes the steps to create non-accident data using the most important features. First, we obtain the feature weights of the positive class (accident data) by using MLR [36]. Then these weights are ranked in a descending order to determine the top 10 important features. We determine the distribution of these 10 important features. Then, we create negative samples mostly outside of the current feature distribution of these features. For the less important features based on the output of MLR, we randomly create feature distribution for negative class using positive data value ranges. Then we assign a new label to the negative class and combine it with the current dataset. The description of the accident data and feature distribution of the most important features are presented in Appendices A and B.

Figure 2. Negative (Non-accident) data generation process.

4. Experimental Results

4.1. Data Description

Experiments are performed using two different real-world accident datasets. Motor vehicle accident data used for accident risk analysis are retrieved from the US National Highway Traffic Safety Administration website, particularly the Fatality Analysis Reporting System [37] and UK Transport for Greater Manchester website [38]. The US dataset contains accident records from 2015 to 2016 for the states of California, Florida, Georgia, North Carolina, and Texas, where the highest number of accident records were found in the US. The UK dataset contains accident records for 2018. Both datasets went through preprocessing procedures by removing instances that have missing, incorrect, or undefined values in the explanatory variables. To avoid bias in the training process, post-crash-related features such as the number of fatalities are removed from datasets. We also applied a standardization process for both datasets to rescale the features due to the differences among their value ranges. Negative/non-accident samples are created by the proposed data generator for the US dataset. The newly generated class has 8104 instances labeled as "5". The US data have 30,484 entries, six classes, and 17 features related to driving conditions [3]. The UK data have 14,593 entries and 10 features related to driving conditions [3]. Table 2 summarizes the information about injury severity levels of accidents. The classes range from no apparent/slight injury level to fatal injury level. All experiments are conducted using Python libraries.

Table 2. Description of injury severity levels in accident datasets.

US Accident Dataset (2015–2016)			UK Accident Dataset (2018)	
Injury Severity		**# of Accidents**	**Injury Severity**	**# of Accidents**
Class 0	No apparent	6405 (21.0%)	Slight	8381 (57.4%)
Class 1	Possible	2697 (8.84%)	Serious	4541 (31.1%)
Class 2	Minor	2967 (9.73%)	Fatal	1671 (11.5%)
Class 3	Serious	1812 (5.95%)		
Class 4	Fatal	8499 (27.8%)		
Class 5	No accident	8104 (26.5%)		

4.2. Feature Extraction and Negative Data Generation

For the US dataset, among all 17 driving-related features, we only picked high-impact features for negative sampling to create non-accident data. The weights of the features are obtained using multinomial logistic regression. The top five features and their corresponding weights are provided in the order in Table 3. The most crucial feature from the minor injury severity level to fatal injury is alcohol. Surface type, surface condition, person type, age, and sex are also common among these levels. For accidents with low injury severities, such as non-fatal and possible injury, light condition, intersection type, and the number of traffic lanes are among the important features.

Table 3. The top five features and corresponding weights of the US dataset.

Non-Fatal Injury		Possible Injury		Minor Injury		Major Injury		Fatal Injury	
Light condition	0.166	Person type	0.264	Alcohol	0.262	Alcohol	0.490	Alcohol	0.918
Lane	0.161	Intersection type	0.213	Person type	0.259	Person type	0.442	Surface type	0.099
Intersection type	0.064	Sex	0.189	Surface condition	0.122	Surface type	0.127	Age	0.013
Holiday	0.016	Lane	0.081	Surface type	0.099	Sex	0.106	Vehicle make	0.005
Accident hour	0.012	Surface condition	0.032	Accident hour	0.004	Surface condition	0.022	Surface condition	0.002

Table 3 identifies the critical factors for the five accident classes for the US dataset. For the non-accident data generation process, the top ten features' combined range of values is examined. These features are alcohol, person type, intersection type, sex, light condition, lane, surface type, surface condition, holiday, and age, respectively, according to their importance values. For instance, the value of the surface condition feature ranges from one to two for all classes. Thus, other surface condition values should range randomly from three to five for the non-accident class. With this information, by using negative sampling surface condition values ranged from three to five in the non-accident class. Similar approaches are applied to other ten important features to generate random values for the non-accident class. For other less significant features, the values are randomly chosen from the combined range value of the five classes.

