Relationship between road traffic features and accidents: An application of two-stage decision-making approach for transportation engineers

Peer-reviewed author version

SHAH, Syyed; Ahmad, Naveed; SHEN, Yongjun; Ahmed Kamal Mumtaz; BASHEER, Muhammad & BRIJS, Tom (2019) Relationship between road traffic features and accidents: An application of two-stage decision-making approach for transportation engineers. In: Journal of safety research, 69, p. 201-215.

DOI: 10.1016/j.jsr.2019.01.001
Handle: http://hdl.handle.net/1942/28338

Relationship between Road Traffic Features and Accidents: An Application of Two-Stage Decision Making Approach for Transportation Engineers

Syyed Adnan Raheel Shah [1,2,*], Naveed Ahmad [1], Yongjun Shen [3], Mumtaz Ahmed Kamal [1], Muhammad Aamir Basheer [2] and Tom Brijs [2]

[1] Taxila Institute of Transportation Engineering, Department of Civil Engineering, University of Engineering & Technology, Taxila 47050, Pakistan; n.ahmad@uettaxila.edu.pk(NA); dr.kamal@uettaxila.edu.pk(MAK)
[2] Transportation Research Institute (IMOB), Hasselt University, Agoralaan, B-3590 Diepenbeek, Belgium; aamirbashir50@yahoo.com (M.A.B.); tom.brijs@uhasselt.be (T.B.)
[3] School of Transportation, Southeast University, Sipailou 2, 210096 Nanjing, China; shenyongjun@seu.edu.cn (YS)

* Correspondence: syyed.adnanraheelshah@uhasselt.be; Tel.: +923007914248

Relationship between Road Traffic Features and Accidents: An Application of Two-Stage Decision Making Approach for Transportation Engineers

Abstract: Introduction: An efficient decision-making process is one of the major necessities of road safety performance analysis for human safety and budget allocation procedure. Method: During the road safety analysis procedure, data envelopment analysis (DEA) supports policymakers in differentiating between risky and safe segments of a homogeneous highway. Cross risk, an extension of the DEA models, provides more information about risky segments for ranking purpose. After identification of risky segments, the next goal is to identify those factors, which are major contributors in making that segment risky. Results: This research proposes a methodology to analyze road safety performance by using a combination of DEA with the decision tree (DT) technique. The proposed methodology not only provides a facility to identify problematic road segments with the help of DEA, but also identifies contributing factors with the help of DT. Practical applications: The applicability of the proposed model will help policymakers to identify the major factors contributing to road accidents and analysis of safety performance of road infrastructure to allocate the budget during the decision-making process.

Keywords: Transportation; Roads; Decision Making; Accidents; Risk Evaluation; DEA-DT

1. Introduction

Within the European region, road accidents on motorways are a major contributing factor in road fatalities. Therefore, it is necessary to conduct a safety performance analysis for European motorways. Traffic stream characteristics and road infrastructure geometric design feature are the interrelated and major focus of studying these factors in combination with safety are to analyze core-contributing factor in increasing road accidents. Worldwide, an estimated 1.2 million people are killed in road crashes each year and as many as 50 million are injured(WHO, 2004). It is necessary to improve the standard of road traffic conditions to improve the road safety conditions for the betterment of human life. In this battle, the European Commission fixed the aspiring goal of halving the number of road fatalities by 2010 in its White Paper “European transport policy for 2010: time to decide” of 2001. “A new target for 2020 to halve the number of road deaths compared to 2010 was set by the EU in its “Road Safety Programme 2011-2020”. It is estimated that the number of road accident fatalities in the EU fell by 42% between 2005 and 2014”(CARE, 2016). Road safety performance of Belgium shows that it is in top ten worst countries among the peer European countries is shown in Fig.1.

