Accident Analysis and Prevention 120 (2018) 211–218

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Predicting interstate motor carrier crash rate level using classification models

Jundi Liu[a], Linda N. Boyle[a,b,*], Ashis G. Banerjee[a,c]

[a] Industrial & Systems Engineering, University of Washington, Seattle, WA 98195, USA
[b] Civil & Environmental Engineering, University of Washington, Seattle, WA 98195, USA
[c] Mechanical Engineering, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Keywords:
Motor carrier safety
Crash rate
Variable selection
Classification and Regression Tree

ABSTRACT

Ensuring safe operations of large commercial vehicles (motor carriers) remains an important challenge, particularly in the United States. While the federal regulatory agency has instituted a compliance review-based rating method to encourage carriers to improve their safety levels, concerns have been expressed regarding the effectiveness of the current ratings. In this paper, we consider a crash rate level (high, medium, and low) rather than a compliance review-based rating (satisfactory, conditional satisfactory, and unsatisfactory). We demonstrate an automated way of predicting the crash rate levels for each carrier using three different classification models (Artificial Neural Network, Classification and Regression Tree (CART), and Support Vector Machine) and three separate variable selection methods (Empirical Evidence, Multiple Factor Analysis, Garson's algorithm). The predicted crash rate levels (high, low) are compared to the assigned levels based on the current safety rating method. The results indicate the feasibility of crash rate level as an effective measure of carrier safety, with CART having the best performance.

1. Introduction

In the United States (U.S.), large truck crashes are a leading cause of death and injuries every year, and result in billions of dollars in medical expenses and productivity loss. There are ongoing efforts at the federal and carrier level to help enhance the safety performance of all motor carriers. However, many safety management systems do not appear to be effective (Mooren et al., 2014). For that reason, it is important to identify the key indicators of safe commercial vehicle operations, which would lead to an effective classification of carrier safety levels.

The U.S. Department of Transportation (DOT) Federal Motor Carrier Safety Administration (FMCSA) conducts on-site compliance reviews of motor carrier operations based on several criteria that include the drivers' service hours, driver qualifications, maintenance and inspection records, and crash reports (Chen, 2008). The motor carriers are then assigned a unique safety rating. However, concerns have been expressed that, "the overall conclusion is that the worse a firm does on a large part of the audit, the better its accident record" (Moses and Savage, 1992). In fact, some of the studies on the effectiveness of the safety audits reveal that some inspection activities are unrelated to the actual safety performances of the motor carriers (Moses and Savage, 1992, 1994).

FMCSA has created the Compliance, Safety, Accountability (CSA) program, a data-driven safety compliance and enforcement program to improve safety and prevent commercial motor vehicle crashes, injuries, and fatalities (Volpe, 2013). This program includes data analyses generated from the original MCMIS dataset to identify non-compliant and unsafe companies to prioritize them for enforcement interventions. The Safety Measurement System (SMS) is a key component of CSA, which uses data from inspections and crash reports to identify and intervene with motor carriers that pose the greatest risk to safety. Many of the concepts used to construct the SMS originated from the SafeStat measurement system. SafeStat was developed under a project plan agreement with the Federal Highway Administrations (FHWA) Office of Motor Carriers, FMCSAs predecessor. It was designed and tested under the Federal/State Performance and Registration Information Systems Management (PRISM) program in the mid-1990s. From the mid-1990s until December 2010, when FMCSA replaced SafeStat with the SMS, SafeStat was implemented nationally to prioritize motor carriers for on-site compliance reviews (Volpe, 2013).

CSA organizes the data into seven Behavior Analysis and Safety Improvement Categories (BASICs). The SMS groups carriers by BASICs with other carriers that have a similar number of safety events and then assigns a percentile to prioritize them for interventions (Volpe, 2013).

* Corresponding author at: Industrial & Systems Engineering, University of Washington, Seattle, WA 98195, USA.
E-mail address: linda@uw.edu (L.N. Boyle).

https://doi.org/10.1016/j.aap.2018.06.005
Received 22 November 2017; Received in revised form 29 May 2018; Accepted 6 June 2018
Available online 29 August 2018

However, some studies show that BASICs are not always effective. For example, Lueck (2012) show that as a carriers' Driver Fitness record improves, its crash rate goes up.

Crashes involving commercial vehicles have greater severity, in terms of injuries and cost, than that of passenger vehicles (Blincoe et al., 2015). Rogers and Knipling (2007) showed that 48% of truck crashes that involved fatalities and injuries were not necessarily due to the faults of the commercial drivers. The study showed that in fatal crashes that involve a truck and a passenger vehicle, 44% of the truck drivers and 56% had a critical reason assigned to the other vehicle or driver. The Driver/Carrier Data Relationship Project showed that drivers with high citation rates were positively correlated with high crash rates (Murray et al., 2005).

Knipling (2009) also found major differences for the CRs assigned to the truck drivers between single-vehicle truck crashes and multiple-vehicle truck crashes. In single-vehicle truck crashes, speed, fatigue, vehicle failure and inattention were the most important factors, whereas in multiple-vehicle crashes, inattention, inadequate surveillance, speed, and illegal maneuvers were most important. Data from the Large Truck Crash Causation Study (LTCCS) showed that 87% of crashes were caused by driver errors (Federal Motor Carrier Safety Administration, 2006a).

