Assessment of Large Trucks Crash Severity on a Rural Interstate Road in Wyoming Using Decision Trees and Structural Equation Model

Sherif M. Gaweesh, Ph.D., P.E.[1]; Irfan U. Ahmed, Ph.D.[2]; and Mohamed M. Ahmed, Ph.D., P.E., M.ASCE[3]

Abstract: Promoting the safety of commercial trucks by identifying countermeasures that eliminate/reduce the effect of factors that increase the severity of truck-related crashes is crucial. Crash causal factors for rural interstate roads, located within the mountain plains, are inherently unique compared to urban interstate roads. This is due to the presence of challenging road geometry coupled with severe weather conditions and high truck traffic volumes. This study investigated Interstate 80 in Wyoming using decision trees, as a data mining approach, and structural equation model (SEM) as a latent factor modeling approach. SEM was employed to clarify the direct and indirect relationships between endogenous and exogenous variables while accounting for the variation and covariation within and between the constructed measurement models. Crash severity data were processed to account for factors affecting single vehicles and multivehicle trucks. The results showed that the interaction with surrounding traffic was the most significant latent variable affecting the crash severity of multivehicle truck crashes, while adverse weather conditions were the most significant latent variable affecting the crash severity of single-truck crashes. The results of this study highlighted the importance of increasing the situational awareness of commercial truck drivers with upcoming hazardous events. This could be performed by communicating information using variable message signs, the 511 application, the commercial vehicle operator portal (CVOP), or the connected vehicle (CV) technologies. **DOI: 10.1061/JTEPBS.TEENG-7446.**

Author keywords: Truck-related crashes; Crash severity analysis; Structural equation model (SEM); Exploratory factor analysis; Confirmatory factor analysis; Latent variables; Decision trees.

Introduction

Traffic safety and mobility of commercial trucks seized the attention of many researchers due to their high importance for freight movement, industrial development, and the national economy. Commercial trucks were responsible for transporting nearly 64% of the total US freight tonnage in 2015, representing 69% of the total freight value (Worth et al. 2016). An increase of 10% in large-truck-related fatal crashes for 2017 was observed compared to 2016 (i.e., 4,251 to 4,657 total crashes) (Federal Motor Carrier Safety Administration 2019). Moreover, increases of 5% and 3% were observed for truck-related injury and property damage-only crashes, respectively. Additionally, large-truck fatalities per 100 million vehicle miles traveled (MVMT) increased by 6% from 1.48 in 2016 to 1.56 fatalities per MVMT in 2017. Trucks have a substantial effect on the surrounding traffic as a result of their interfering effect (Moridpour et al. 2015). It is expected that the freight tonnage will be doubled in the coming 30 years (DOT 2020). This increase in truck traffic may pose traffic safety issues.

The National Highway Traffic Safety Administration (NHTSA) indicated that Wyoming had an increase of 3% in fatal crashes in 2017 compared to 2016 (*2018 Fatal Motor Vehicle Crashes: Overview* 2019). According to the Wyoming Department of Transportation (WYDOT), a 13% increase in truck-related crashes for the year 2018 was reported (Wyoming DOT 2018). Interstate 80 (I-80) is considered one of the main freight corridors in Wyoming. It has a total length of 402 miles and is located in southern Wyoming. It connects the east and west borders of Wyoming. The corridor is characterized by adverse weather conditions, challenging roadway geometry, and high truck traffic. The percentage of truck traffic on I-80 reaches 55% of the total annual average daily traffic (AADT). A steady increase in truck-related crashes was observed on I-80 in Wyoming. The crash rate per MVMT in 2012 was 0.67, which increased to 0.76 in 2014 and spiked to 1.02 crashes per MVMT in 2016. In 2018, the crash rate for truck-related crashes on Wyoming's I-80 increased to become 1.04 crashes per MVMT.

Several studies were conducted on I-80 with a focus on promoting the safety and mobility of commercial trucks, emphasizing the need to thoroughly investigate the factors affecting truck crashes in Wyoming (Ahmed et al. 2019, 2020; Gaweesh et al. 2021a, b, 2019; Gaweesh and Ahmed 2020; Irfan Ahmed et al. 2020; Khoda Bakhshi et al. 2021; Raddaoui et al. 2020).

Several statistical techniques have been used to investigate the severity of historical crashes. Observed variables are mainly

[1]Assistant Professor, Dept. of Civil Engineering, Univ. of North Dakota (UND), Grand Forks, ND 58201; formerly, Road Safety Professional, Assistant Professor of Research, Research Center for Energy Systems Research (CESR), Tennessee Tech Univ., Cookeville, TN 38505 (corresponding author). ORCID: https://orcid.org/0000-0001-7977-6378. Email: sherif.gaweesh@und.edu

[2]Traffic Engineer, Henningson, Durham & Richardson (HRD) Engineering Inc., 10450 Holmes Rd., Kansas City, MO 64131. ORCID: https://orcid.org/0000-0003-2315-6262. Email: irfan.ahmed@hdrinc.com

[3]Professor and Director, Dept. of Civil and Architectural Engineering and Construction Management Transportation Center, Univ. of Cincinnati, Cincinnati, OH 45221. Email: ahmed2mh@ucmail.uc.edu

Note. This manuscript was submitted on March 28, 2022; approved on May 12, 2023; published online on October 20, 2023. Discussion period open until March 20, 2024; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering, Part A: Systems***, © ASCE, ISSN 2473-2907.

J. Transp. Eng., Part A: Systems, 2024, 150(1): 05023008

analyzed to estimate the direct relationship between the indicator variables and the dependent variable. Moreover, the discrete model analysis was utilized to conduct the crash severity analysis and is considered the most dominating approach. Despite the selection of the model type and its underlying assumptions, the complex interrelationship between crash variables cannot be observed (Lee et al. 2008). To overcome this limitation alternative statistical approaches could be used to provide further insights about the investigated crash data. Latent variable analysis as well as data mining approaches could be potential alternatives to investigate the severity of crashes. For instance, a structural equation model (SEM) is a technique in which indicator variables are factored together to measure unobserved latent variables. It can resolve the complex relationship between the indicator variables (Wang and Qin 2014). Additionally, it assesses the effectiveness of the indicator variable to measure latent constructs. Data mining techniques are superior in predicting future crash trends and severities compared to traditional statistical approaches. They have the ability to analyze large and complex datasets, while explaining the relationships between the indicators and the response variable.

Crashes on rural roads are generally more severe compared to urban road crashes due to differences in operating speeds, road geometry, functionality, and enforcement levels. Adding the complication of encountering severe weather events as well as high truck traffic volumes will raise the need to conduct several studies to alleviate the safety of such rural corridors. However, several studies focused on crash severity in rural interstates (Cafiso et al. 2010; Chen et al. 2016; de Oña et al. 2013; Siskind et al. 2011), but none of the investigated corridors had the unique characteristics of the I-80 corridor in Wyoming. Additionally, most of the utilized statistical approaches accounted for observed variables only, which might not clarify the indirect effects that might increase the crash severities of large trucks.

This study utilized decision trees as a data mining technique as well as the SEM to identify the variables that affect the crash severity of large trucks. It developed insights into the most effective indicator variables that affect the severity of truck crashes. In addition, observed variables that might be used to measure the latent variables affecting crash severity were identified. Unlike other studies that utilized one observed variable to indicate the crash severity, this research introduced a new measure to quantify the severity of truck crashes. Direct and indirect effects on large-truck crash severity were identified from the developed SEMs. Furthermore, not only hypothetical assumptions were used to develop the measurement model but also exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to provide a systematic approach to identify latent variables to develop the path model. The research considered both; crash and vehicle level in the analysis, in which commercial trucks were the considered vehicle type in the analysis. Additionally, the difference between the factors affecting single-truck crashes and multivehicle truck crashes was accounted for by developing two separate models for the two vehicle-level analyses.

The advantage of utilizing the SEM is its ability to interpret causal relationships involving factors that cannot be directly measured or observed. These unobserved variables are called latent variables. The adoption of SEM is needed to investigate factors associated with the crash severity of large trucks as not all crash factors can be directly accounted for despite having numerous variables and data indicators.

Background

Many studies were conducted with a focus on determining the causal factors and predicting the severity of truck-related crashes. A broad variety of statistical approaches were employed to identify factors that might increase truck crashes. Parametric, nonparametric, Bayesian, and latent class analysis were the main statistical approaches adopted to conducting the analysis. The mentioned statistical approaches analyze the crash data differently, clarifying the factors that might increase the outcome severity of the investigated type of crash from multiple perspectives.

Parametric Approaches

Truck crashes are usually investigated by separating the data into two main categories, single and multivehicle truck crashes, as the factors affecting the outcome crash injury severity were different (Chen and Chen 2011, 2013; Geedipally and Lord 2010; Islam et al. 2014; Lemp et al. 2011; Naik et al. 2016).

A study included driver behavior factors along with variables extracted from crash reports utilizing an ordered probit model (Zhu and Srinivasan 2011). Among the several significant variables, dummy variables that indicated missing data showed a strong significance toward increasing crash injury severity, which could be due to the gaps in crash reporting or due to the unobserved heterogeneity in the data. However, the study concluded a limitation of the small sample size used in the analysis. Another study provided a comparative analysis of factors affecting the crash severity of large trucks using the ordered logit model (Taylor et al. 2018). The study showed that variables related to roadway, crash, vehicle, and driver had a significant effect on the investigated crash severity. Among the significant variables, the season, manner of collision, lighting conditions, driving under the influence, and percentage of truck traffic were significant. Uddin and Huynh (2017) used a random parameters logit model to investigate the crash severity of crashes involving a truck (Uddin and Huynh 2017). The study concluded that adverse weather and challenging roadway geometry were significant variables in increasing the severity of crashes involving a truck. The results of the study were in accordance with Naik et al. (2016). The latter study utilized random parameters ordinal and multinomial regression models.

Nonparametric Approaches

Data mining approaches were also adopted to investigate the crash severity of commercial trucks. A recent study used gradient boosting to evaluate the truck crash injury severity (Zheng et al. 2018). The analyzed crash data were obtained from two states, North Dakota and Colorado. The study accounted for the scale of the trucking company as a predicting variable. The results showed that small-scale companies had the lowest probability of crash risk.

Another study accounted for the spatial concentrations of large-truck crashes in which granular level land use and urban design factors were considered in the analysis (Tahfim and Yan 2022). The study utilized the density-based spatial clustering of application with noise (DBSCAN) to explore the effect of housing, population, employment, and road network density attributes along with the crash characteristics on increasing the severity of truck-related crashes. The study showed that high road network density and medium and high population density were associated with nonsevere injuries.