We created a side-by-side histogram of feature distributions for positive and negative data for the most important ten features. Figure 3a–d show an example of distributions for age, alcohol, intersection, and light condition features. Figure 3a depicts the distribution of intersection variables for both classes. As indicated before, we used mostly outside of the positive instance range to create negative data. While the intersection variable consists of some of the positive data variables (i.e., 1–3), most of the variables are outside of the range. For example, the number six indicates a round-about intersection and due to their geometry, round-about intersections cause less collision [4–6]. Our negative (non-accident) data generation supports the values/categories that cause less collision. Similarly, the light condition variable contains feature values mostly outside of the positive range. The number four and five indicates dusk and dawn times, which is around 8 pm and 5 am, respectively. Most of the accidents happen during rush hours (6–10 am and 3–7 pm) and in daylight. Since dusk and dawn times are outside of the rush hour and the number of vehicles may be less than regular traffic, the likelihood of an accident is less compared to other times. For the age and alcohol involvement variables, negative and positive classes have a

similar distribution. The mean and standard deviation of the age variable is 38.35 ± 20.06 for accident data and 43.37 ± 16.33 for non-accident (negative) data. All categorical variables in the US dataset are encoded as an integer by US National Highway Traffic Safety Administration [37]. The descriptions of the features are provided in Appendix A. The distribution of other variables is illustrated in Appendix B.
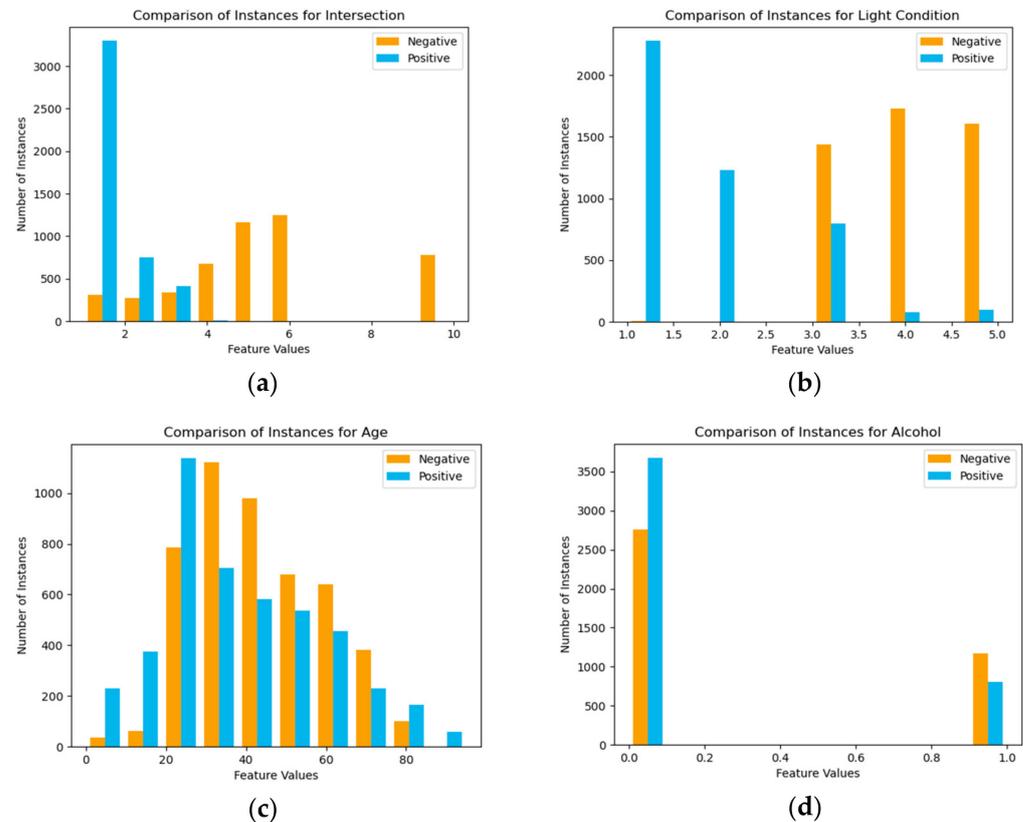

Figure 3. The distribution of variables for accident and non-accident data. (**a**) Intersection, (**b**) Light Condition, (**c**) Age, (**d**) Alcohol Involvement.