Findings from many carrier-based studies have been based on logistic regression models. Lantz and Loftus (2005) used a logistic regression model to examine the association between driver violations and carrier crash rates, and Blower and Green (2009) used this modeling technique with LTCSS data to show that vehicle defects (brakes, tires, steering) were significantly associated with crash events as they could negatively impact the ability of the driver to respond appropriately. They also showed that the maintenance of large trucks were an important contributor to large truck crashes.

The Motor Carrier Management Information System (MCMIS) dataset has also been used to identify predictors of crash rates, mostly using Poisson regression (Moses and Savage, 1992), negative binomial regression (Moses and Savage, 1994), and ordinary least squares (OLS) (Corsi et al., 1984; Corsi and Fanara, 1988a,b). The negative binomial is often preferred given the over-dispersed nature of the crash data. All of these models belong to the category of generalized linear models (GLMs). However, based on the FMCSA compliance reviews and crash rates, the motor carriers may be more appropriately categorized into subgroups, which lends itself to a classification rather than a regression approach. In recent years, models such as classification trees, support vector machines, and artificial neural networks have been used to successfully classify large and heterogeneous datasets comprising both continuous and discrete or categorical variables. By learning suitable model parameters and structures, they have been effective in many challenging computer vision, natural language processing, and user behavior detection applications. In this study, we consider the potential of classifying the safety outcomes of motor carriers using the MCMIS dataset, which assigns a unique safety rating for each carrier based on the U.S. DOT FMCSA compliance review. The research objective is, therefore, to examine the suitability of classification models in predicting carrier safety.

2. Materials and methods

2.1. Datasets

The MCMIS dataset assigns a unique safety rating for each carrier based on the U.S. DOT FMCSA compliance review (Federal Motor Carrier Safety Administration, 2017). The safety rating consists of three levels: satisfactory (S), conditional satisfactory (C), and unsatisfactory (U). Although not perfect, this rating provides information on the safety performance of the carriers. This database contains detailed records of all the reported crash events, census studies, and inspection results.

The Carrier Safety Measurement System (CSMS) is another system that serves as a potential carrier safety identification tool. There are four tables used in this study: three are from MCMIS (REVIEW, CENSUS and CRASH_MASTER) and one is from CSMS (CARRIER).

The CRASH_MASTER table contains 70 data elements for 3,022,849 crash events from 1989 to 2015. After removing all the crash events with missing DOT numbers, 2,073,489 events are available for a total of 410,259 different motor carriers. The data elements include the location and time of any specific crash event, basic information of the vehicles involved in the crash, environmental condition (weather, road, etc.) of the crash event, detailed information about the reporting process, and injuries and fatalities associated with the crash event. 12 of these elements are of interest to us, which include the severity of the crash, configurations of the crash vehicles, and environmental conditions. Detailed data element names and definitions are available at https://ask.fmcsa.dot.gov/app/mcmiscatalog/d_crash3 (see Table A.11 in Appendix). Because of the size of the dataset, all crash events with missing fields are eliminated. After removing the missing values, there are 1,274,472 crash events related to 293,788 motor carriers.

The REVIEW table contains information on the compliance review, which is then used for the FMCSA designated safety ratings of the carriers. Three (3) safety ratings are listed in the MCMIS database: overall rating (RATING_OVERALL), provisional rating (PROVISIONAL_RATING), and safety rating (SAFETY_RATING). Little information is available for the provisional rating. Hence, RATING_OVERALL and SAFETY_RATING is considered for the subsequent analysis.

The review dates for these ratings vary from 1987 to 2010. The overall rating is available only until 2002, at which time it is no longer collected. Fig. 1a shows the percentage of carriers in each overall rating level from 1990 to 2001. Compared to the overall rating, the safety rating is maintained up to the most recent years, and is, thus, more useful for our purpose. Fig. 1b shows the percentage of carriers in each safety rating level from 2003 to 2010. Both figures show that the proportion of carriers in the different rating levels remain more or less stable over the observed time periods.

The CENSUS table contains the carrier operation types; this includes interstate carriers, intrastate carriers transporting hazardous materials, and intrastate carriers transporting non-hazardous materials. In total, there are 2,464,729 motor carriers with unique operations. 61.9% (n = 1,525,993) of the carriers provide interstate services, and the other 38.1% (n = 938,736) provide intrastate services, of which 96.7% (n = 908,207) are involved in non-hazardous materials transportation.

The CARRIER table contains the average power units for each motor carrier and is defined as the weighted averages of the power units as reported six (6) months and eighteen (18) months prior to the snapshot date FMCSA (2018). This variable is used as a measure of carrier size. There are 1,640,195 carriers in this table.

After preprocessing, a training dataset is constructed from the four tables by merging the unique DOT numbers for each motor carrier. After merging, the crash events of the most recent 6 years (from 2010 to 2015) is pulled out and the crash rate levels are assigned. As noted earlier, there is no overall rating as of 2002. We also note that the safety rating does not change substantially from 2003 to 2010. Both of the ratings follow a stable trend for different compliance levels. Therefore, it is reasonable to assume that the percent of carriers in different safety levels do not change much in the observed years. For overall rating, the average percent is 55.2% for S, 30.4% for C, and 14.4% for U. For safety rating, the corresponding average percents are 67.5%, 27.1%, and 5.4% for S, C, and U, respectively. Moving forward, the accuracy of the predictions for the crash rate levels will be based on SAFETY_RATING given that more recent data is available.