Bayesian Approaches

Additionally, Bayesian logistic models were adopted to conduct crash severity analysis (Ahmed et al. 2018). The study included a factor expressing the presence of large trucks in the model and utilized interactions to account for factors affecting the severity of truck crashes. The results of the study showed that adverse weather and steep downgrades would increase the severity of truck-related crashes. Another study utilized the hierarchical Bayesian models to investigate the between-crash variance and within-crash correlations for large trucks on rural corridors (Chen et al. 2016). The study clarified that road curve, vehicle damage in a crash, number of vehicles in a crash, wet road surface, vehicle type, driver age, driver gender, driver, seatbelt use, and driver alcohol or drug involvement are among the significant variables that increase the severity of truck crashes. Missing data and interactive effects between variables were among the limitations mentioned in the study.

Latent Class Models

Latent variables, which are unobserved variables, could help in clarifying the complex interrelationship between the crash indicator variables. The SEM could be considered a promising statistical approach that accounts for these interrelationships. It could quantify latent variables that cannot be directly measured or observed. SEM was previously used to analyze survey data and to assess driver behavior questionnaires (Ambak et al. 2010; Farag et al. 2007; Hamdar et al. 2008; Hassan and Abdel-Aty 2013; Shaaban et al. 2018, 2020). Recently, SEM was adopted to investigate the resultant severity of crashes (Barman and Bandyopadhyaya 2020; Cho et al. 2017; Dong et al. 2022; Wang and Qin 2014). The study concluded that injury severity and vehicle damage could be used as indicator variables to measure crash severity. Three SEM models with one, two, and three latent variables were developed. The results showed that the SEM with the two latent variables provided the best model fit. Another study developed an SEM to estimate truck-crashes severity (Cho et al. 2017). The developed SEM was factored into five latent variables named crash, environment, road, driver, and severity. However, the development of the measurement model was based on hypothetical assumptions, and it provided reliable results. The results showed that crash severity could be measured using the number of deaths, the number of injured, and the number of cars involved in the crash. Kim et al. (2011) examined the effect of accessibility on crash severity (Kim et al. 2011). It was found that accessibility had a reverse effect on crash severity. Increased accessibility would reduce the crash severity. A recent study conducted a comprehensive analysis framework utilizing SEM as well as random parameters model to investigate the severity of crashes involving large trucks on mountainous interstate roads (Gaweesh et al. 2023). The study identified direct and indirect effects on the crash severity of large trucks in which challenging roadway conditions were found to have an indirect effect on crash severity for single-truck crashes. Khattak and Targa (2004) utilized an ordinary least-square regression (OLS) to examine the risk factor affecting the large-truck-related crashes (Khattak and Targa 2004). The authors found that dangerous truck-driving behaviors, speeding, and reckless driving would increase the probability of truck rollover. The OLS is a similar statistical technique to the SEM. However, the SEM is considered a superior statistical model due to the better model fit, higher precision, and accuracy. The measures used to assess the model fitness and precision/accuracy were Chi-Square/df and root mean square error of approximation (RMSEA) in which values from the SEM were lower than the values obtained from the OLS (Nazim and Ahmad 2013).

Data Collection

Truck-related crash data on Interstate 80 (I-80) in Wyoming was extracted from the crash reports preserved by the WYDOT. Truck-related crashes that occurred on I-80 for the years 2009 to 2016 were utilized in this study. Truck-related crashes refer to crashes that involved at least one truck in the crash. Crash data were processed and subdivided into two datasets for further investigation. The two subdivisions were (1) single-truck crashes where only one truck was involved in a crash; and (2) multivehicle truck crashes, where more than one vehicle was involved in the crash including at least one truck. Crashes with missing outcome severity were eliminated from the utilized dataset and represented 2.3% of the dataset. For missing values for indicator variables, data imputation using mean substitution was adopted. Mean substitution has the advantage of retaining the sample mean for that variable.

The total number of single-truck crashes used in this study was 2044 crashes, while 1968 multivehicle truck crashes were included in the analyses. In addition to crash data, roadway geometry data, extracted from the Wyoming Roadway Data Portal (WRDP), were linked to the crash data. This dataset includes information about the roadway geometry, pavement type, number of lanes, median characteristics, and countermeasures information. Traffic data were also extracted from the monthly traffic data reports published by the WYDOT. It should be clarified that the traffic data used in this study were the hourly traffic volumes. Unlike aggregate traffic data, the hourly traffic volume data express the actual traffic encountered while the crash occurred. Although an extensive effort was conducted to collect the hourly volume of crashes from 2009 to 2016, it provided a disaggregate level of traffic interaction, providing more reliable results compared to aggregate traffic data.

Fig. 1 shows a heat map for the truck-related crashes that occur on I-80. The upper section of Fig. 1 shows the distribution of crashes along I-80 for single-truck crashes and the lower part shows the multivehicle truck crashes. The provided heatmap is weighted using the crash severity. It could be observed from the developed heatmap that the concentration of single-truck crashes was located in the Elk Mountain section between milepost (MP) 235 and MP 290. Additionally, the concentration of multivehicle truck crashes was located at the Laramie–Cheyenne section, MP 316 to MP 335, and the Green River section, MP 85 to MP 110, as well as the Elk Mountain section. These sections are characterized by challenging roadway geometry and harsh weather conditions compared to the other sections of the corridor.

Table 1 shows the descriptive statistics of the collected datasets used in this study. Data were categorized into several categories to easily explore the indicator variables. The first category included indicator variables related to crash injury severity. These variables included the manner of collision and the number of injured personnel in the crash as well as the number of fatalities. The roadway factors category included several indicator variables that express the roadway geometry, pavement type, and cross section elements. The temporal category included the traffic volumes and the season in which the crash occurred. The season variable was considered in the analysis as it accounts for the seasonal variation in crash frequencies. Each year was divided into two seasons: summer and winter, in which each crash was assigned to the summer or winter according to the date it occurred. Crashes that occurred from April 15th to October 15th were considered summer season crashes, while other crashes were considered winter crashes. Lighting conditions, roadway surface conditions, and weather reported in the crash reports were the indicator variables for the environment data category. The truck type was included as one of the indicator variables for the crash characteristics category. It was categorized as a binary variable, heavy truck, and other. A heavy truck is identified as a heavy vehicle with a gross vehicle weight rating (GVWR) greater than 26,000 lbs. The medium and light trucks were combined into the other level of the truck type variable. Driver and roadway treatments were the last two categories in the dataset. The driver category indicates the truck driver characteristics,

J. Transp. Eng., Part A: Systems, 2024, 150(1): 05023008

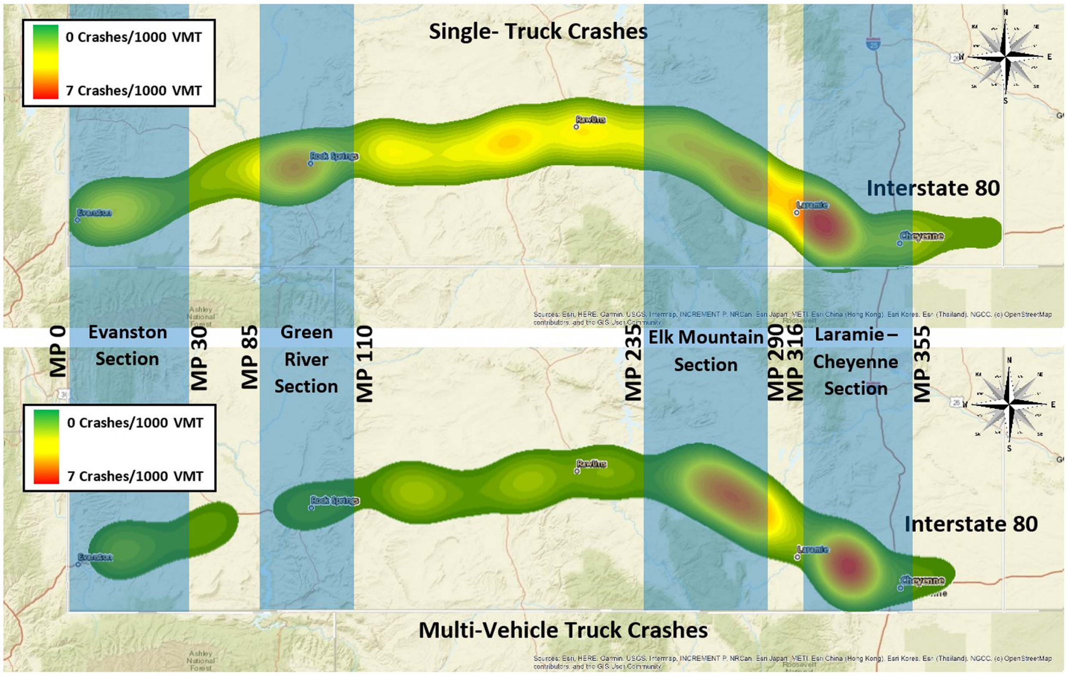

Fig. 1. Crash frequency heatmap for single and multivehicle truck crashes on I-80. [Sources: Esri, HERE, Garmin, USGS Intermap INCREMENT P. NRCan, Esri Japan, METI Esri China (Hong Kong), Esri (Thailand), NGCC, © OpenStreetMap contributors and the GIS User Community.]

Table 1. Descriptive statistics of the investigated indicator variables

		Description		MV truck (1,968 crashes)				SV truck (2,044 crashes)			
Category	Name	Variable	Categories	Avg. (%)	SD	Min.	Max.	Avg. (%)	SD	Min.	Max.
Crash severity	Fatal	Outcome severity	253 crashes	7.72	—	—	—	4.94	—	—	—
	Injury		586 crashes	19.87	—	—	—	9.54	—	—	—
	PDO		3,173 crashes	72.41	—	—	—	85.52	—	—	—
	Total		4,012 crashes	49.05	—	—	—	50.95	—	—	—
	NInj	Number of injured (integer)		0.181	0.478	0	4	0.403	0.834	0	10
	Nkill	Number of killed (integer)		0.003	0.058	0	1	0.023	0.190	0	5
Crash type	ManCol	Manner of collision (categorical)	1: Front-to-Rear	40.70	—	—	—	—	—	—	—
			2: Front-to-Front	1.52	—	—	—	—	—	—	—
			3: Angle	18.65	—	—	—	—	—	—	—
			4: Sideswipe	31.15	—	—	—	—	—	—	—
			5: Other	7.98	—	—	—	100	—	—	—
Roadway factors	Grade	Road grade-%- (Cont.)		0.06	2.02	−6.21	5.49	−0.11	1.87	−6.59	5.49
	Delta	Deflection angle-degrees- (Cont.)		9.32	17.29	0	88.47	9.60	17.00	0	88.47
	MedWid	Median width-feet- (integer)		101	110	18	956	127	145	18	956
	MedTyp	Median type (nominal)	0: Depressed	68.26	—	—	—	72.78	—	—	—
			1: Raised	31.74	—	—	—	27.22	—	—	—
	Nlanes	Number of lanes (integer)		2.095	0.312	2	5	2.079	0.275	2	5
	SHTYP	Shoulder pavement type (binary)	0: Asphalt	72.96	—	—	—	78.67	—	—	—
			1: Concrete	27.04	—	—	—	21.33	—	—	—
	LnTyp	Lanes pavement type (nominal)	0: Asphalt	48.43	—	—	—	53.52	—	—	—
			1: Concrete	51.57	—	—	—	46.48	—	—	—
Temporal	Season	Season of the year (binary)	0: Summer	20.69	—	—	—	34.15	—	—	—
			1: Winter	79.31	—	—	—	65.85	—	—	—
	HrVol	Hourly volume-VPH- (Cont.)		519.7	556.6	22.9	4,518	494.4	588.1	27.4	4,391
	TVol	Truck hourly volume-VPH-(Cont.)		243.1	272.2	10.4	2,426	225.7	272.4	12.1	2,148
Environ.	Lcond	Lighting condition	0: Daytime	71.72	—	—	—	56.73	—	—	—
			1: Nighttime	28.28	—	—	—	43.27	—	—	—
	Rcond	Road surface condition (nominal)	0: Dry	39.00	—	—	—	25.42	—	—	—
			1: Adverse	61.00	—	—	—	74.58	—	—	—
	Rweath	Reported weather (binary)	0: Clear	52.03	—	—	—	37.36	—	—	—
			1: Adverse	47.97	—	—	—	62.64	—	—	—