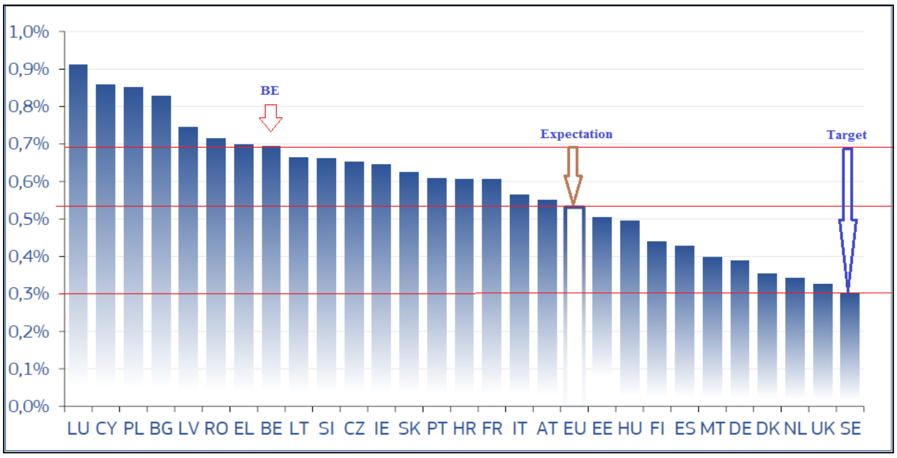

Figure 1. Percentage of road crash fatalities of all fatalities by country (2014)(CARE, 2016)

However realistic approach tells the story that it should be nearest to European average and target should be at least to be the safest country in Schengen region. Road safety administration professionals have constantly endeavored to depict an unequivocal point of view in both quantitative and subjective road safety assessment criteria by setting reliable systems.

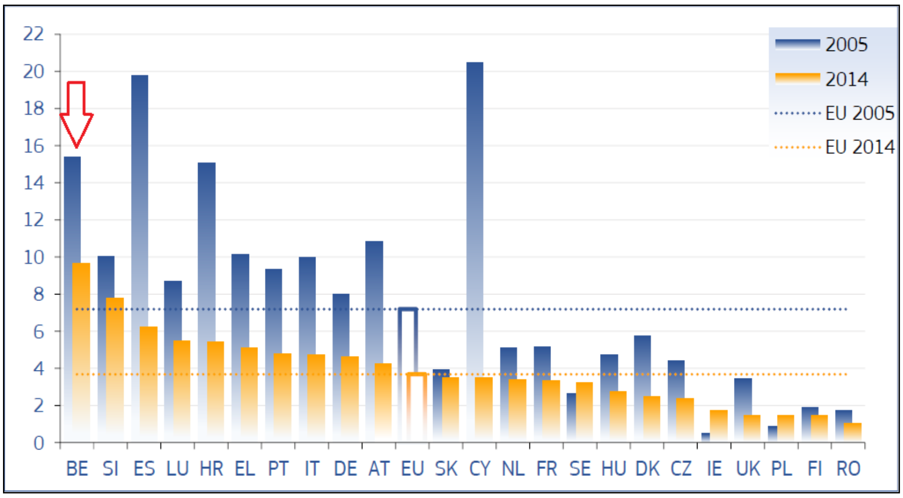

Figure 2. Fatalities rates per million population by country for Motorways in Europe (2005 & 2014)(EU, 2016)

Usually, this type of analysis is dependent on the stage-wise safety performance of a country at almost four levels: federal, regional, provisional and municipal level. At the federal level, decision making is related to motorways and overall performance of Belgium among European countries is not satisfactory on motorways. Previously focus of road safety investigation research was on local level in Europe including Belgium(Eksler & Lassarre, 2008; Mohan, Bangdiwala, & Villaveces, 2017). "Almost 26.000 people were killed in road accidents on motorways in the European Union countries between 2005 and 2014. This number corresponds to 7% of all road fatalities in those countries. There were 3.558 road accident fatalities on motorways in 2005, and the number fell by 48% in 2014 (1.865).The total number of road accident fatalities in the European Union countries also fell significantly over the same decade, by 42%. Although the overall number of road fatalities decreased rather steadily, the trend for motorway fatalities has been more variable. The most significant reduction of the number of fatalities on motorways occurred between 2007–2008. Spain had the highest percentage of fatalities on motorways in 2014 in the EU (17%), followed by Belgium (15%), Slovenia (13%) and the Netherlands (12%). By contrast, the lowest proportion of fatalities occurring on motorways was in Romania (1%) and Poland (2%)"(EU, 2016). According to the European Road Safety Observatory, in 2014 fatality rate was higher in Belgium (9.6) than in the other European countries and hence higher than the average rate (3.7) of the EU countries (EU, 2016) as shown in Fig.2. Therefore, it is necessary to focus on the motorways of Belgium.