2.2. Calculating crash rates

We hypothesize that using the crash rate levels is a better indicator of carrier safety since it captures the actual crash rate characteristics. The crash rate of a carrier is defined in Eq. (1). We note here that a

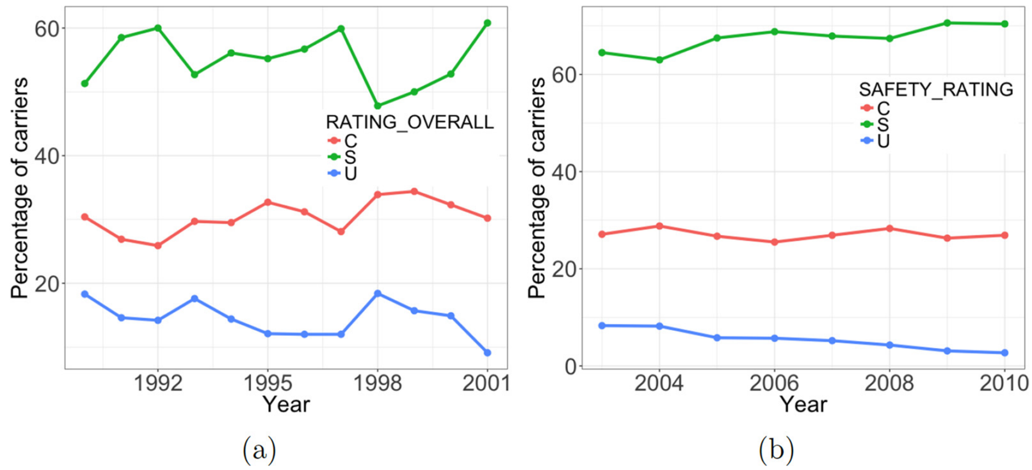

Fig. 1. Proportion of motor carriers given the overall rating levels from 1990 to 2001 (a) and safety rating levels from 2003 to 2010 (b).

Table 1
Mean crash rates per hundred power units for each rating level.

MCMIS variable name	Years	S	C	U
RATING_OVERALL	1991–2002	1.10	0.45	0.38
SAFETY_RATING	2003–2010	1.26	0.54	0.36

Note: S = Satisfactory, C = Conditional Satisfactory, U = Unsatisfactory.

satisfactory (S) rating level does not necessarily mean the carrier will have a low crash rate. In fact, Table 1 shows the exact opposite trend in that the average crash rates for both rating methods actually decrease progressively from S to C to U. We therefore use the computed crash rates to create categories of crash rate levels, denoted as (High [H], Medium [M], and Low [L]). Initially, the percent of carriers categorized as H, M, and L is set to the same percents observed for S, C, and U in the safety rating.

$$\text{crash rate} = \frac{\text{Number of crash events in } a \text{ given time period}}{\text{Carrier size}} \tag{1}$$

2.3. Classification models

It is difficult to create a universal classification model to predict crash rate level as different carrier types are assumed to have different safety performances. The subsequent analysis is focused on interstate truck carriers as they represent over 75% of all motor carriers. After cleaning, the final dataset included 281,716 crash events from 81,340 interstate truck carriers. There were ten predictor variables (Table 4) and one outcome variable (crash rate level).

Classification tree, Artificial Neural Network (ANN), and Support Vector Machine (SVM) are widely used classification models and are considered in this study. They share the common goal of mapping the input (predictor) variables to outcomes (responses) that are categorical variables.

Classification tree uses a tree structure to partition the dataset recursively, and assigns specific class types (labels) at the leaf nodes of the tree for different conjunctions[1] of the predictor variables values at the interior tree nodes. In other words, an interior node splits up or branches off into two or more child nodes such that all the constraints on the predictor variables for the node under consideration as well as its parent nodes are satisfied. The splitting terminates when no further data points or observations are found satisfying all the constraints, yielding leaf nodes with labeled classes. The constraints are typically of the forms that the variables are greater than, or less than or equal to specified values, or lie within defined ranges. We specifically employ the Classification and Regression Tree (CART) algorithm, which uses Gini impurity to identify the most suitable predictor variable for splitting any interior node into two branches. Gini impurity, $I_G(f)$, computes the probability of incorrectly classifying a randomly chosen sample as shown in Eq. (2).

$$I_G(f) = \sum_{i=1}^{n} f_i(1 - f_i) = 1 - \sum_{i=1}^{n} f_i^2 = \sum_{i \neq k} f_i f_k \tag{2}$$

Here, n is the number of classes, $i \in \{1, b, \ldots, n\}$, and f_i is the fraction of samples belonging to class i. Using the CART algorithm, the classification tree model can be trained efficiently for both two and three-class crash rate level prediction.

Artificial Neural Network (ANN) is another widely-used classification model that assumes a more complicated structure and is learned using a variety of methods. It comprises a layer of input nodes, one or more layers of hidden nodes, and a layer of output nodes. The hidden nodes perform (typically) non-linear operations on the sum of all the inputs, often with weights to adjust the strengths of the inputs, and sometimes with thresholds to process only those inputs whose strengths exceed the thresholds. The nodes mimic animal brain neurons and require three sets of hyperparameters to define the complete network. These hyperparameters are the connections among the nodes, connection weights, and an activation function to transform any node's weighted inputs to an output in a smooth manner. The first two hyperparameters are automatically learned during the training process. For the third hyperparameter, we select the sigmoid function, a widely used activation function (see Eq. (3)).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

The training process for ANN is more time consuming than for the classification tree model because it is inherently more complex. Hence, many efficient learning algorithms like the backpropagation algorithm, evolutionary methods (Castelletti et al., 2005), and simulated annealing (Da and Xiurun, 2005) have been developed for this purpose. We adopt a basic backpropagation algorithm as studies have shown it to be efficient for a wide variety of problems (Rumelhart et al., 1988).