J. Transp. Eng., Part A: Systems, 2024, 150(1): 05023008

Table 1. (*Continued.*)

Category	Name	Description		MV truck (1,968 crashes)				SV truck (2,044 crashes)			
		Variable	Categories	Avg. (%)	SD	Min.	Max.	Avg. (%)	SD	Min.	Max.
Crash Char.	TrTyp	Truck type (Ordinal)	0: Other	3.19	—	—	—	9.45	—	—	—
			1: Heavy	96.81	—	—	—	90.55	—	—	—
	VehN	# of crashed vehicles (integer)		2.12	0.75	2.00	5.00	1.00	0.00	1.00	1.00
	Grdrail	Guardrail	0: Absent	92.12	—	—	—	96.70	—	—	—
			1: Presented	7.88	—	—	—	3.30	—	—	—
	Speeding	Above speed limit	0: No	92.68	—	—	—	93.47	—	—	—
		+5 (binary)	1: Yes	7.32	—	—	—	6.53	—	—	—
	Wild “FHE”	Wildlife crash	0: No	99.88	—	—	—	99.41	—	—	—
		(binary)	1: Yes	0.12	—	—	—	0.59	—	—	—
	FxdObj “FHE”	Hitting a fixed object	0: No	93.79	—	—	—	82.52	—	—	—
		(binary)	1: Yes	6.21	—	—	—	17.48	—	—	—
	WrkZn “FHE”	Work zone crash	0: No	99.68	—	—	—	97.63	—	—	—
		(binary)	1: Yes	0.32	—	—	—	2.37	—	—	—
Driver	DUI	Driving under the	0: No	98.49	—	—	—	99.70	—	—	—
		influence (binary)	1: Yes	1.51	—	—	—	0.30	—	—	—
	Gender	Gender (binary)	0: Male	84.36	—	—	—	93.88	—	—	—
			1: Female	15.64	—	—	—	6.11	—	—	—
	Age	Driver age (Cont.)		44.22	17.94	16	90	43.27	12.08	17	78
Roadway treatments	RS	Rumble strips	0: No	69.11	—	—	—	58.21	—	—	—
		(binary)	1: Yes	30.89	—	—	—	41.79	—	—	—
	SF	Snow fence (binary)	0: No	74.43	—	—	—	69.19	—	—	—
			1: Yes	25.57	—	—	—	30.81	—	—	—
	VSL	Variable speed limit	0: No	52.61	—	—	—	51.55	—	—	—
		(binary)	1: Yes	47.39	—	—	—	48.45	—	—	—
	CL	Climbing lanes	0: No	96.46	—	—	—	98.16	—	—	—
		(binary)	1: Yes	3.54	—	—	—	1.84	—	—	—

Note: MV = multitruck vehicle crashes; SV = single-vehicle truck crash; Ave = average; SD = standard deviation; Min. = minimum; Max. = maximum; Cont. = continuous; Char. = characteristics; ManCol = manner of collision; NInj = number of injured; NKill = number of killed; MedWid = median width; MedTyp = median type; Nlanes = number of lanes; SHTYP = shoulder type; LnTyp = lane pavement type; HrVol = hourly volume; TVol = truck volume; Lcond = lighting condition; Rcond = road surface condition; Rweath = reported weather; TrTyp = truck type; VehN = number of crashed vehicles; Grdrail = presence of guardrail; FHE = first harmful event; WrkZn = work zone; DUI = driving under influence; RS = rumble strips; SF = snow fence; VSL = variable speed limit; and CL = climbing lanes.

while the roadway treatments indicate the existing countermeasures implemented at the crash location. Table 1 provides the percentage of each level for the categorical indicator variables, the mean, standard deviation, minimum, and maximum for the continuous and integer indicator variables.

Methodology and Data Analysis

Decision Trees

A decision tree (DT) is a type of data mining method that has been extensively utilized in traffic safety classification and regression analysis (Song and Lu 2015). The analysis results in a status of the class variable (response) that shows the highest number of cases in the leaf node analyzed. Therefore, the number of rules can be identified with the number of terminal nodes in the tree. This DT is considered a classification analysis as the response variable is nominal. Several researchers have utilized classification trees in their safety studies, for instance, classification and regression tree (CART) (Abellán et al. 2013; Chang and Chien 2013; Montella et al. 2012) and Chi-square automatic interaction detection (CHAID) (Badea-Romero and Lenard 2013; Mohamadi Hezaveh et al. 2018; Prati et al. 2017). CHAID DT technique utilizes the $\chi 2$-test of association to construct the tree by repeatedly splitting subsets of the space into two or more child nodes beginning with the entire set. The best split at any node is determined by merging any allowable pair of categories of the predictor variables until there is no statistically significant difference within the pair with respect to the target variable (Ture et al. 2009). In this study, the CHAID classification task is undertaken to investigate single-vehicle and multivehicle truck crashes separately.

Exploratory Factor Analysis

EFA was mainly used in this study to prespecify and develop the exogenous and endogenous latent variables for the CFA. Application of EFA and CFA on the same dataset should be avoided, as it might hinder the external validity of the obtained factors (Hurley et al. 1997). Accordingly, the EFA was conducted on 20% randomly selected crash data (409 observations for the single-truck crash data and 394 observations for the multivehicle truck crashes) to provide a priori hypothesized patterns. A minimum of three indicator variables were selected to measure a single latent variable to avoid convergence issues. The extraction method used to conduct the EFA was the generalized least squares (GLS) method, with a Varimax orthogonal rotation method. A cutoff value of 0.4 was used for the factor loading values (Hatcher and O’Rourke 2013). The obtained Kaiser–Meyer–Olkin value (KMO) was found to be 0.695 and 0.764 for the single and multivehicle truck crash analysis, respectively. The latent variables are considered well-factored if the KMO value is above 0.5 (Shaaban et al. 2018, 2020). Table 2 shows the obtained factor loading for the EFA.

Confirmatory Factor Analysis

The CFA is distinguished from the EFA as it provides a more parsimonious solution, which generates error variances resulting from

Table 2. EFA obtained latent variables from the EFA analysis

	Factor number and loading value							
	Single-truck crashes				Multivehicle truck crashes			
Indicator variable	Factor#1	Factor#2	Factor#3	Factor#4	Factor#1	Factor#2	Factor#3	Factor#4
Number of vehicles	—	—	—	—	0.526	—	—	—
Truck type	—	—	—	—	0.400	—	—	—
Manner of collision	—	—	—	—	0.463	—	—	—
DUI	0.431	—	—	—	0.458	—	—	—
Number of injured	0.751	—	—	—	—	—	—	—
Number of killed	0.447	—	—	—	—	—	—	—
Season		0.710	—	—	—	0.657	—	—
Road condition	—	0.824	—	—	—	0.847	—	—
Reported weather	—	—	—	—	—	0.597	—	—
Speeding		0.667	—	—	—	—	—	—
Lighting condition	—	—	−0.803	—	—	—	−0.782	—
Hourly volume	—	—	0.921	—	—	—	0.879	
Truck hourly volume	—	—	0.943	—	—	—	0.913	—
Lanes surface type	—	—	—	0.775	—	—	—	0.795
Median width	—	—	—	−0.678	—	—	—	−0.727
Median type	—	—	—	0.405	—	—	—	0.577
Shoulder surface type	—	—	—	0.786	—	—	—	0.844

Table 3. Item reliability index for the investigated indicator variables

	Obtained factors and item reliability index (R^2)							
	Single-truck crashes				Multivehicle truck crashes			
Indicator variable	Crash severity	Adverse driving conditions	Interaction with traffic	Roadway factors	Crash severity	Adverse driving conditions	Interaction with traffic	Roadway factors
Number of vehicles	—	—	—	—	0.570	—	—	—
Manner of collision	—	—	—	—	0.630	—	—	—
Number of vehicles	—	—	—	—	0.456	—	—	—
DUI	0.418	—	—	—	—	—	—	—
Number of injured	0.451	—	—	—	—	—	—	—
Number of killed	0.525	—	—	—	—	—	—	—
Season	—	0.524	—	—	—	0.567	—	—
Road condition	—	0.815	—	—	—	0.399	—	—
Reported weather	—	0.424	—	—	—	0.567	—	—
Speeding	—	0.474	—	—	—	0.692	—	—
Lighting condition	—	—	0.395	—	—	—	0.397	—
Hourly volume	—	—	0.886	—	—	—	0.879	—
Truck hourly volume	—	—	0.984	—	—	—	0.936	—
Lanes surface type	—	—	—	0.475	—	—	—	0.396
Median width	—	—	—	0.410	—	—	—	0.886
Median type	—	—	—	0.529	—	—	—	0.503
Shoulder surface type	—	—	—	0.737	—	—	—	0.656

the conditional relationship among the indicator variables (Brown 2015). CFA was used to determine the measurement model used to conduct the path analysis of the SEM. The absence of multicollinearity was checked by developing a multicollinearity matrix where the r-squared was less than 0.8 (Brown 2015; Hatcher and O'Rourke 2013). The selection of the indicator variables should be supported by theory and the results of prior research evidence. Hence, the selected variables were based on the EFA, results of previous studies (Cho et al. 2017; Khattak and Targa 2004; Wang and Qin 2014), and engineering judgment. The obtained results from the CFA were nearly similar to the EFA, which was expected. The reliability of the indicator variables was assessed based on their contribution to the measurement model. The percent of variance for each indicator that was scrutinized by its latent variable is the index of the item reliability, which is referred to as R^2. The indicator variables with an item reliability index greater than 0.39 are considered ideal (Hatcher and O'Rourke 2013). Table 3 shows the obtained values for the item reliability indices (R^2), latent variables, and factored indicator variables.