4.3. Experimental Results, Comparisons, and Discussion

This section presents the experimental results of the accident datasets. In all experiments, 10-fold cross-validation is applied to avoid the effect of randomness. The dataset randomly is divided into 80% for training and 20% for testing. Using NN hyperparameters, different architectures were created and after the experiments, the top eight NN architectures were presented in Table 4. The learning rate was taken as adaptive. SGD and Adam algorithms were used as a solver. Tanh and ReLU activation functions were adopted for hidden layers. After various experiments, three and five hidden layers with a different number of neurons were also used in our experiments. To examine and compare the performances of different neural network architectures along with ordinal regression models and three existing methods, mean squared error (MSE), and class accuracy (ACC), are used as performance evaluation criteria. The results are shown in Table 4 with MSE values, and the best scores are shown in bold. Blue-colored rows and white-colored rows present MSE values for the US and UK datasets, respectively. Three hidden layers architecture adopted seventeen neurons each and hyperbolic tangent activation function and Adam solver provided the best MSE score, which is 0.254 ± 0.038 for the US (third architecture) dataset and 0.173 ± 0.016 for the UK (third architecture) dataset.

Table 4. Classification results of different neural network architectures. Blue color: US dataset, White color: UK dataset.

Architecture	Hidden Layer	Neuron	Solver	Activation Function	MSE
1		17 neuron each	SGD	ReLu	US dataset: 0.264 ± 0.040 [a]
					UK dataset: 0.252 ± 0.078
2		17 neuron each	SGD	Tanh	0.258 ± 0.053
					0.297 ± 0.081
3		17 neuron each	Adam	Tanh	**0.254 ± 0.038**
	3				**0.173 ± 0.016**
4		50 neuron each	SGD	Tanh	0.283 ± 0.044
					0.208 ± 0.054
5		100, 50, 25	Adam	Tanh	0.368 ± 0.026
					0.176 ± 0.027
6		100, 50, 25	SGD	Tanh	0.311 ± 0.037
					0.183 ± 0.034
7		25, 50, 50, 50, 100	SGD	Tanh	0.283 ± 0.035
	5				0.236 ± 0.051
8		100 neuron each	Adam	Tanh	0.339 ± 0.030
					0.175 ± 0.024

[a] Mean Squared Error ± Standard Deviation.

Increasing the number of hidden layers adversely affected MSE scores. Our results show that it is not necessary to have too many hidden layers in the neural network (seventh and eighth architectures) to obtain good prediction performance. Therefore, the experiments were diversified into three hidden layers using different optimizers and activation functions. Throughout the experiments, it is observed that the hyperbolic tangent activation function provided better performance.

The bagging method, also called bootstrap aggregating, is an ensemble meta-algorithm, proposed by [39] to improve the performance of the weak classifiers. In the current study, the bagging method is also implemented in both datasets by applying ordinal regression models. Performance results of the bagging method on ordinal regression models are shown in Table 5 with MSE scores. Logistic all-threshold and ordinal ridge achieved the best MSE score for US and UK datasets. Specifically, the bagging method provided a 2.1% and 2.4% improvement compared to the no-bagging method for both datasets. The comparison among ordinal regression models indicates that the UK dataset performed the best MSE.

Comprehensive comparisons among NNs, ordinal regression, and three existing methods, namely decision tree, Linear SVM, and logistic regression are presented in Table 5. Table 5 indicates NN outperformed ordinal regression algorithms and other methods with lower MSE and higher accuracy scores. For the US dataset, the best and worst NN architecture outperformed other methods. DT has the second-best MSE score for the UK dataset. The third architecture with three hidden layers presents the best MSE scores and the highest accuracy values. In the US dataset, ordinal regression algorithms have the worst performance values among all methods. Feature values associated with the US dataset overlap each other. Thereby, ordinal regression algorithms are not successful in distinguishing them and predicting injury severity classes. Specifically, the logistic IT method has the worst MSE for the US dataset since the algorithm only predicted three classes among six classes. This algorithm failed to classify injury severities as multiclass, which is not desired in an advanced transportation safety system. When comparing the performance of machine learning models on two different data, we infer that UK data have better prediction results since it has distinguishable feature values that reduce the overlapping among classes and increase the model performance. Our experiments support that NNs are more robust than other existing and ordinal regression models for accident

prevention systems, as NN provides higher accuracy values per class and lower MSE scores on both datasets despite their differences.