Support Vector Machine (SVM) is based on the idea that classification of data points becomes easier if they can be mapped to a high-dimensional feature space. This model includes a suitable set of features defined as functions of the predictor variables based on the underlying structures present in the data. Since explicit mapping of points to the

[1] A conjunction is a logical operator that returns true only when all of its operands are true.

Table 2
Confusion matrix. Assigned class refers to the computed data label (carrier crash rate level) and predicted class denotes the predicted label using one of the classification models discussed above.

	Assigned class	
	Positive	Negative
Predicted class		
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

feature space is quite challenging, a solution is to employ a suitable kernel function to implicitly map the data to the feature space using just the inner products of pairwise combinations of the points. This kernel function, also known as the similarity function, quantifies the similarity or closeness between any two points, and typically assumes some form of an inverse of a distance metric. The intuition behind this operation is that similar points cluster together in the feature space (even though these clusters are not detected easily in the input space), enabling us to obtain separating hyperplanes to differentiate the clusters, thereby, classifying the points appropriately. Eq. (4) shows the mathematical representation of the radial basis function (RBF) kernel K used in this paper, where $||\mathbf{x} - \mathbf{x}'||^2$ denotes the squared Euclidean distance between two input data vectors $\mathbf{x}$ and $\mathbf{x}'$, and σ is a free parameter.

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right). \tag{4}$$

Once the classification models are trained, confusion matrices are generated for systematic evaluation using several performance metrics that include overall prediction accuracy, precision, recall, and F_1 score for a particular class. (see Table 2 and Eq. (5)). Accuracy provides an overall measure of how many crash events are correctly predicted in their assigned crash rate levels. Precision, also known as Positive Predictive Value (PPV), is the proportion of crash events that are correctly classified in any predicted crash rate level. In other words, it enumerates how many of the identified events in any crash rate level are actually relevant (true members of that level). Recall, also termed as True Positive Rate (TPR), is the proportion of correctly predicted crash events for a given (assigned) crash rate level. Therefore, it measures how many relevant events are identified for the corresponding crash rate level. The F_1 score is the harmonic mean of PPV and TPR, with higher values indicating greater model accuracy for a particular crash rate level.

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{Precision (PPV)} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{Recall (TPR)} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ F_1 = 2\cdot\frac{\text{PPV}\cdot\text{TPR}}{\text{PPV} + \text{TPR}} &= \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \end{aligned} \tag{5}$$

2.4. Variable selection

Reducing the dimensionality of the input (predictor) variables through variable selection is an important step for classification models. In particular, for large datasets, removing the redundant dimensions decreases model training time significantly without affecting performance.

Empirical evidence is often used as an efficient and relatively accurate method for variable selection. The variables, INJURIES and FATALITIES, provide a similar measure of the severity of a crash event. However, many of the crash events do not result in injuries or fatalities and, therefore, result in no values being recorded. We exclude them from the initial set of predictor variables. Of the remaining eight variables, CARGO_BODY_TYPE_ID (Table A.12) and VEHICLE_CONFIGURATION_ID (Table A.13) both describe vehicle types. We retain the vehicle configuration as they are more indicative of the characteristics of interstate carriers.

Table 3
Spearman correlation analysis of weather condition and road surface condition.

Weather condition	Road surface condition	rho	*p*-Value
No adverse condition	Dry	0.73	< 0.0001
Rain	Wet	0.73	< 0.0001
Snow	Snow	0.64	< 0.0001

Table 4
The four groups of predictor variables based on Multiple Factor Analysis.

Predictor variables	Variable type	Groups
INJURIES	Continuous	Severity
FATALITIES	Continuous	
AVG_POWER_UNIT	Categorical	Carrier
CARGO_BODY_TYPE_ID	Categorical	Vehicle
VEHICLE_CONFIGURATION_ID	Categorical	
ACCESS_CONTROL_ID	Categorical	
TRAFFICWAY_ID	Categorical	Environment
LIGHT_CONDITION_ID	Categorical	
WEATHER_CONDITION_ID	Categorical	
ROAD_SURFACE_CONDITION_ID	Categorical	

ROAD_SURFACE_CONDITON_ID and WEATHER_CONDITON_ID are another pair of strongly correlated predictors. To quantify the correlation between them, we carry out the Spearman correlation analysis. It is high if two variables are similar in their relative position labels of the observations within the variable. Furthermore, statistical testing is conducted to test whether an observed value of the correlation coefficient ρ is significantly different from zero. Table 3 confirms that ROAD_SURFACE_CONDITON_ID and WEATHER_CONDITON_ID are strongly correlated.

Multiple Factor Analysis (MFA) is an extension of Principal Component Analysis (PCA). It is used to reduce the dimensionality of the predictors by clustering them into groups. The groups are constructed based on a) the relationships among the predictors; and b) the similarities among the carriers with respect to all the predictors. It is a useful technique for our dataset given the complexity of our data tables, and the inclusion of both quantitative and qualitative data. In our case, the ten predictor variables are divided into four groups: severity, carrier, vehicle, and environment (see Table 4).