As a traffic engineer and decision maker during the safety analysis procedure, contributing factors like geometric design, speed control, and traffic flow characteristics are of major concern. In previous studies, relationship of road crashes fatalities (Dadashova, Arenas-Ramírez, Mira-McWilliams, & Aparicio-Izquierdo, 2016; Givehchi, Hemmativaghef, & Hoveidi, 2017; Golob, Recker, & Pavlis, 2008; Golob, Recker, & Alvarez, 2004; Imprialou, Quddus, Pitfield, & Lord, 2016; Kaye, Lewis, & Freeman, 2018; Lord, Manar, & Vizioli, 2005; Luoma & Sivak, 2007; Ogwueleka, Misra, Ogwueleka, & Fernandez-Sanz, 2014; Pande & Abdel-Aty, 2006; C. Wang, Quddus, & Ison, 2009; Yasin Çodur & Tortum, 2015) have been studied in context of volume/capacity (Golob & Recker, 2004; Lord, Manar, et al., 2005; Pande & Abdel-Aty, 2006; Zhou & Sisiopiku, 1997), vehicles miles travelled (Abdel-Aty, Lee, Siddiqui, & Choi, 2013; Dadashova et al., 2016; Frantzeskakis & Iordanis, 1987; Jovanis & Chang, 1986; Y. Wang & Kockelman, 2013), vehicles hours travelled (Martin, 2002; Yeo, Jang, Skabardonis, & Kang, 2013; Zhou & Sisiopiku, 1997), Speed (Aljanahi, Rhodes, & Metcalfe, 1999; Elvik, 1997; Elvik, Christensen, & Amundsen, 2004; Garber & Ehrhart, 2000; Golob & Recker, 2004; Hamzeie, Savolainen, & Gates, 2017; Imprialou et al., 2016; Kononov, Lyon, & Allery, 2011; Malyshkina & Mannering, 2008; Pande & Abdel-Aty, 2006), flow (Aljanahi et al., 1999; Garber & Ehrhart, 2000; Golob et al., 2008; Golob & Recker, 2003; Golob et al., 2004; Kononov et al., 2011) and geometric design (Fu, Guo, Yuan, Feng, & Ma, 2011; Garber & Ehrhart, 2000; Karlaftis & Golias, 2002; Milton & Mannering, 1996; Noland & Oh, 2004; B. Wang, Hallmark, Savolainen, & Dong, 2017). For motorways, during road safety analysis procedure it is necessary to prioritize the problematic segments of a road and then to the identify contributing factors as well for an econometric based decision-making.

In this paper, we propose a two-stage DEA-DT approach to analyze the road safety condition of motorways. The proposed methodology will be applied to overcome the safety problem for motorways. An illustrative application of the methodology to two motorways of Limburg province in Belgium is given as a case study.

2. Literature Review

2.1. Road Accidents and Safety Management

Road Accidents have been perceived as a standout amongst the most noteworthy human life threatening health problem nowadays. Without new or enhanced improvements, it is anticipated that with the increase in use of vehicles, it will be the fifth leading cause of death by the end of 2030 (OECD, 2002; WHO, 2004, 2006, 2015). Worldwide the statistics about the road accident deaths is 1.2 million per year and an addition of that 20 to 50 million are injuries due to the same reason. One of the leading indicators used for road safety performance measurement is fatality rate (fatalities per no. of population or no. of registered vehicles)(Wegman et al., 2008) (Al Haji, 2005). In last decade, due to lack of information about road safety performance indicators, development of a composite road safety index has developed an interest globally against the traditional approach (Al Haji, 2005; ETSC, 2001; Wegman et al., 2008). Road safety and its management is a complex system which is governed in five major dimensions (Haddon Jr, 1980; J.-j. Wang & Chen, 2012a, 2012b). These five dimensions are not completely autonomous of each other, and each dimensions is impacted by numerous components and indicators(Al Haji, 2005; Rumar, 1999; J.-j. Wang & Chen, 2012a, 2012b). Choosing an arrangement of safety performance indicators, which should fill in as strong tool of benchmarking for policymakers(Wegman et al., 2008), is an unpredictable issue. The decision of every indicator is pivotal and primarily relies upon the sort, accessibility, and nature of information/data being gathered(Papadimitriou, Yannis, Bijleveld, & Cardoso, 2013). In addition, the chosen indicator should be able to be fit for joining in a composite index, which can consolidate all the significant parameters in a compact and comprehensive way (Wegman et al., 2008). Moreover, the composite index should be designed to be too adjusted as acceptable and balanced (Al Haji, 2005).So researchers applied the method of risk calculation having basic concept as shown in eq.1(Shah et al., 2018)

$$Risk = \frac{Road\ Safety\ Outcome}{Exposure} \tag{1}$$

With the increase of thoughtful indicators, number of indicators have increased to produce a composite road safety index (Al Haji, 2005; ETSC, 2001; Hakkert, Gitelman, & Vis, 2007; Wegman et al., 2008). Researchers have explained the criteria of building a complete composite road safety index to replace the conventions methods of measuring road safety performance (Wegman et al., 2008).