Structural Equation Model

The SEM is mainly used as a statistical technique to analyze survey questionnaire datasets (Brown 2015; Hassan and Abdel-Aty 2013; Hurley et al. 1997; Shaaban et al. 2018). Recently, few studies utilized this statistical technique to analyze crash data and to perform real-time risk assessment (Cho et al. 2017; Kim et al. 2011; Ramos 2014; Wang and Qin 2014; Xu et al. 2018). SEM has several advantages as it can handle complex relationships (indirect, multiple, and reverse relationships) between exogenous and endogenous variables (Kervick et al. 2015). It quantifies unmeasurable variables by developing latent variables using the observed variables. Moreover, it simultaneously estimates the path coefficients of the relationships between the latent variables in the context of a full model.

Model Development

SEM could be defined as a multivariate statistical analysis that analyzes structural relationships between the measurement model and latent constructs in which endogenous and exogenous variables are analyzed (Kaplan 2000). The SEM assumes a directional relationship between the developed latent variables resulting from the CFA. The SEM consists of two major components; (1) the measurement model, which specifies the significant variables that can measure each constructed latent variable with an exogenous model (x-measurement model) and an endogenous model (y-measurement model), and (2) the structural model, which specifies the significant direction of prediction between the exogenous model and the endogenous model.

The measurement model is basically developed in the CFA phase. In the structural model, simultaneous equations are formed by linking the exogenous and endogenous variables (Hassan 2011). To avoid identification and convergence issues, a minimum number of three indicator variables should measure each developed latent variable (Byrne 1998). Additionally, it is advised to use a maximum of 30 indicator variables to obtain a converged model and to evade model fitting issues (Hatcher and O'Rourke 2013).

It is worth mentioning that the estimation method used to develop the SEM was the diagonal weighted least square (DWLS) method. DWLS method does not have a specific distributional assumption toward the data (Li 2016; Rhemtulla et al. 2012). Among the other estimation methods, DWLS was designed to deal specifically with ordinal data. Several studies investigated the validity of using the DWLS estimation method with ordinal data and found that it led to unbiased results (Beauducel and Herzberg 2006; Forero et al. 2009; Lei 2009; Yang-Wallentin et al. 2010).

The sample size is one of the important factors to develop SEM, as it is based on the large sample theory (Hatcher and O'Rourke 2013). Various studies had asserted the required minimum sample size to conduct a SEM. A study showed that a minimum of 300 observations is required to develop the SEM (Hatcher and O'Rourke 2013). However, other studies showed that a minimum sample size of 200 would be adequate to meet the assumptions of the large sample theory (Hatcher and O'Rourke 2013). Another study showed that a ratio of 10:1 for the number of observations to the number of investigated indicator variables should be achieved to obtain an adequate sample size (Suhr 2006). To measure the adequacy of the sample size used in the analysis, it was suggested to assess the statistical power of the developed SEM (Hatcher and O'Rourke 2013). To determine the statistical power of the model, the confidence intervals surrounding the RMSEA should be evaluated as well as the RMSEA value. An RMSEA value less than or equal to 0.08 suggests adequate statistical power. Fig. 2 shows the structure map and the different elements of the SEM. The measurement models could be expressed as shown in Eq. (1), and the structural model is given in Eq. (2) (Lee et al. 2008; Wang and Qin 2014)

$$\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} \lambda_y & 0 \\ 0 & \lambda_x \end{bmatrix} \begin{bmatrix} \eta \\ \xi \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \delta \end{bmatrix} \tag{1}$$

$$\eta = \beta\eta + \Gamma\xi + \zeta \tag{2}$$

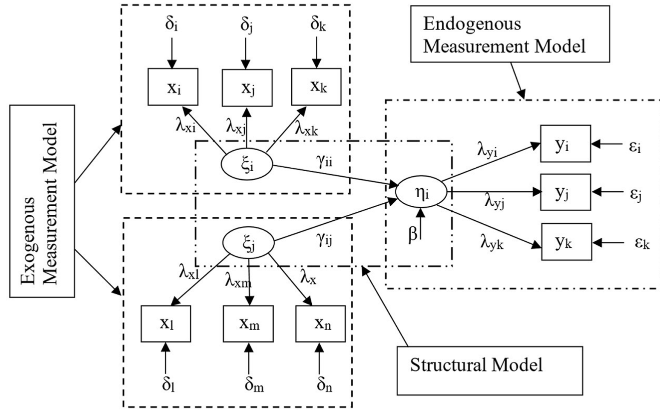

Fig. 2. Structural map and the elements of the SEM.

Several trials were attempted in developing the models where variables were used in several forms and transformations. Initially, the utilized variable used was in their original recorded format without clustering. Additionally, clustering was tested with important variables reviled from the results obtained from the CART model. While SEM was initially derived to consider only continuous and discrete variables, several studies have applied SEM with categorical variables, especially with survey analysis. Categorical variables could be utilized in the SEM as exogenous or endogenous variables. Having a binary, nominal, or ordinal categorical variable might cause a problem as there cannot be a single path arrow for each category with a specific coefficient. Accordingly, this behavior presents a challenge for parameterizing path diagrams. To overcome this limitation, for binary variables, the values were coded as 0 and 1, where the model was set as numeric. For other nominal and ordinal variables, dummy variables were created for each level in which the model treats them as a set of binary variables. In this research, the approach of converting nominal and ordinal variables was feasible as the maximum number of levels in most of the investigated variables was two levels, except for only one variable that had five levels. With other categorical variables that include more than five levels, other alternative approaches such as the Lavaan procedure could be utilized (Rosseel 2014).

Classification Tree Results

Crash severity was selected as the dependent variable for the developed models, using two severity levels of fatal and injury (F+I) representing and property damage only (PDO) crashes of Fig. 3 represent the CHAID model for the single-vehicle truck crashes, while Fig. 4 is for the multivehicle truck crashes. The CHAID resulting tree for the single-vehicle truck crashes, in Fig. 3, had a depth level of five with 19 nodes. The model selected the variables of road surface conditions, gender, guardrails, rumble strips, reported weather conditions, season, truck hourly volume, median type, and visibility level to classify the single-vehicle truck crashes. Road surface conditions was selected as the first variable to affect the severity of the single vehicle truck crashes, in which most of the crashes occurs in wet surface conditions (a total of 72.7%). When road surface is dry, the absence of guardrails at clear weather conditions is involved in 19.7% of the single-vehicle truck crashes. The multivehicle truck crashes model had 24 nodes with a tree depth of 4 levels as presented in Fig. 4. Collision type, age, speeding, pavement surface type, road surface condition, traffic volumes, lighting conditions, and day of week were the variables considered in the

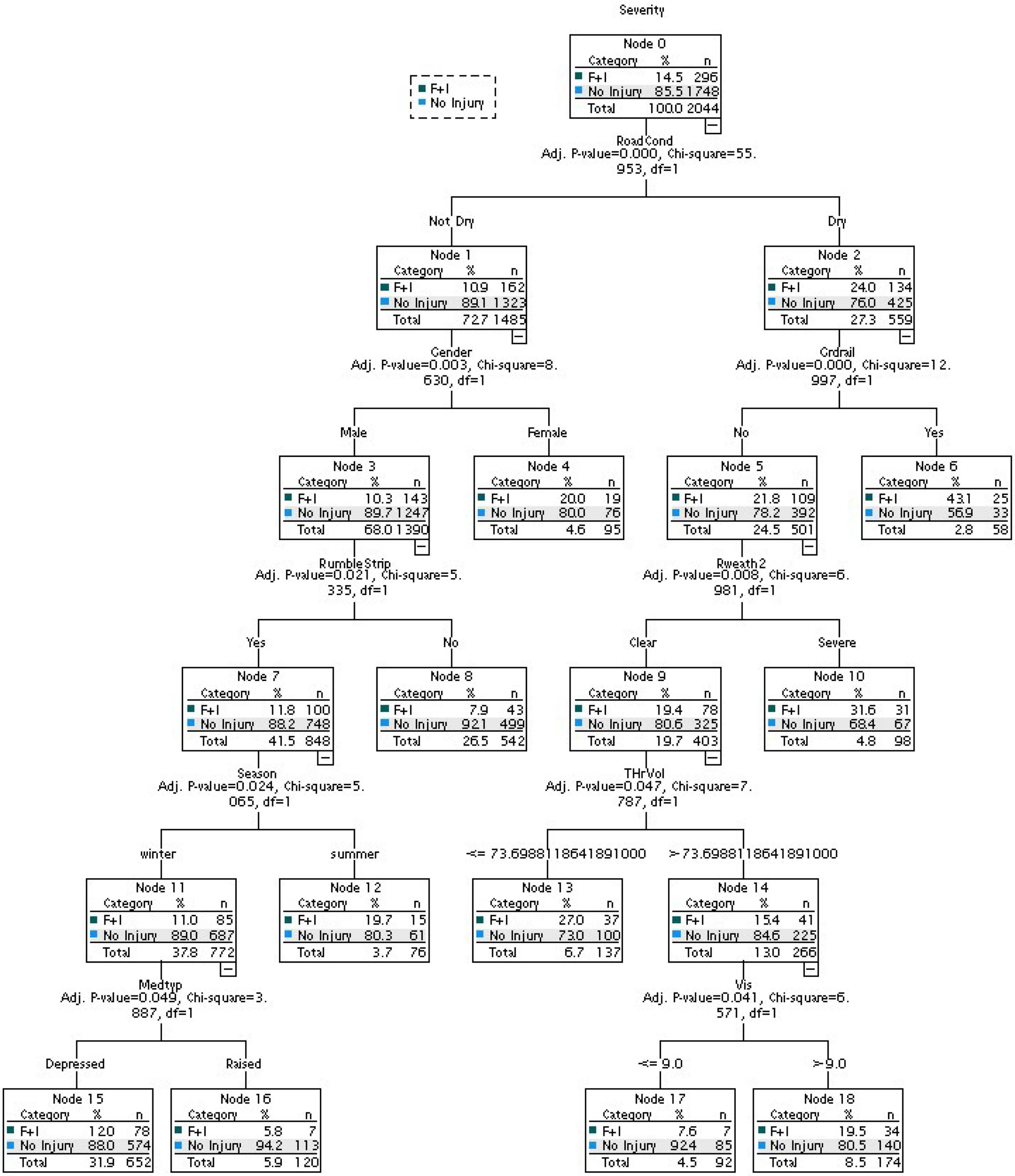

Fig. 3. Single-vehicle truck crashes classification tree.

developed model. Fig. 4 shows that collision type was selected as the most important variable affecting the multivehicle truck crashes, resulting into three initial nodes. It is worth mentioning that different variables describe each crash type selected by the CHAID model. Most angle crashes involved older age groups (17.3%) were only 5.4% involved young age groups, where the identified age threshold was 27 years old. Road surface condition was the factor affecting the severity of front crashes, in which 27% of them occurred in slippery surface conditions, and 16.1% in dry conditions. On the other hand, speeding was the variable affecting