Variable selection in ANN is conducted using Garson's algorithm. Garson proposed an algorithm (Garson, 1991) to calculate the relative importance of the input node of a neural network that was later modified by Goh (1995). Eq. (6) calculates this relative importance, R_{ik}, of the input node i with respect to the output node k in a single hidden layer neural network, where w_{ij} denotes the weight connecting the ith input node to the jth hidden node and v_{jk} denotes the weight connecting the jth hidden node to the kth output node. The input variables with the highest relative importances are selected.

$$R_{ik} = \frac{\sum_{j=1}^{L} |w_{ij}v_{jk}| / \sum_{r=1}^{N} |w_{rj}|}{\sum_{i=1}^{N}\sum_{j=1}^{L} (|w_{ij}v_{jk}| / \sum_{r=1}^{N} |w_{rj}|)} \tag{6}$$

3. Results

The experiments were carried out on a MAC Sierra version 10.12.4 Operating System using the R programming language version 3.3.2. A 2.6 GHz Intel Core i7 quad-core processor with 16 GB 2133 MHz LPDDR3 RAM was used. R packages nnet (Venables and Ripley, 2002) and NeuralNetTools (Beck, 2016), caret (Kuhn, 2016), e1071 (Meyer et al., 2015), and other supportive packages like ggplot2 (Wickham,

Table 5
Confusion matrices for the 2-level classifiers (L,H) using SAFETY_RATING as the categorization method with all ten predictor variables.

Model name		CART		ANN		SVM	
Predicted level		L	H	L	H	L	H
Assigned level	L	69,787	10,723	62,728	17,782	65,408	15,102
	H	43,766	56,540	40,438	59,868	43,890	56,416

2009) were employed for implementing ANN models, CART models, SVM models, and results visualization, respectively.

Classification models were built using the percentages computed from the variable, SAFETY_RATING, for the carrier crash rate level. The training set consisted of 100,900 crash events for 41,972 interstate carriers for the period 2010–2013. The test set comprised of 180,816 crash events for 58,464 interstate carriers for the period 2014 to 2015. Ten-fold cross validation was employed during the training process. Some of the carriers appeared in both the training and test data sets, for a total of 81,340 unique carriers. All classification models included the same set of ten predictor variables listed in Table 4.

For the ANN model, the prediction accuracy is based on the number of nodes in the hidden layer and the weight decay in the back-propagation algorithm. The parameters were selected based on empirical studies that resulted in the best training accuracy. 5 to 15 hidden nodes were examined, with 13 being the most suitable choice. The best decay parameter was 0.07 (0.025–0.10 was tested). We also tested the ANN model for different number of hidden layers. We selected the most parsimonious model as no appreciable improvement in prediction accuracy was observed even when the number of layers grew from 3 to 5.

A 3-level classification is initially considered for all the learning models, CART, ANN, and SVM. However, the performance of all the classifiers in the high crash rate level is substantially inferior when compared to the medium and low crash rate level due to the low number of carriers identified as having a high crash rate (5.4%). To achieve better results, we combine the medium and high crash rate carriers into one category, denoted as high crash rate (H).

The confusion matrices for the 2-level classification models are shown in Table 5. The diagonal elements are dominant in each row and column, which confirms that it is both reasonable and practical to combine the M and H level carriers into one level. Table 6 shows the performance metrics for the 2-level classification models. In general, the metrics are better than those for the 3-level classifiers. CART outperforms the ANN and SVM except for the recall rate of H level carriers. The F_1 scores for L and H level carriers are both above 65%.

The categorical variables included dummy variables during the training process. For the 10 predictors, we have 56 input variables with 53 of these being dummy variables. Hence, input dimension reduction is necessary to not only reduce the training time but also extract a concise set of variables that significantly affect classification performance.

The SVM training time is substantially higher than the other two models and Garson's algorithm is only applicable for the ANN model. Given that all the classifiers yielded comparable performance metrics, the variable selection methods were tested with ANN only. Based on Section 2.4, the input variables are first selected using Empirical Evidence, then MFA is considered, followed by Garson's algorithm. The resulting confusion matrix for the three algorithms are shown in Table 9, and the performance metrics and training times are reported in Table 10.

For the algorithm based on Empirical Evidence. the selected variables consist of TRAFFICWAY_ID, ACCESS_CONTROL_ID, WEATHER_CONDITION_ID, VEHICLE_CONFIGURATION_ID, LIGHT_CONDITION_ID, and AVG_POWER_UNIT. The MFA is applied to the ten predictor variables. Fig. 2 visualizes the analysis result, and Table 7 lists

Table 6
Performance metrics and training times for the classifiers with all the predictor variables.

Model name	Precision (%)		Recall (%)		F_1 (%)		Accuracy (%)	Training time (min)
	L	H	L	H	L	H		
CART	61.5	84.1	86.7	56.4	72	67.5	70	1.11
ANN	60.8	77.1	77.9	59.7	68.3	67.3	67.8	1.69
SVM	59.8	78.9	81.2	56.2	68.9	65.6	67.4	23.36

the coordinates of the variables, which are categorized into 4 groups (dimensions) based on their underlying relations as explained in Section 2.4. Each vector in Fig. 2 denotes the weight of the corresponding variable along the two most dominant (principal) factors. The proportion of information explained by a dominant factor is expressed as a percent. Therefore, the variables with maximum corresponding vector lengths (L_2 norms) are selected due to their high relative importances. The selected variables comprise INJURIES, AVG_POWER_UNIT, CARGO_BODY_TYPE_ID, and LIGHT_CONDITON_ ID.