Table 5. Classification results and comparison of machine learning algorithms.

Data	Method		MSE	Class Accuracy					
				Class 0	Class 1	Class 2	Class 3	Class 4	Class 5
US Dataset	NN	# 3 (Best)	**0.254 ± 0.038**	**0.963 ***	0.974	**0.977 ***	0.820	**0.979 ***	**1.000 ***
		# 5 (Worst)	0.368 ± 0.026	0.923	**0.978 ***	0.977	**0.834 ***	0.970	0.999
	OR Models	Ordinal Ridge	NB: 1.177 ± 0.097 B: 1.158 ± 0.094	0.178	0.262	0.289	0.426	0.238	0.703
		LAD	1.193 ± 0.106 1.174 ± 0.102	0.332	0.255	0.321	0.426	0.237	0.829
		Logistic IT	1.793 ± 0.189 1.686 ± 0.184	0.927	0.000	0.000	0.000	0.917	0.997
		Logistic AT	0.948 ± 0.135 0.928 ± 0.132	0.701	0.269	0.220	0.195	0.762	0.993
	Other Methods	DT Linear SVM LR	0.472 ± 0.136 0.797 ± 0.067 0.773 ± 0.043						
UK Dataset	NN	# 3 (Best)	**0.173 ± 0.016**	**0.833 ***	**0.658 ***	0.969			
		# 2 (Worst)	0.297 ± 0.081	0.829	0.556	0.895			
	OR Models	Ordinal Ridge	NB: 0.372 ± 0.025 B: 0.363 ± 0.022	0.620	0.534	0.771			
		LAD	0.585 ± 0.035 0.501 ± 0.092	0.451	0.141	**0.974 ***			
		Logistic IT	0.438 ± 0.062 0.426 ± 0.059	0.624	0.272	0.890			
		Logistic AT	0.396 ± 0.022 0.387 ± 0.023	0.620	0.430	0.831			
	Other Methods	DT Linear SVM LR	0.205 ± 0.052 0.387 ± 0.071 0.430 ± 0.038						

NB: No Bagging, B: Bagging, * Indicates the best class accuracy.

In Pradhan and Sameen's study [21], similar approaches are used to predict injury severity using real-world data containing 1138 observations with seven explanatory variables. Their linear SVM method outperformed the deep neural network and other SVM models with a 71.34% accuracy score. Compared to our study, higher accuracy scores are obtained using two different real-world datasets that contain more (30,484 and 14,593, respectively) observations with seventeen and ten explanatory variables. In addition, we provided more consistent and unbiased predictions by removing post-crash-related features such as collision type, number of fatalities, etc.

Confusion matrices for the neural network and best performed ordinal regression algorithms are shown in Figures 4 and 5 for the US and the UK data, respectively. These matrices show how classes often are confused with each other. For example, while serious class (class 1) is often confused with fatal injury (class 2) class by having 625 misclassified instances, the ordinal ridge algorithm identified slight injury (class 0) well, as seen in Figure 4 left. The confusion matrix in Figure 5 right presents that NN performed better by successfully classifying slight and fatal injury levels. Figure 5 shows class 0 (no apparent injury), class 4 (fatal injury), and class 5 (no accident) were predicted well by the logistic all-threshold method. Other classes were confused with each other. NN confusion matrix indicated that all classes predicted well with high true positive instances.
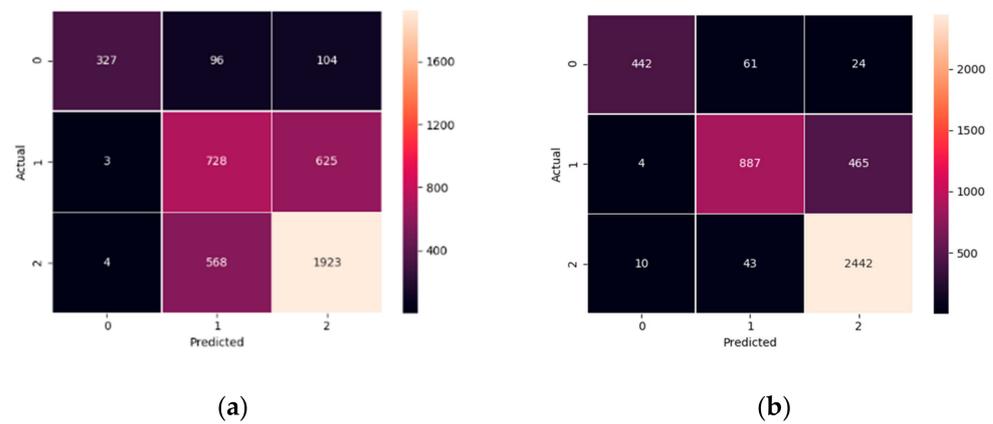
(**a**)　　　　　　　　　　　　　　　　　　(**b**)