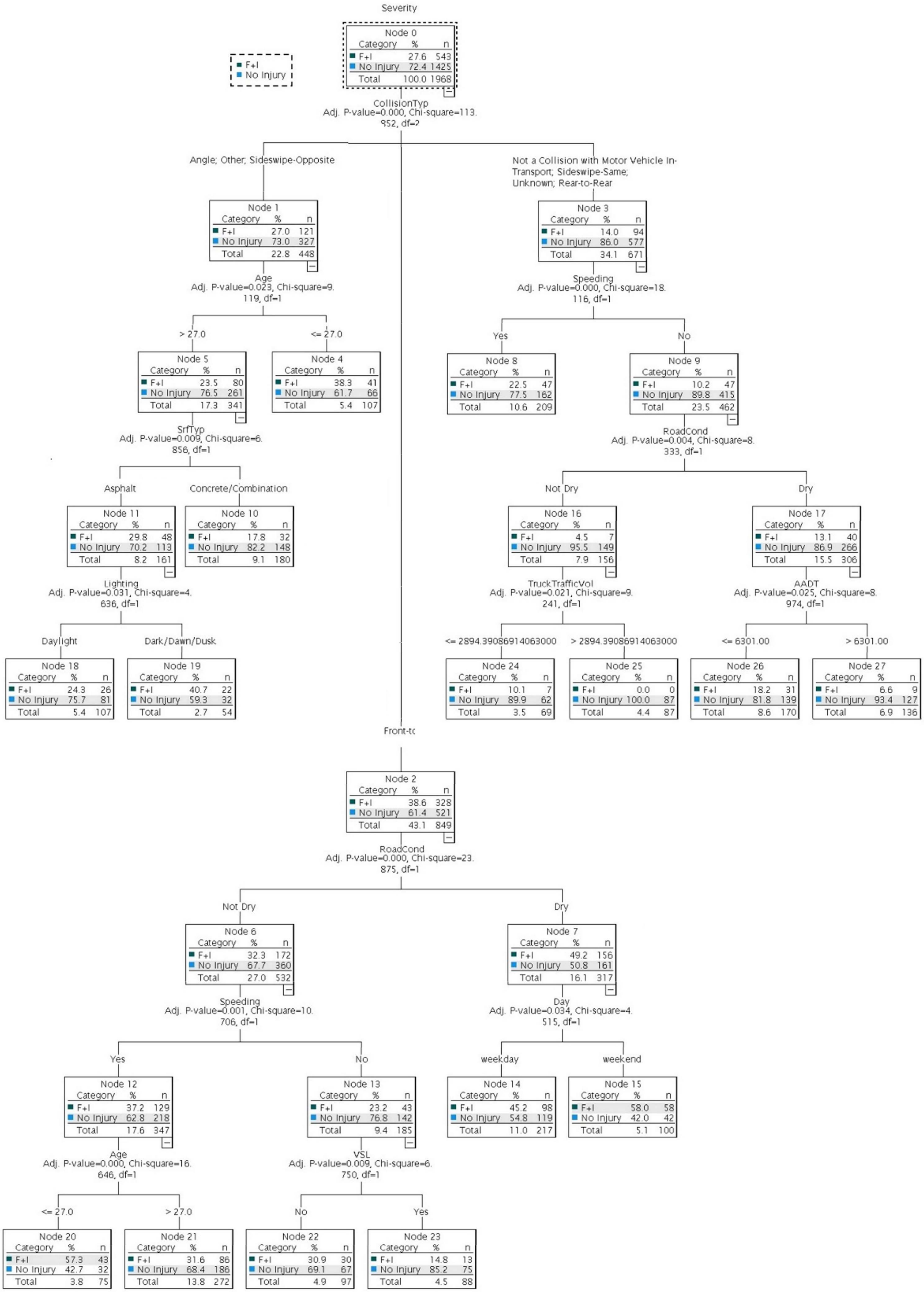

Fig. 4. Multivehicle truck crashes classification tree.

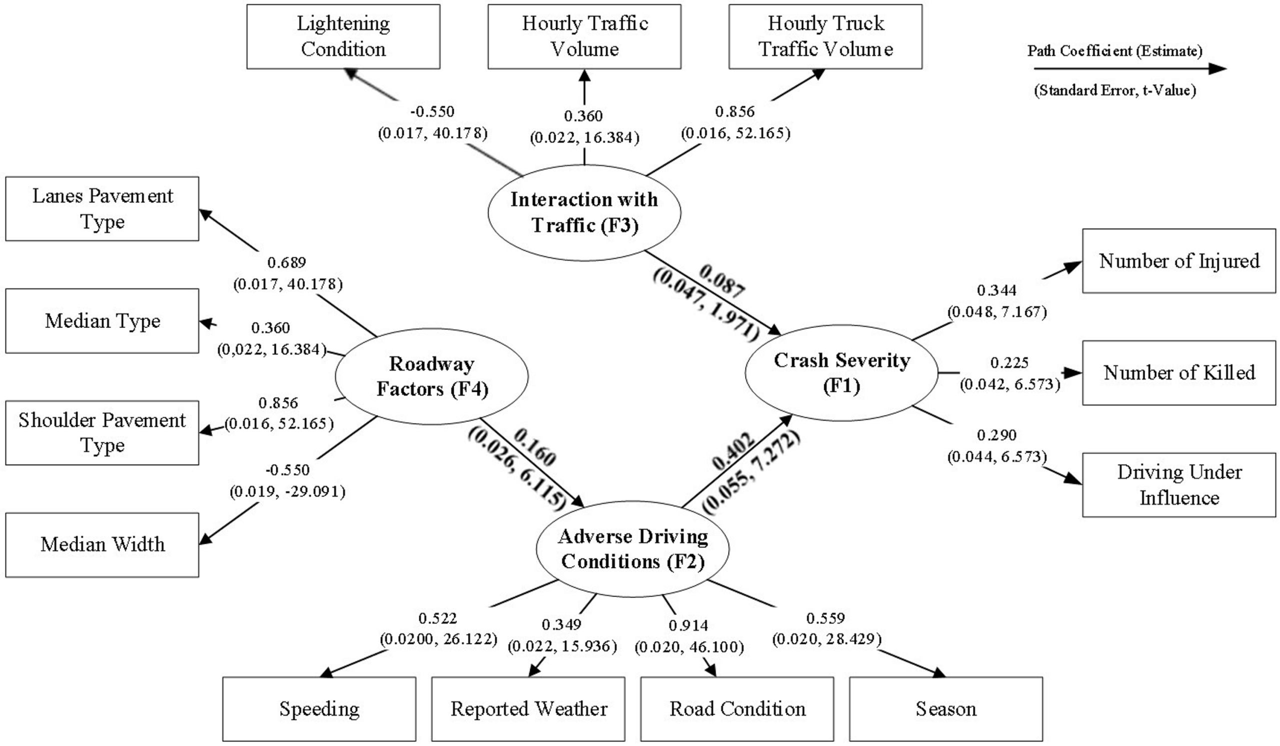

Fig. 5. SEM of single-truck related crashes.

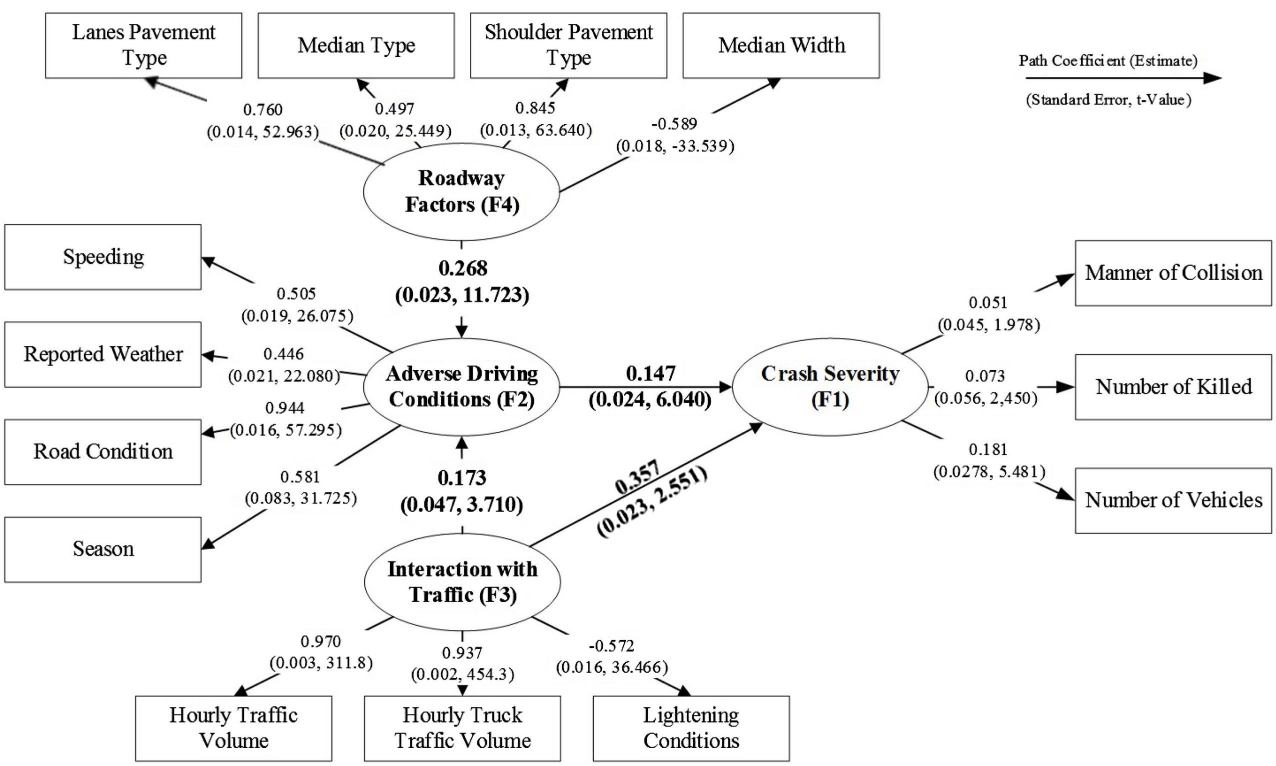

Fig. 6. SEM of multivehicle truck-related crashes.

collisions with a fixed object representing 10.6%. In addition, Fig. 4 shows a total of 7.9% of multivehicle truck crashes occurred in slippery conditions and involving no speeding were a fixed object crashes

SEM Results for Single and Multivehicle Truck Crashes

Figs. 5 and 6 show the developed path model to estimate the crash severity of track-related crashes. The latent variables are presented with oval shapes, rectangular shapes represent the indicator variables, and the arrows represent the direction of the model. Path coefficients (estimate) are provided on the top of the path arrows. The obtained standard error and the significance level presented in the form of the t-value are provided below the path arrow. For the two developed models, the crash severity latent variable was considered the endogenous variable. Direct and indirect relationships could be observed from the developed models. Direct relationship occurs when the exogenous variable is directly connected to the endogenous latent variable (i.e., interaction with traffic is directly connected to the crash severity). Likewise, the indirect relationship occurs when the exogenous latent variable is connected to the endogenous latent variable through an intermediate exogenous latent variable (i.e., the roadway information is indirectly connected to the crash severity). The provided path coefficients demonstrate the standardized estimates for the linear equations, in which all the provided coefficients were significant at a 95% confidence level.