Finally, we apply Garson's algorithm to the ANN model to compute the relative importance of each predictor variable. Since the categorical variables are represented as dummy variables, the relative importance is calculated by taking the average of the dummy variables. Table 8 presents the relative importance of each variable in descending order. The relative importance denotes the importance of a predictor in the training process. Larger the relative importance, the more significant is its role during model generation. It is expected that the variables with high relative importance values would have dominant effects in predicting carrier crash rate levels given our dataset and choice of classes. As an example, AVG_POWER_UNIT is expected to be the most significant variable in determining whether a particular carrier's crash events are classified[2] as L or H. Therefore, it is necessary to consider only the top few variables with highest importance values, and avoid the problem of model overfitting to our training data without compromising classification performance. These variables, consisting of ACCESS_CONTROL_ID, CARGO_BODY_TYPE_ID, VEHICLE_CONFIGURATION_ID, FATALITIES, LIGHT_CONDITION_ID and AVG_POWER_UNIT, are selected as the predictors.

From Table 10, both MFA and Garson's algorithm outperform Empirical Evidence except for the F_1 score of L-level carriers, thereby highlighting the effectiveness of the automated variable selection methods. MFA yields maximum precision for H-level carriers and highest recall for L-level carriers. Among all the methods, Garson's algorithm has the best accuracy. Since MFA selects the fewest number of variables, it has the least training time.

4. Discussion

This study provides a systematic comparison of various classification models using real-world interstate motor carrier data replete with missing values and uncertainty in crash level assignments. The focus of this paper is on predictive modeling (supervised machine learning) rather than traditional inferential models such as a Negative Binomial Model. That is, rather than identifying the highest risk carriers, the goal is to identify the most important predictors of crash risk. From Table 8, we observe that average power unit is the most important predictor of crash risk followed by cargo body type.

The results point out several noteworthy trends that demonstrate

[2] Note that high significance does not necessarily imply positive or negative correlation of the variable to either L or H class. Instead, it means that this variable is most useful (among all the available variables) in predicting whether the crash rate level would be L or H.

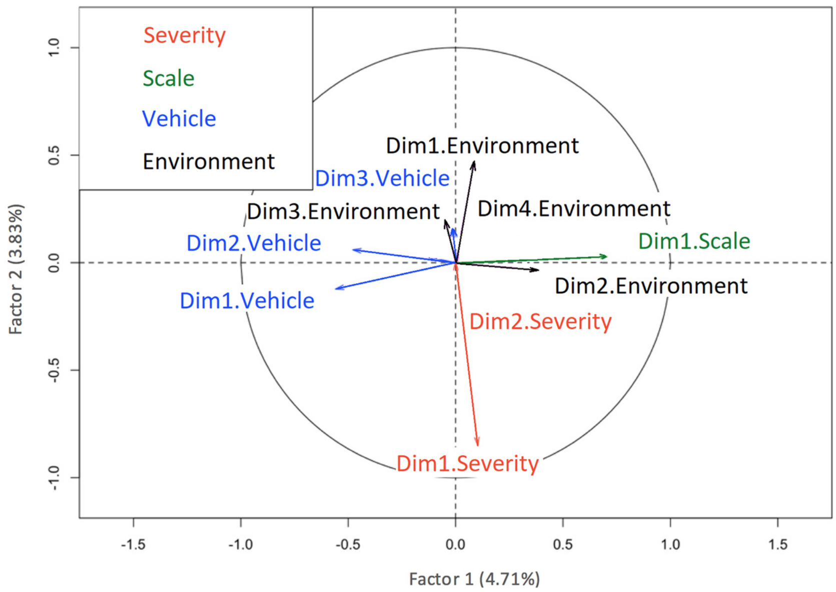

Fig. 2. Visualization of the MFA results with the 10 predictor variables categorized into four groups based on their underlying relations.

Table 7
MFA results listing the coordinates of all the grouped predictor variables with respect to the two principal factors.

MFA variable name	Corresponding variable name	Factor 1	Factor 2
Dim1.Severity	INJURIES	0.104	−0.851
Dim2.Severity	FATALITIES	0.000	0.004
Dim1.Scale	AVG_POWER_UNIT	0.705	0.028
Dim1.Vehicle	CARGO_BODY_TYPE_ID	−0.560	−0.123
Dim2.Vehicle	VEHICLE_CONFIGURATION_ID	−0.479	0.060
Dim3.Vehicle	ACCESS_CONTROL_ID	−0.013	0.158
Dim1.Environment	TRAFFICWAY_ID	0.087	0.471
Dim2.Environment	LIGHT_CONDITION_ID	0.386	−0.036
Dim3.Environment	WEATHER_CONDITION_ID	−0.050	0.199
Dim4.Environment	ROAD_SURFACE_CONDITION_ID	0.014	0.039

Table 8
Relative importance of the input variables using Garson's algorithm.

Variable name	Relative importance
AVG_POWER_UNIT	0.024
CARGO_BODY_TYPE_ID	0.021
FATALITIES	0.019
ACCESS_CONTROL_ID	0.018
VEHICLE_CONFIGURATION_ID	0.018
LIGHT_CONDITION_ID	0.015
TRAFFICWAY_ID	0.014
WEATHER_CONDITION_ID	0.014
INJURIES	0.012
ROAD_SURFACE_CONDITION_ID	0.011

the usefulness of the proposed method to promote safe carrier operations. The crash risk can be predicted using any of the three algorithms discussed. From a practitioners' perspective, a CART model would be the most reasonable choice when examining the carrier crash rates globally. It provides the best overall performance for 2-level classification (High, Low). In fact, all the three classification models were better in predicting 2-level crash rates when compared to 3-level crash rates, indicating that it is better to adopt the simplest possible bi-level classification of carrier safety given the available data. If more parsimonious models with fewer predictor variables are desired, the ANN model with either MFA or Garson's algorithm-based variable selection is suitable.