Figure 4. The UK dataset confusion matrices. (**a**) Ordinal Ridge, (**b**) NN.
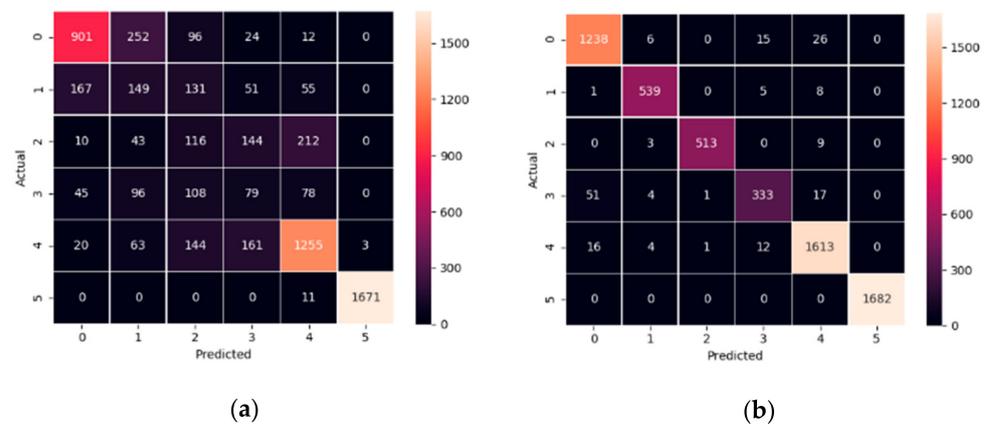

(**a**)　　　　　　　　　　　　　　　　　　(**b**)

Figure 5. The US dataset confusion matrices. (**a**) Logistic All-threshold, (**b**) NN.

5. Conclusions

In this paper, we proposed an accident prevention system, which is a significant matter in developing advanced transportation safety. Provided that injury severity levels cause deaths or disability, predicting accident risks and timely precautions could reduce casualties and increase safety in society. To provide a robust prediction model, we investigated the use of a deep neural network and the effect of its hyperparameters in estimating injury severity. We also generated non-accident data based on positive instances by using feature weights. Hence, we overcome the disadvantages of traditional sampling techniques and imbalanced learning by proposing a naïve data generator. Experimental results on two real-world datasets from the US National Highway Traffic Safety Administration and UK Transport for Greater Manchester are used to demonstrate the feasibility and robustness of our proposed framework. The study also analyzed the effect of data distribution and quality on the model performance. The differences in the data sets in terms of the number of classes and features and the distinguishability characteristics of the explanatory variables affected the model performance. All models have achieved better performance in the UK dataset compared to the US dataset. The US dataset has many overlapping instances and features value that belongs to different classes whereas the UK dataset has more distinguishable feature values that ease the classification problem.

We investigated the effect of hyperparameters of NNs on prediction performance. We analyze the number of hidden layers, the number of hidden neurons, activation functions, and optimizers. Our results show that it is unnecessary to have too many hidden layers (e.g., three hidden layers is good enough) in the NN to obtain a good prediction performance on injury severity. An increase in the number of hidden layers caused overfitting, which decreased the models' performance, by learning details and noise in the training set. Tanh activation function and Adam optimizer also showed better performance than other

activation functions and optimizers. Moreover, our comprehensive empirical performance comparison shows that NN outperforms four variants of ordinal regression and existing methods based on the MSE and accuracy measures on both datasets. Hence, a 3-hidden layered NN risk prediction model can be added to the proposed accident prevention and alert system in intelligent vehicles to alert drivers and trigger safety functions to reduce the risks of accidents.