Interpretation of the Measurement Models

For both developed SEMs, three indicator variables were factored in to express the latent variable "interaction with traffic." Two of these three indicator variables had a positive estimate, indicating a higher interaction with traffic was obtained with the increase in the hourly traffic and the hourly truck traffic. The estimates showed that for every unit increase in hourly traffic and hourly truck traffic, the interaction latent variable increases by 0.36 and 0.856 units for the single-truck crashes model and 0.97 and 0.937 units for multivehicle truck crashes, respectively. Lighting conditions had a negative sign, indicating a lower interaction with traffic was obtained at nighttime. This result is due to the lower traffic volumes obtained at nighttime compared to daytime.

The exogenous latent variable named "adverse driving conditions" was measured with four indicator variables, in which all of them had a positive estimate. The estimate of the season variable indicates that more challenging driving conditions were obtained in winter compared to summer, with a coefficient of 0.559 for the single-truck model and 0.581 for multivehicle truck model. Moreover, adverse road surface conditions increase the adverse driving conditions, which could be due to the lower coefficient of friction. Furthermore, the increase in the severity of the reported weather provided more complex driving conditions. Additionally, speeding was found to affect the adverse driving conditions latent variable. The estimate showed that motorists tended to drive above appropriate driving speeds in adverse weather conditions.

The fourth exogenous latent variable named "roadway factors" showed that rigid pavement (concrete pavement) for carriageways and shoulders was among the factors that increased the crash severity for trucks, with coefficients equal to 0.689 and 0.760 for single-truck and multivehicle truck SEMs, respectively. Additionally, depressed and wide medians increased the crash injury severity of truck-related crashes. The measurement model for the single-vehicle truck crashes showed that three indicator variables form the severity of the crash as an endogenous latent variable. The three indicator variables were the number of injured, the number of killed, and driving under the influence (DUI). The coefficient estimates showed that the crash severity would increase with the increase in the three indicator variables. The DUI variable was a binary variable with a zero assigned to no DUI involvement and one for DUI involved in the crash. The endogenous latent variable for the multivehicle truck crashes showed that one-unit increase in the number of killed and the number of vehicles involved in the crash would increase the severity of the multivehicle truck crash by 0.073 and 0.181 units, respectively. Additionally, the manner of collision has a significant effect on the severity of multivehicle truck crashes.

Interpretation of the Structural Models

The path model for the single-truck crashes showed that the roadway factors had an indirect effect on the crash severity of the single-truck-related crashes. The positive sign of the path coefficients for the roadway factors and the adverse driving conditions showed that a one-unit increase in roadway factors would worsen the driving conditions by 0.160 units and indirectly contribute to increasing the single-truck-related crashes by 0.402 units. Additionally, the increase in traffic interaction would increase the single-truck-related crashes by 0.087 units.

The path model of the multivehicle truck crashes showed that the interaction with traffic and adverse driving conditions had a significant effect on increasing the outcome severity for truck crashes. The coefficients showed that for each unit increase in the interaction with traffic, the crash severity increases by 0.357 units and 0.147 units for each unit increase in adverse driving conditions. Roadway factors had an indirect effect on crash severity, as pavement type (concrete), median type (depressed), and increase in median width increase the adverse driving conditions, increasing the truck crash severity. Moreover, the path model showed that the traffic interaction latent variable affects the adverse driving conditions.

Goodness of Fit and Statistical Power

To assess the model fit, several thresholds should be met. Hooper et al. (2008) provided guidelines to assess the model fit, which was considered as the assessment golden rules (Hooper et al. 2008). The several provided model fit indices reflect a different aspect of model fit. Table 4 concluded the obtained model fit indices and the threshold for each index. Several path models were developed, however, and the model with the lowest Akaike information criterion (AIC) was reported in this study.

Although some of the obtained model fit indices did not meet the threshold minimum/maximum limits, the model fit for both the developed models provide an acceptable fit, given the slight

Table 4. Model fit indices and statistical power summary for the developed SEMs

Model fit index	Obtained values of indices		Threshold values
	Single-truck crash	Multivehicle truck crash	
Standardized root mean square residual (SRMR)	0.0544	0.0506	<0.050
Goodness of fit index (GFI)	0.9201	0.9126	>0.900
Parsimony index—adjusted GFI (AGFI)	0.8950	0.8909	>0.900
RMSEA estimate	0.0573	0.0603	<0.055
Bentler comparative fit index (CFI)	0.9029	0.8990	>0.900
Akaike information criterion (AIC)	1,353	1,925	Lower is better

Table 5. Direct, indirect, and total effects for single-truck and multivehicle truck SEM

Latent variable	Direct effect Estimate	Direct effect Standard error	Indirect effect Estimate	Total effect Estimate
Effects on the crash severity of the single-truck crash SEM				
Interaction with traffic	0.087	0.024	NA	0.087[a]
Adverse driving conditions	0.402	0.048	NA	0.402[a]
Roadway factors	NA	NA	0.064	0.064[b]
Effects on the crash severity of multivehicle truck crash SEM				
Interaction with traffic	0.357	0.032	0.025	0.382
Adverse driving conditions	0.147	0.024	NA	0.147[a]
Roadway factors	NA	NA	0.039	0.039[b]

[a]The total expresses the direct effect only (no indirect effect is presented).
[b]The total expresses the indirect effect only (no direct effect is presented).

variation from the limits. The obtained SRMRs were above the threshold limit; however, values below 0.08 would provide a good model fit (Hu and Bentler 1999). Additionally, the parsimony index was slightly below the limits of the acceptable threshold. The statistical power of the model could be measured using the RMSEA value. Even though the RMSEA was slightly above the threshold, it is still considered to have adequate statistical power as the obtained value is below 0.08. The obtained RMSEA confidence intervals provided a significance level of 90%. Additionally, the RMSEA is considered a less preferable index to assess the goodness of fit when having relatively large sample sizes (Hu and Bentler 1999).

Comparative Discussion for the Single-Truck and the Multivehicle Truck Crash Models

Similar latent variables constructed to form the measurement models were found for the two investigated crash types; single-truck crashes, and multivehicle truck crashes. However, their direct/indirect effects, as well as the magnitude of estimates on the crash severity of large trucks were different. Table 5 concludes the direct, indirect, and total effects of the latent variables on the investigated crash severity for large trucks.

To determine the indirect effect of a latent variable on the endogenous latent variable, the estimated coefficients on the path should be multiplied. For example, the indirect effect of the interaction with traffic on crash severity of multivehicle truck crashes is 0.025, which is the product of 0.173 multiplied by 0.147. The total effect is the calculated general effect for the direct and indirect effects on the endogenous variable. The total effect could be calculated using the estimates obtained from the path model by adding the direct and indirect effects. For the same example, adding the direct (0.357) and indirect effects (0.025) for the interaction with traffic on crash severity of multivehicle truck crashes would provide a total effect of 0.382.

Interaction with traffic was significant for both models. It had only a direct effect on single-truck crashes, while it had a direct and indirect effect on the outcome severity of multivehicle truck crashes. Its total effect on single-truck crashes was lower compared to its effect on multivehicle truck crashes. This result shows that the impact of one-unit increase in the "interaction with surrounding traffic" on the severity of multivehicle truck crashes is 4.3 times higher compared to its impact on the severity of single-truck crashes.

Adverse driving conditions had the highest impact on single-truck crashes with a total effect of 0.402, which is more than 2.7 times its impact on multivehicle truck crashes. The increased effect of adverse conditions on the severity of single-truck crashes could be due to the high center of gravity (CG) for trucks. The high truck's CG could worsen the controllability of trucks when encountering slippery road surfaces or high wind speeds (Gaweesh et al. 2022; Irfan Ahmed et al. 2020).

The roadway factors had an indirect total effect on truck crash severity for both models. Its impact was nearly 1.6 times higher on the truck crash severity for single-truck crashes compared to the multivehicle truck crashes. For the developed models, the roadway factors affect the crash severity of trucks through the adverse driving conditions. This implies that roadway factors become a significant indicator when encountering adverse driving conditions explained by severe weather conditions, adverse road surface, and speeding.

Suggested Improvements to Enhance Truck Traffic Safety

The goal of this study is to evaluate the safety of commercial trucks by identifying observed variables as well as latent variables that might increase the crash severity of trucks. This will help local and regional transportation agencies in regions that share similar characteristics to determine countermeasures and treatments that could help in enhancing traffic safety and operations for commercial vehicles.

The utilized approaches clarified several factors that contribute to an increased truck crash severity on interstate roads. As commonly known, the traffic as the main crash exposure was the main factor contributing to increased crash severity. However, this study utilized hourly volume, as well as truck hourly volume, to show their impact on increasing crash severity. The analysis showed that other factors were, directly or indirectly, coupled with traffic volumes in causing an increased crash severity. Usually, visibility and weather conditions when accompanied by higher traffic volumes increase crash injury severity. The results showed that the presence of guardrails is important to single-truck crashes, especially in reduced visibility. This implies that guardrails and lateral obstructions as roadside elements might help truck drivers in identifying roadway alignment in limited visibility conditions. Although delineators are always presented at the roadside, enhancing their maintenance by installing better reflective sheets would help to increase their conspicuity. This would help drivers to better detect the road alignment, especially in adverse weather events. Moreover, longitudinal bright color strips to the roadside barriers would help to easily detect the road alignment at locations with challenging road geometry at events with reduced visibility, and at locations that encounter increased run-off road crashes.

This result is in line with the importance of rumble strips to identify road alignment for truck drivers. In the winter season, with the presence of rumble strips, increased crash severity was observed.

However, this result seems contradicting because in the winter season, rumble strips on I-80 are completely covered with snow, potentially diminishing their role in alerting drivers when departing the roadway. With the excessive number of rural roadway miles and limited winter maintenance resources, it is challenging to timely clear travel lanes and shoulders. Several DOTs adopt tow plows (TPs) to increase the efficiency and productivity of the winter maintenance operation. It would be suitable to include TPs in winter operations to include shoulders in snow removal operations instead of using them as snow storage areas.