The crash rate levels assigned based on safety rating match well with the predicted levels, as evident from the dominant diagonal elements in the confusion matrices. This finding suggests that the crash rate (crashes/carrier size) could serve as a more effective measure of carrier safety as compared to the compliance review-based rating.

The performance metrics (Precision, Recall, and F_1 score) could be used to update the overall safety level of a carrier as new crash events unfold. Suppose there are x new crash events for a carrier over a certain time period. If x is sufficiently large, we can run the 2-level CART model to obtain y predictions of H. If the ratio y/x is the same as the CART precision value for H specified in Table 6 within some allowable margin, we can conclude that the current carrier safety rating is reasonable. Otherwise, the crash rating can be changed to "low" or a decision can be made to conduct additional reviews to verify that the high crash rating is still appropriate. This opens up a new way of dynamically rating carrier operation safety using actual crash events as the basis for deciding the ratings, which can alleviate some of the concerns of the existing rating mechanism.

The study was limited to the MCMIS dataset and the initial categories that were provided within these datasets. There are many factors that contribute to crashes, many of which may not be included in MCMIS (National Academies of Sciences, Engineering, and Medicine, 2017). Further, our analysis considered interstate carriers only and it is important to consider the findings of this study in that context. Differences among carrier types (Federal Motor Carrier Safety

Table 9
Confusion matrices for the ANN model using three different variable selection methods.

Model name		Empirical evidence		MFA		Garson's algorithm	
Predicted level		L	H	L	H	L	H
Assigned level	L	65,101	15,409	67,677	12,833	57,566	22,944
	H	45,182	55,124	48,548	51758	35,524	64,782

Table 10
Performance metrics and training times using the ANN model for the three variable selection methods.

Variable selection methods	Precision (%)		Recall (%)		F_1 (%)		Accuracy (%)	Training time (min)
	L	H	L	H	L	H		
Empirical evidence	59	78.2	80.9	55	68.2	64.6	66.5	46.4
MFA	58.2	80.1	84.1	51.6	67.4	62.8	66.1	38
Garsons algorithm	61.8	73.8	71.5	64.6	66.3	68.9	67.7	47

Administration, 2006b; Blower and Green, 2010), cargo type, experience (Cantor et al., 2017), and union issues (Corsi et al., 2012) should be further explored given their potential implications on safety.

It is important to note that carrier safety is also more than just using crash rates. Carriers that do not have any crashes in the examined time period may not actually be the safest or even, generally, safe. The Crash Indicator computed by CSA only considers crash severity based on fatalities and injuries, which clearly have higher weights.

The BASIC measures within the CSA program are computed from the various violation types in MCMIS, and account for exposure using vehicle miles traveled and average power unit. The percentiles for the BASIC measures are updated with new carrier entries. However, BASICS does not distinguish carrier safety based on carrier types, traffic conditions, or even environmental differences due to geographic regions. It is also not a prediction model of carrier safety, and, therefore, accuracy and precision are not quantified.

Our models consider crash severity in the context of vehicle, carrier, and environmental conditions, and provides a more holistic perspective to examining carrier safety. The modeling approach described in this paper provides a way of classifying the crash rates into various levels, which is useful for comparing model performance and continually updating a carrier's rating based on new crash event occurrences.

Appendix A. Data element summary

Table A.11
Selected variables in CRASH_MASTER table.

Data element	Definition
REPORT_DATE	The date on which the crash occurred. (MM/DD/YY)
DOT_NUMBER	This is the number assigned by MCMIS to a census record. It is sometimes referred to as the USDOT number. Each motor carrier should have only one active census number. The census number has no internal coded structure. Numbers are issued sequentially as carriers are added to the system
TRUCK_BUS_IND	Indication of whether the vehicle involved in the crash was a truck (T) or bus (B)
TRAFFICWAY_ID	The degree of trafficway division at the place of the crash
ACCESS_CONTROL_ID	The degree that access to abutting land, light, air, or view in connection with highway is fully controlled by public authority
ROAD_SURFACE_CONDITION_ID	The condition that affected traction on the road surface at the time and location of the crash
CARGO_BODY_TYPE_ID	The cargo body type of the motor vehicle
WEATHER_CONDITION_ID	The predominant weather condition at the time and place of the crash
VEHICLE_CONFIGURATION_ID	The configuration of the motor vehicle
LIGHT_CONDITION_ID	Light condition at the time and place of the crash
FATALITIES	Number of persons killed inside or outside a vehicle at the scene of the crash
INJURIES	Number of persons injured inside or outside a vehicle at the scene of the crash who were transported to a medical facility for immediate medical attention

Table A.12
Table of contents of CARGO_BODY_TYPE_ID.

Variable names	CARGO_BODY_TYPE_ID
Contents	1 = Bus (seats for 9–15 people, including driver) 2 = Bus (seats more than 15 people, including driver) 3 = Van/Enclosed Box 4 = Cargo Tank 5 = Flatbed 6 = Dump 7 = Concrete Mixer 8 = Auto Transporter 9 = Garbage/Refuse 10 = Grain, chips, gravel 11 = Pole 12 = Not applicable 13 = Intermodal 14 = Logging 15 = Vehicle Towing another Vehicle 98 = Other

Table A.13
Table of contents of VEHICLE_CONFIGURATION_ID.