The proposed prediction framework can be integrated into an accident prevention and alert system to be used by drivers. Additionally, we defined significant factors and patterns causing road accidents. These patterns as well as driver behavior patterns can assist the real-time alerting messages to reduce the accidents and develop a better design of autonomous vehicles and enhance advanced transportation safety in future work. Future work should also include the integration of object detection in the system to alert drivers of inevitable events, especially in blind spots.

Appendix A

Table A1. Description of independent variables for the US dataset.

Variable	Description	
Atmospheric Condition	1—Clear 2—Rain 3—Sleet 4—Snow	5—Fog 6—Severe crosswinds 10—Cloudy
Holiday Related	0—No Holiday 1—New Year 2—M. Luther King 3—JR Day 4—President's Day 5—Memorial Day	6—Independence Day 7—Labor Day 8—Veterans Day 9—Thanksgiving 10—Christmas
Light Condition	1—Daylight 2—Dark 3—Dark-lighted	4—Dawn 5—Dusk
Intersection Type	1—Not intersection 2—Fourway 3—T-intersection 4—Y-intersection	5—Traffic circle 6—Roundabout 10—L-intersection
Traffic Lane	1–7—Actual number of lanes in a road	

Table A1. *Cont.*

| Variable	Description	
Age	001–120—Actual ages	
Person Type	1—Driver 2—Passenger	
Sex	1—Male 2—Female	
Travel Speed	000–151—Reported speed up to 151 mph 998—Not Reported 999—Unknown	
Vehicle Make	01–94—Actual make 97—Not reported	98—Other make 99—Unknown make
Alcohol Involvement	0—No 1—Yes	
Surface Condition	1—Dry 2—Wet 3—Snow	4—Ice 5—Sand
Surface Type	1—Concrete 2—Asphalt 3—Brick	4—Stone 5—Dirt

Appendix B
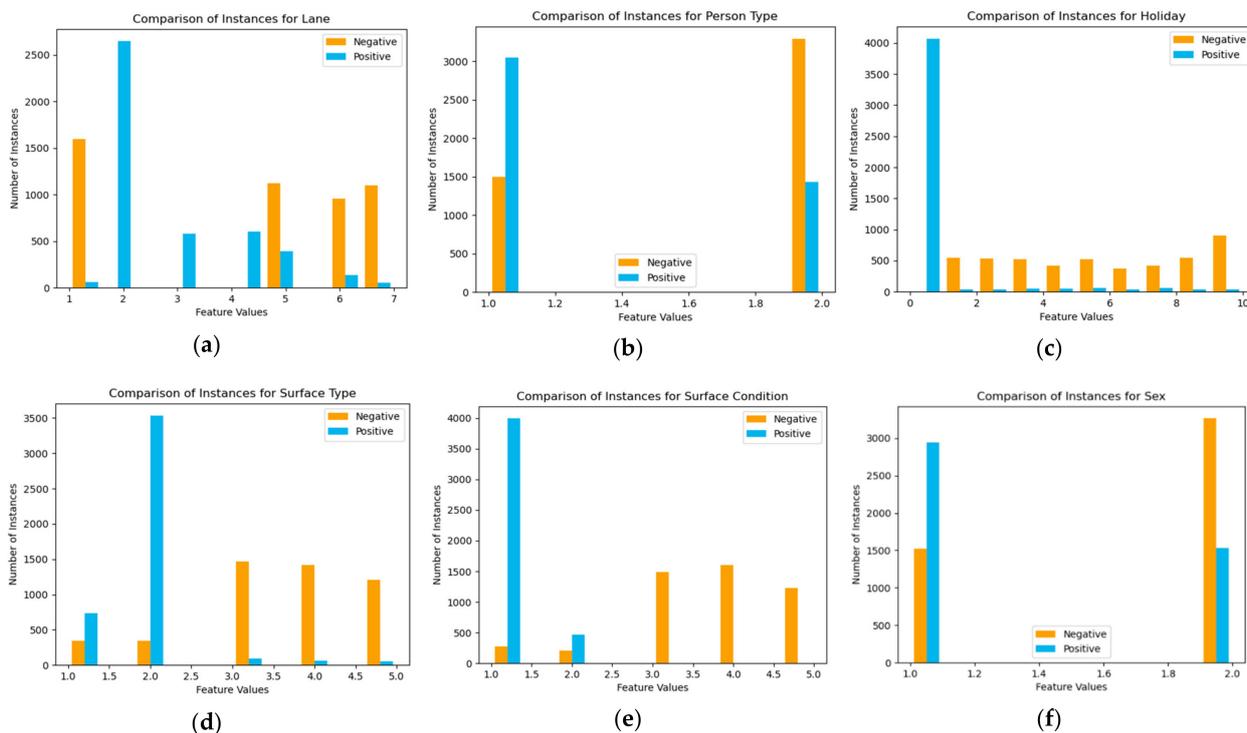
Figure A1. The distribution of the most important variables of the US dataset. (**a**) Lane, (**b**) Person Type, (**c**) Holiday, (**d**) Surface Type, (**e**) Surface Condition, (**f**) Sex.