Speeding was also observed as one of the factors coupled with truck volumes, the presence of VSL, and reduced visibility to cause an increased truck crash severity. This result clarifies the low compliance rates for the provided speed limits within I-80. Endorsement of increased speed enforcement is recommended, as well as conducting public campaigns to educate motorists about the dangerous outcomes of speeding during adverse weather conditions, which is in agreement with previous studies (Ahmed et al. 2011; Ahmed and Abdel-aty 2012; Gaweesh and Ahmed 2019; Yu et al. 2013). Additionally, this might be correlated with the disseminated variable speed limits within I-80. If the provided speeds do not match the current driving conditions, motorists tend to over speed resulting in high-speed variability. The increased speed variability within the corridor sections may result in increased crash likelihood. Thus, enhancing the current VSL algorithms as well as adding more sections with VSL corridors would be recommended.

The results of this study highlighted the importance of increasing the situational awareness of commercial truck drivers about upcoming hazardous events. This could be performed by communicating information regarding upcoming adverse weather events, bad road surface conditions, challenging roadway geometry, and traffic conditions in the connected vehicle (CV) pilot deployment on I-80. By communicating such information, truck drivers' preparedness for encountering dangerous events will be enhanced.

Conclusions and Discussions

The trucking industry has a significant effect on the US economy. Commercial trucks are responsible for transporting a significant amount of freight tonnage. With the anticipated increase in truck traffic, concerns related to traffic safety are raised. Investigating factors that might reduce crash frequency and/or crash severity would provide insights to develop mitigation plans and effective countermeasures. Several studies were conducted with the focus of determining the influential factor that increases the severity of truck-related crashes. Most of these studies utilized discrete choice models to examine the relationship between the severity of a crash and the indicator variables by developing a direct relationship between them. However, the discrete choice model cannot explain the complex interrelationship between the indicator variables and the crash severity.

This study helped in enhancing traffic safety for commercial trucks on Wyoming interstate roads by identifying the most important factors as well as the latent factors that influence the severity level of truck crashes. This study presented a systematic approach to developing the measurement model of the SEM, in which decision trees, EFA, CFA, and engineering judgment were used to select variables that will be used to develop the measurement models. The variation and covariation within and between the constructed measurement models were accounted for by using the SEM as it is considered one of its advantages. From the dataset perspective, unlike other studies, a disaggregate level of the traffic volumes as the main crash exposure was used along with several datasets (i.e. crash data, environmental data, temporal data, roadway factors, and driver factors) to account for other confounding factors. Furthermore, direct and indirect relationships between factors that lead to increased crash severity of large trucks were determined. Direct and indirect effects would clarify which latent variables has a significant effect only with the presence of other variables.

The SEM clarified that the indicator variables measuring the crash severity are mostly related to the nature of the crash. The number of vehicles involved in the crash was the most significant variable that measure the crash severity of multivehicle truck crashes. On the other hand, the number of injured people in the crash was the most significant variable that measures the crash severity of single-truck crashes. Additionally, the developed path model highlights the difference between the natures of the two crash types. The interaction with the surrounding traffic was the most significant latent variable affecting the crash severity of multivehicle truck crashes and the adverse weather conditions were the most significant latent variable affecting the crash severity of single-truck crashes. The results showed that speeding in adverse road and weather conditions were among the variables that increased the severity of truck-related crashes.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

The authors would like to thank the WYDOT for providing the data that were used in this study, and for funding this research—Award No. RS04220. All opinions and results are solely those of the authors.

References

Abellán, J., G. López, and J. de Oña. 2013. "Analysis of traffic accident severity using decision rules via decision trees." *Expert Syst. Appl.* 40 (15): 6047–6054. https://doi.org/10.1016/j.eswa.2013.05.027.

Ahmed, M., H. Huang, M. Abdel-Aty, and B. Guevara. 2011. "Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway." *Accid. Anal. Prev.* 43 (4): 1581–1589. https://doi.org/10.1016/j.aap.2011.03.021.

Ahmed, M. M., and M. A. Abdel-aty. 2012. "The viability of using automatic vehicle identification data for real-time crash prediction." *IEEE Trans. Intell. Transp. Syst.* 13 (2): 459–468. https://doi.org/10.1109/TITS.2011.2171052.

Ahmed, M. M., R. Franke, K. Ksaibati, and D. S. Shinstine. 2018. "Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models." *Accid. Anal. Prev.* 117 (Apr): 106–113. https://doi.org/10.1016/j.aap.2018.04.011.

Ahmed, M. M., S. Gaweesh, and G. Yang. 2019. "A preliminary investigation into the impact of connected vehicle human-machine interface on driving behavior." *IFAC-PapersOnLine* 51 (34): 227–229. https://doi.org/10.1016/j.ifacol.2019.01.051.

Ahmed, M. M., G. Yang, and S. Gaweesh. 2020. "Assessment of drivers' perceptions of connected vehicle—Human machine interface for driving under adverse weather conditions: Preliminary findings from Wyoming." *Front. psychol.* 11 (Aug): 1889. https://doi.org/10.3389/fpsyg.2020.01889.

Ambak, K., R. Ismail, R. A. Abdullah, and M. N. Borhan. 2010. "Prediction of helmet use among Malaysian motorcyclist using structural equation modeling." *Aust. J. Basic Appl. Sci.* 4 (10): 5263–5270.

Badea-Romero, A., and J. Lenard. 2013. "Source of head injury for pedestrians and pedal cyclists: Striking vehicle or road?" *Accid. Anal. Prev.* 50 (Jun): 1140–1150. https://doi.org/10.1016/j.aap.2012.09.024.

Barman, S., and R. Bandyopadhyaya. 2020. "Crash severity analysis for low-speed roads using structural equation modeling considering shoulder- and pavement-distress conditions." *J. Transp. Eng. A Syst.* 146 (7): 1–10. https://doi.org/10.1061/jtepbs.0000373.

Beauducel, A., and P. Y. Herzberg. 2006. "On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA." *Struct. Equation Model.* 13 (2): 186–203. https://doi.org/10.1207/s15328007sem1302_2.

Brown, T. A. 2015. *Confirmatory factor analysis for applied research*. New York: Guilford Publications.

Byrne, B. M. 1998. *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum Associates.

Cafiso, S., A. di Graziano, G. di Silvestro, G. la Cava, and B. Persaud. 2010. "Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables." *Accid. Anal. Prev.* 42 (4): 1072–1079. https://doi.org/10.1016/j.aap.2009.12.015.

Chang, L.-Y., and J.-T. Chien. 2013. "Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model." *Saf. Sci.* 51 (1): 17–22. https://doi.org/10.1016/j.ssci.2012.06.017.

Chen, C., G. Zhang, X. C. Liu, Y. Ci, H. Huang, J. Ma, Y. Chen, and H. Guan. 2016. "Driver injury severity outcome analysis in rural interstate highway crashes: A two-level Bayesian logistic regression interpretation." *Accid. Anal. Prev.* 97 (Jul): 69–78. https://doi.org/10.1016/j.aap.2016.07.031.

Chen, F., and S. Chen. 2011. "Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways." *Accid. Anal. Prev.* 43 (5): 1677–1688. https://doi.org/10.1016/j.aap.2011.03.026.

Chen, F., and S. Chen. 2013. "Differences in injury severity of accidents on mountainous highways and non-mountainous highways." *Procedia Soc. Behav. Sci.* 96 (Oct): 1868–1879. https://doi.org/10.1016/j.sbspro.2013.08.212.

Cho, S.-H., D.-K. Kim, and S.-Y. Kho. 2017. "Latent factors of severity in truck-involved and non-truck-involved crashes on freeways." *Int. J. Transp. Veh. Eng.* 11 (7): 920–927. https://doi.org/10.5281/zenodo.1131423.

de Oña, J., G. López, R. Mujalli, and F. J. Calvo. 2013. "Analysis of traffic accidents on rural highways using latent class clustering and Bayesian networks." *Accid. Anal. Prev.* 51 (Aug): 1–10. https://doi.org/10.1016/j.aap.2012.10.016.

Dong, X., K. Xie, and H. Yang. 2022. "How did COVID-19 impact driving behaviors and crash Severity? A multigroup structural equation modeling." *Accid. Anal. Prev.* 172 (Dec): 106687. https://doi.org/10.1016/j.aap.2022.106687.

DOT. 2020. "National freight strategic plan." Accessed September 4, 2020. https://www.transportation.gov/sites/dot.gov/files/2020-09/NFSP_fullplan_508_0.pdf.

Farag, S., T. Schwanen, M. Dijst, and J. Faber. 2007. "Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping." *Transp. Res. Part A Policy Pract.* 41 (2): 125–141. https://doi.org/10.1016/j.tra.2006.02.003.

Federal Motor Carrier Safety Administration. 2019. *Large truck and bus crash facts 2017*. Washington, DC: Federal Motor Carrier Safety Administration.

Forero, C. G., A. Maydeu-Olivares, and D. Gallardo-Pujol. 2009. "Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation." *Struct. Equation Model.* 16 (4): 625–641. https://doi.org/10.1080/10705510903203573.

Gaweesh, S., M. N. Khan, and M. M. Ahmed. 2021a. "Development of a novel framework for hazardous materials placard recognition system to conduct commodity flow studies using artificial intelligence Alexnet convolutional neural network." *Transp. Res. Rec.* 2675 (11): 1357–1371. https://doi.org/10.1177/03611981211026653.

Gaweesh, S. M., I. Ahmed, and M. M. Ahmed. 2023. "Analysis framework to assess crash severity for large trucks on rural interstate roads utilizing the latent class and random parameter model." *Transp. Res. Rec.* 10 (Jun): 158–175. https://doi.org/10.1177/03611981231158627.

Gaweesh, S. M., I. U. Ahmed, M. M. Ahmed, and S. S. Wulff. 2022. "Developing statewide safety performance functions for commercial trucks transporting hazardous materials on interstate rural roads in Wyoming." *Transp. Res. Rec.* 2022 (1): 036119812211032. https://doi.org/10.1177/03611981221103231.

Gaweesh, S. M., and M. M. Ahmed. 2019. "Evaluating the safety effectiveness of a weather-based variable speed limit for a rural mountainous freeway in Wyoming." *J. Transp. Saf. Security* 10 (Jun): 1–26. https://doi.org/10.1080/19439962.2019.1583707.

Gaweesh, S. M., and M. M. Ahmed. 2020. "Evaluating the safety effectiveness of a weather-based variable speed limit for a rural mountainous freeway in Wyoming." *J. Transp. Saf. Security* 12 (10): 1205–1230. https://doi.org/10.1080/19439962.2019.1583707.

Gaweesh, S. M., M. M. Ahmed, and A. V. Piccorelli. 2019. "Developing crash prediction models using parametric and nonparametric approaches for rural mountainous freeways: A case study on Wyoming Interstate 80." *Accid. Anal. Prev.* 123 (Dec): 176–189. https://doi.org/10.1016/j.aap.2018.10.011.