Variable names	VEHICLE_CONFIGURATION_ID
Contents	1 = Passenger Car (Only if Vehicle displays HM Placard) 2 = Light Truck (Only if Vehicle displays HM Placard) 3 = Bus (Seats for 9–15 People, Including Driver) 4 = Bus (Seats for > 15 People, Including Driver) 5 = Single-Unit Truck (2-Axle, 6-Tire) 6 = Single-Unit Truck (3 or More Axles) 7 = Truck/Trailer 8 = Truck Tractor (Bobtail) 9 = Tractor/Semitrailer 10 = Tractor/Double 11 = Tractor/Triple 99 = Unknown Heavy Truck > 10,000 lbs, Cannot Classify

References

Beck, M., 2016. NeuralNetTools: Visualization and Analysis Tools for Neural Networks. R package version 1.5.0.

Blincoe, L.J., Miller, T.R., Zaloshnja, E., Lawrence, B.A., 2015. The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised), (Report No. DOT HS 812 013). National Highway Traffic Safety Administration.

Blower, D., Green, P.E., 2009. Truck mechanical condition and crashes in the large truck crash causation study, (Report No. UMTRI-2009-09). University of Michigan, Ann Arbor, Transportation Research Institute.

Blower, D., Green, P., 2010. Type of motor carrier and driver history in fatal bus crashes. Transp. Res. Record: J. Transp. Res. Board 37–43.

Cantor, D.E., Corsi, T.M., Grimm, C.M., 2017. The impact of new entrants and the new entrant program on motor carrier safety performance. Transp. Res. E: Logist. Transp. Rev. 97, 217–227.

Castelletti, A., de Rigo, D., Rizzoli, A.E., Soncini-Sessa, R., Weber, E., 2005. A selective improvement technique for fastening neuro-dynamic programming in water resource network management. IFAC Proc. Vol. 38, 7–12.

Chen, G.X., 2008. Impact of federal compliance reviews of trucking companies in reducing highway truck crashes. Accid. Anal. Prevent. 40, 238–245.

Corsi, T.M., Fanara Jr., P., 1988a. Driver management policies and motor carrier safety. Logist. Transp. Rev. 24, 153.

Corsi, T., Fanara, P., 1988b. Effects of new entrants on motor carrier safety. Transportation Deregulation and Safety Conference.

Corsi, T.M., Fanara, P., Roberts, M.J., 1984. Linkages between motor carrier accidents and safety regulation. Logist. Transp. Rev. 20.

Corsi, T.M., Grimm, C.M., Cantor, D.E., Sienicki, D., 2012. Safety performance differences between unionized and non-union motor carriers. Transp. Res. E: Logist. Transp. Rev. 48, 807–816.

Da, Y., Xiurun, G., 2005. An improved PSO-based ANN with simulated annealing technique. Neurocomputing 63, 527–533.

Federal Motor Carrier Safety Administration, 2006a. Future Truck and Bus Safety Research Opportunities, vol. 38 Transportation Research Board.

Federal Motor Carrier Safety Administration, 2006b. Report to Congress on the Large Truck Crash Causation Study, (Report No. MC-R/MC-RRA). the National Technical Information Service, Springfield, Virginia 22161.

Federal Motor Carrier Safety Administration, 2017. SAFER Help. https://safer.fmcsa.dot.gov/saferhelp.aspx.

Federal Motor Carrier Safety Administration (FMCSA), 2018. Safety Measurement System (SMS) Methodology: Behavior Analysis and Safety Improvement Category (BASIC) Prioritization Status:version 3.09.

Garson, D.G., 1991. Interpreting neural network connection weights. AI Expert 6.

Goh, A., 1995. Back-propagation neural networks for modeling complex systems. Artif. Intell. Eng. 9, 143–151.

Knipling, R.R., 2009. Three large truck crash categories: what they tell us about crash causation. Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Kuhn, M., 2016. caret: Classification and Regression Training. R package version 6.0-73.

Lantz, B., Loftus, J., 2005. Development and implementation of a driver safety history indicator into the roadside inspection selection system. J. Saf. Res. 36, 489–490.

Lueck, M.D., 2012. Compliance, safety, accountability: Analyzing the relationship of scores to crash risk. American Tansportation Research Institute.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2015. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, R package version 1.6-7.

Mooren, L., Grzebieta, R., Williamson, A., Olivier, J., Friswell, R., 2014. Safety management for heavy vehicle transport: a review of the literature. Saf. Sci. 62, 79–89.

Moses, L.N., Savage, I., 1992. The effectiveness of motor carrier safety audits. Accid. Anal. Prevent. 24, 479–496.

Moses, L.N., Savage, I., 1994. The effect of firm characteristics on truck accidents. Accid. Anal. Prevent. 26, 173–179.

Murray, D.C., Lantz, B., Keppler, S.A., 2005. Predicting truck crash involvement: developing a commercial driver behavior-based model and recommended countermeasures. Transp. Res. Forum.

National Academies of Sciences, Engineering, and Medicine, 2017. Improving Motor Carrier Safety Measurement. National Academies Press.

Rogers, W.C., Knipling, R.R., 2007. Commercial driver human factors. Transportation Research Circular E-C117, 92–112.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., et al., 1988. Learning representations by back-propagating errors. Cogn. Model. 5, 1.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Springer.

John, A., Volpe National Transportation System Center (Volpe), 2013. Carrier Safety Measurement System (CSMS) Methodology: version 3.0.1. Federal Motor Carrier Safety Administration.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.