References

1. National Center for Statistics and Analysis. 2015 Motor Vehicle Crashes: Overview. *Traffic Saf. Facts Res. Note* **2016**, *2016*, 1–9.
2. Han, S.; Wang, X.; Xu, L.; Sun, H.; Zheng, N. Frontal object perception for intelligent vehicles based on radar and camera fusion. In Proceedings of the 35th Chinese Control Conference (CCC), Chengdu, China, 27–29 July 2016. [CrossRef]

3. Alicioglu, G.; Sun, B.; Ho, S.S. Assessing accident risk using ordinal regression and multinomial logistic regression data generation. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020. [CrossRef]

4. Severino, A.; Pappalardo, G.; Curto, S.; Trubia, S.; Olayode, I.O. Safety Evaluation of Flower Roundabout Considering Autonomous Vehicles Operation. *Sustainability* **2021**, *13*, 10120. [CrossRef]

5. Macioszek, E. Roundabout Entry Capacity Calculation—A Case Study Based on Roundabouts in Tokyo, Japan, and Tokyo Surroundings. *Sustainability* **2020**, *12*, 1533. [CrossRef]

6. Macioszek, E. The Comparison of Models for Critical Headways Estimation at Roundabouts. In *Contemporary Challenges of Transport Systems and Traffic Engineering Lecture Notes in Networks and Systems*; Macioszek, E., Sierpiński, G., Eds.; Springer: Cham, Switzerlands, 2017; Volume 2. [CrossRef]

7. Thabtah, F.A.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [CrossRef]

8. Mujalli, R.O.; Oña, J.D. A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks. *J. Saf. Res.* **2011**, *42*, 317–326. [CrossRef] [PubMed]

9. Yuan, Z.; Zhou, X.; Yang, T.; Tamerius, J. Predicting traffic accidents through heterogeneous urban data: A case study. In Proceedings of the International Workshop on Urban Computing (KDD), Halifax, NS, Canada, 13–17 August 2017.

10. Jeong, H.; Jang, Y.; Bowman, P.J.; Masoud, N. Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. *Accid. Anal. Prev.* **2018**, *120*, 250–261. [CrossRef]

11. Pérez-Ortiz, M.; Gutiérrez, P.A.; García-Alonso, C.R.; Salvador-Carulla, L.; Salinas-Perez, J.A.; Hervás-Martínez, C. Ordinal classification of depression spatial hot-spots of prevalence. In Proceedings of the 11th International Conference on Intelligent Systems Design and Applications, Cordoba, Spain, 22–24 November 2011. [CrossRef]

12. Aci, C.; Ozden, C. Predicting the severity of motor vehicle accident injuries in Adana-Turkey using machine learning methods and detailed meteorological data. *Int. J. Intell. Syst. Appl. Eng.* **2018**, *6*, 72–79. [CrossRef]

13. Wang, Y.; Ho, I.W. Joint Deep Neural Network Modelling and Statistical Analysis on Characterizing Driving Behaviors. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018. [CrossRef]

14. Kahng, M.; Andrews, P.Y.; Kalro, A.; Chau, D. ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 88–97. [CrossRef]

15. Chatzimparmpas, A.; Martins, R.M.; Jusufi, I.; Kucher, K.; Rossi, F.; Kerren, A. The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations. *Comput. Graph. Forum* **2020**, *39*, 713–756. [CrossRef]