Gaweesh, S. M., A. Khoda Bakhshi, and M. M. Ahmed. 2021b. "Safety performance assessment of connected vehicles in mitigating the risk of secondary crashes: A driving simulator study." In *Proc., Transportation Research Board 100th Annual Meeting*. Washington, DC: Transportation Research Board.

Geedipally, S. R., and D. Lord. 2010. "Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models." *Accid. Anal. Prev.* 42 (4): 1273–1282. https://doi.org/10.1016/j.aap.2010.02.004.

Hamdar, S. H., H. S. Mahmassani, and R. B. Chen. 2008. "Aggressiveness propensity index for driving behavior at signalized intersections." *Accid. Anal. Prev.* 40 (1): 315–326. https://doi.org/10.1016/j.aap.2007.06.013.

Hassan, H. 2011. *Improving traffic safety and drivers' behavior in reduced visibility conditions*. Orlando, FL: Univ. of Central Florida.

Hassan, H. M., and M. A. Abdel-Aty. 2013. "Exploring the safety implications of young drivers' behavior, attitudes and perceptions." *Accid. Anal. Prev.* 50 (Jun): 361–370. https://doi.org/10.1016/j.aap.2012.05.003.

Hatcher, L., and N. O'Rourke. 2013. *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.

Hooper, D., J. Couglan, and M. R. Mullen. 2008. "Evaluating model fit: A synthesis of the structural equation modelling literature." In Vol. 2008 of *Proc., 7th European Conf. on Research Methodology for Business and Management Studies*, 195–200. South Oxfordshire, UK: Academic Conferences and Publishing.

Hu, L., and P. M. Bentler. 1999. "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." *Struct. Equation Model.* 6 (1): 1–55. https://doi.org/10.1080/10705519909540118.

Hurley, A. E., T. A. Scandura, C. A. Schriesheim, M. T. Brannick, A. Seers, R. J. Vandenberg, and L. J. Williams. 1997. "Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives." *J. Organ. Behav.* 18 (6): 667–683. https://doi.org/10.1002/(SICI)1099-1379(199711)18:6<667::AID-JOB874>3.0.CO;2-T.

Irfan Ahmed, S., M. Gaweesh, and M. Ahmed. 2020. "Exploration of hazardous material truck crashes on Wyoming's interstate roads using a novel Hamiltonian Monte Carlo Markov chain Bayesian inference." *Transp. Res. Rec.* 2674 (9): 661–675. https://doi.org/10.1177/0361198120931103.

Islam, S., S. L. Jones, and D. Dye. 2014. "Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama." *Accid. Anal. Prev.* 67 (Sep): 148–158. https://doi.org/10.1016/j.aap.2014.02.014.

Kaplan, D. 2000. *Structural equation modeling. Foundations and extensions*. London: SAGE.

Kervick, A. A., M. J. Hogan, D. O'Hora, and K. M. Sarma. 2015. "Testing a structural model of young driver willingness to uptake smartphone driver support systems." *Accid. Anal. Prev.* 83 (Jun): 171–181. https://doi.org/10.1016/j.aap.2015.07.023.

Khattak, A. J., and F. Targa. 2004. "Injury severity and total harm in truck-involved work zone crashes." *Transp. Res. Rec.* 1877 (1): 106–116. https://doi.org/10.3141/1877-12.

Khoda Bakhshi, A., S. M. Gaweesh, and M. M. Ahmed. 2021. "The safety performance of connected vehicles on slippery horizontal curves through enhancing truck drivers' situational awareness: A driving simulator experiment." *Transp. Res. Part F Traffic Psychol. Behav.* 79 (Jun): 118–138. https://doi.org/10.1016/j.trf.2021.04.017.

Kim, K., P. Pant, and E. Yamashita. 2011. "Measuring influence of accessibility on accident severity with structural equation modeling." *Transp. Res. Rec.* 2236 (1): 1–10. https://doi.org/10.3141/2236-01.

Lee, J.-Y., J.-H. Chung, and B. Son. 2008. "Analysis of traffic accident size for Korean highway using structural equation models." *Accid. Anal. Prev.* 40 (6): 1955–1963. https://doi.org/10.1016/j.aap.2008.08.006.

Lei, P. W. 2009. "Evaluating estimation methods for ordinal data in structural equation modeling." *Qual. Quant.* 43 (3): 495–507. https://doi.org/10.1007/s11135-007-9133-z.

Lemp, J. D., K. M. Kockelman, and A. Unnikrishnan. 2011. "Analysis of large truck crash severity using heteroskedastic ordered probit models." *Accid. Anal. Prev.* 43 (1): 370–380. https://doi.org/10.1016/j.aap.2010.09.006.

Li, C. H. 2016. "Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares." *Behav. Res. Methods* 48 (3): 936–949. https://doi.org/10.3758/s13428-015-0619-7.

Mohamadi Hezaveh, A., M. Azad, and C. Cherry. 2018. "Pedestrian crashes in Tennessee: A data mining approach." In *Proc., Transportation Research Board 97th Annual Meeting*. Washington, DC: Transportation Research Board.

Montella, A., M. Aria, A. D'Ambrosio, and F. Mauriello. 2012. "Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery." *Accid. Anal. Prev.* 49 (Apr): 58–72. https://doi.org/10.1016/j.aap.2011.04.025.

Moridpour, S., E. Mazloumi, and M. Mesbah. 2015. "Impact of heavy vehicles on surrounding traffic characteristics." *J. Adv. Transp.* 49 (Aug): 535–552. https://doi.org/10.1002/atr.1286.

Naik, B., L. W. Tung, S. Zhao, and A. J. Khattak. 2016. "Weather impacts on single-vehicle truck crash injury severity." *J. Saf. Res.* 58 (Jul): 57–65. https://doi.org/10.1016/j.jsr.2016.06.005.

Nazim, A., and S. Ahmad. 2013. "A comparison between ordinary least square (OLS) and structural equation modeling (SEM) methods in estimating the influencial factors of 8th grades student's mathematics achievement in Malaysia." *Int. J. Sci. Eng. Res.* 4 (7): 717–722.

Prati, G., L. Pietrantoni, and F. Fraboni. 2017. "Using data mining techniques to predict the severity of bicycle crashes." *Accid. Anal. Prev.* 101 (Jan): 44–54. https://doi.org/10.1016/j.aap.2017.01.008.

Raddaoui, O., M. M. Ahmed, and S. M. Gaweesh. 2020. "Assessment of the effectiveness of connected vehicle weather and work zone warnings in improving truck driver safety." *IATSS Res.* 44 (3): 1–11. https://doi.org/10.1016/j.iatssr.2020.01.001.

Ramos, A. 2014. *Simultaneous equation modeling for crash rate of freeway segments*. Las Vegas: Univ. of Nevada.

Rhemtulla, M., P. É. Brosseau-Liard, and V. Savalei. 2012. "When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions." *Psychol. Methods* 17 (3): 354–373. https://doi.org/10.1037/a0029315.

Rosseel, Y. 2014. *The Lavaan tutorial*. Ghent, Belgium: Ghent Univ.

Shaaban, K., S. Gaweesh, and M. Ahmed. 2020. "Investigating in-vehicle distracting activities and crash risks for young drivers using structural equation modeling." *PLoS One* 15 (7): e0235325. https://doi.org/10.1371/journal.pone.0235325.

Shaaban, K., S. Gaweesh, and M. M. Ahmed. 2018. "Characteristics and mitigation strategies for cell phone use while driving among young drivers in Qatar." *J. Transp. Health* 8 (Mar): 6–14. https://doi.org/10.1016/j.jth.2018.02.001.

Siskind, V., D. Steinhardt, M. Sheehan, T. O'Connor, and H. Hanks. 2011. "Risk factors for fatal crashes in rural Australia." *Accid. Anal. Prev.* 43 (3): 1082–1088. https://doi.org/10.1016/j.aap.2010.12.016.

Song, Y., and Y. Lu. 2015. "Decision tree methods: Applications for classification and prediction." *Shanghai Arch. Psychiatry* 27 (2): 130. https://doi.org/10.11919/J.ISSN.1002-0829.215044.

Suhr, D. 2006. *The basics of structural equation modeling*. Irvine, CA: SAS User Group of the Western Region of the United States.

Tahfim, S. A. S., and C. Yan. 2022. "Analysis of spatial concentrations of large-truck crashes using data mining methods." *Int. J. Saf. Security Eng.* 12 (1): 55–64. https://doi.org/10.18280/ijsse.120107.

Taylor, S. G., B. J. Russo, and E. James. 2018. "A comparative analysis of factors affecting the frequency and severity of freight-involved and non-freight crashes on a major freight corridor freeway." *Transp. Res. Rec.* 2672 (34): 49–62. https://doi.org/10.1177/0361198118776815.

Ture, M., F. Tokatli, and I. Kurt. 2009. "Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients." *Expert Syst. Appl.* 36 (2): 2017–2026. https://doi.org/10.1016/j.eswa.2007.12.002.

Uddin, M., and N. Huynh. 2017. "Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways." *Accid. Anal. Prev.* 108 (Mar): 44–55. https://doi.org/10.1016/j.aap.2017.08.009.

Wang, K., and X. Qin. 2014. "Use of structural equation modeling to measure severity of single-vehicle crashes." *Transp. Res. Rec.* 2432 (1): 17–25. https://doi.org/10.3141/2432-03.

Worth, M., S. Guerrero, and A. Meyers. 2016. *2016 freight quick facts report*. Rep. No. FHWA-HOP-16-083. Washington, DC: Federal Highway Administration.

Wyoming DOT. 2018. *Wyoming Strategic Highway Safety Plan*. Cheyenne, WY: Wyoming DOT.

Xu, C., D. Li, Z. Li, W. Wang, and P. Liu. 2018. "Utilizing structural equation modeling and segmentation analysis in real-time crash risk assessment on freeways." *KSCE J. Civ. Eng.* 22 (7): 2569–2577. https://doi.org/10.1007/s12205-017-0629-3.

Yang-Wallentin, F., K. G. Jöreskog, and H. Luo. 2010. "Confirmatory factor analysis of ordinal variables with misspecified models." *Struct. Equation Model.* 17 (3): 392–423. https://doi.org/10.1080/10705511.2010.489003.

Yu, R., M. Abdel-Aty, and M. Ahmed. 2013. "Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors." *Accid. Anal. Prev.* 50 (Mar): 371–376. https://doi.org/10.1016/j.aap.2012.05.011.

Zheng, Z., P. Lu, and B. Lantz. 2018. "Commercial truck crash injury severity analysis using gradient boosting data mining model." *J. Saf. Res.* 65 (Jul): 115–124. https://doi.org/10.1016/j.jsr.2018.03.002.

Zhu, X., and S. Srinivasan. 2011. "A comprehensive analysis of factors influencing the injury severity of large-truck crashes." *Accid. Anal. Prev.* 43 (1): 49–57. https://doi.org/10.1016/j.aap.2010.07.007.