16. Azodi, C.B.; Tang, J.; Shiu, S. Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends Genet.* **2020**, *36*, 442–455. [CrossRef]

17. Çodur, M.Y.; Tortum, A. An Artificial Neural Network Model for Highway Accident Prediction: A Case Study of Erzurum, Turkey. *Promet-Traffic Transp.* **2015**, *27*, 217–225. [CrossRef]

18. Chong, M.; Abraham, A.; Paprzycki, M. Traffic accident analysis using machine learning paradigms. *Informatica* **2005**, *29*, 89–98.

19. Iranitalab, A.; Khattak, A.J. Comparison of four statistical and machine learning methods for crash severity prediction. *Accid. Anal. Prev.* **2017**, *108*, 27–36. [CrossRef] [PubMed]

20. Zhu, M.; Li, Y.; Wang, Y. Design and experiment verification of a novel analysis framework for recognition of driver injury patterns: From a multi-class classification perspective. *Accid. Anal. Prev.* **2018**, *120*, 152–164. [CrossRef]

21. Pradhan, B.; Sameen, M.I. Modeling Traffic Accident Severity Using Neural Networks and Support Vector Machines. In *Laser Scanning Systems in Highway and Safety Assessment*; Springer: Cham, Switzerlands, 2020; pp. 111–117. [CrossRef]

22. Delen, D.; Tomak, L.; Topuz, K.; Eryarsoy, E. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. *J. Transp. Health* **2017**, *4*, 118–131. [CrossRef]

23. Liao, Y.; Zhang, J.; Wang, S.; Li, S.; Han, J. Study on Crash Injury Severity Prediction of Autonomous Vehicles for Different Emergency Decisions Based on Support Vector Machine Model. *Electronics* **2018**, *7*, 381. [CrossRef]

24. Zeng, Q.; Huang, H. A stable and optimized neural network model for crash injury severity prediction. *Accid. Anal. Prev.* **2014**, *73*, 351–358. [CrossRef]

25. Fernández-Navarro, F.; Campoy-Muñoz, P.; Paz-Marin, M.L.; Hervás-Martínez, C.; Yao, X. Addressing the EU sovereign ratings using an ordinal regression approach. *IEEE Trans. Cybern.* **2013**, *43*, 2228–2240. [CrossRef]

26. Landschoot, S.; Waegeman, W.; Audenaert, K.; Haesaert, G.; Baets, B.D. Ordinal regression models for predicting deoxynivalenol in winter wheat. *Plant Pathol.* **2013**, *62*, 1319–1329. [CrossRef]

27. Gao, X.; Feng, Y. Penalized weighted least absolute deviation regression. *Stat. Its Interface* **2018**, *11*, 79–89. [CrossRef]

28. Xia, F.; Zhou, L.; Yang, Y.; Zhang, W. Ordinal regression as multiclass classification. *Int. J. Intell. Control. Syst.* **2007**, *12*, 230–236.

29. Zahid, F.M.; Ramzan, S. Ordinal ridge regression with categorical predictors. *J. Appl. Stat.* **2012**, *39*, 161–171. [CrossRef]

30. Aggarwal, C.C. *Neural Networks and Deep Learning*; Springer: Cham, Switzerlands, 2018. [CrossRef]

31. Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Prentice Hall: New York, NY, USA, 2009.

32. Kalogirou, S.A. *Solar Energy Engineering*, 2nd ed.; Elsevier: Amsterdam, The Netherlands; Academic Press: Cambridge, MA, USA, 2014. [CrossRef]

33. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 2007. [CrossRef]

34. Bottou, L. Stochastic Gradient Descent Tricks. In *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*; Montavon, G., Orr, G.B., Müller, K.R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7700, pp. 421–436. [CrossRef]

35. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR, San Diego, CA, USA, 7–9 May 2015.
36. Williams, R. Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *Stata J.* **2006**, *6*, 58–82. [CrossRef]
37. National Highway Traffic Safety Administration. Available online: https://www-fars.nhtsa.dot.gov (accessed on 18 February 2019).
38. UK Transport for Greater Manchester. Available online: https://data.gov.uk/ (accessed on 20 December 2019).
39. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]