2.2. Relationship between Safety and Road Traffic Features

The Common way to deal with traffic safety investigation(Abdel-Aty & Pande, 2007) has been to set up connections between the activity attributes (e.g., flow, speed), roadway and natural conditions (e.g., geometry of the expressway etc), and crash event (Abdel-Aty & Pande, 2005). The inadequacy of the vast majority of the models created utilizing this approach is that they depend upon total measures of movement speed (e.g., speed limit, travel speed) and volume (e.g., AADT or hourly volumes) and thus are most certainly not adequate to distinguish the continuous blackspots (i.e., areas having a high likelihood of accidents), made because of the cooperation of encompassing traffic conditions with the geometric attributes of interstate portions (Abdel-Aty & Pande, 2005). From previous studies it is evident that traffic engineers have always focused on studying relationship of road crashes (accidents and fatalities) (Dadashova et al., 2016; Golob et al., 2008; Golob et al., 2004; Imprialou et al., 2016; Lord, Manar, et al., 2005; Ogwueleka et al., 2014; Pande & Abdel-Aty, 2006; C. Wang et al., 2009; Yasin Çodur & Tortum, 2015) in context of volume/capacity (Golob & Recker, 2004; Lord, Manar, et al., 2005; Pande & Abdel-Aty, 2006; Zhou & Sisiopiku, 1997), vehicles miles travelled (Abdel-Aty et al., 2013; Dadashova et al., 2016; Frantzeskakis & Iordanis, 1987; Jovanis & Chang, 1986; Y. Wang & Kockelman, 2013), vehicles hours travelled (Martin, 2002; Yeo et al., 2013; Zhou & Sisiopiku, 1997), Speed (average travel speed, posted speed etc.) (Aljanahi et al., 1999; Elvik, 1997; Elvik et al., 2004; Garber & Ehrhart, 2000; Golob & Recker, 2004; Imprialou et al., 2016; Kononov et al., 2011; Malyshkina & Mannering, 2008; Pande & Abdel-Aty, 2006), flow (vehicles/hour) (Aljanahi et al., 1999; Chimba & Kutela, 2014; Garber & Ehrhart, 2000; Golob et al., 2008; Golob & Recker, 2003; Golob et al., 2004; Kononov et al., 2011) and geometric design (tangent, horizontal, vertical curves) (Fu et al., 2011; Garber & Ehrhart, 2000; Karlaftis & Golias, 2002; Milton & Mannering, 1996; Noland & Oh, 2004).In this investigation, identification of high risk segments for traffic crashes (i.e., distinguishing areas with high constant crash potential) has been focused. Different Data sources were adopted during traffic data and safety analysis such as provisional road administrations (Wegman, 2014) National Highway Traffic Safety Administration (NHTSA) Applying circle information to anticipate crashes in ongoing is still in preparatory stages. The idea of crash antecedents what's more, guessed that the probability of a crash is fundamentally influenced by the transient turbulence of movement stream(C. Lee, Saccomanno, & Hellinga, 2002). They thought of elements like speed variety along the length of the roadway (i.e., distinction between the speeds upstream and downstream of the crash area) and furthermore over the three paths at the crash area(C. Lee et al., 2002).

2.3. Road Accident Data Analysis Models

Previously, different statistical techniques were utilized for accident data analysis. Those techniques were Multiple Linear Regression (FB Mustakim & Busu, 2007; Fajaruddin Mustakim & Fujita, 2011), Poisson regression (Joshua & Garber, 1990; Lord, Washington, & Ivan, 2005; Y. Wang & Kockelman, 2013), Binary Logistic Regression (Al-Ghamdi, 2002; Dissanayake & Lu, 2002; Kim, Lee, Washington, & Choi, 2007; Sze & Wong, 2007; H.

Wang, Li, Chen, & Ni, 2011), Multinomial Regression (Shankar & Mannering, 1996; Ye & Lord, 2011), Tobit (Anastasopoulos, 2016; Anastasopoulos, Mannering, Shankar, & Haddock, 2012; Anastasopoulos, Shankar, Haddock, & Mannering, 2012; Anastasopoulos, Tarko, & Mannering, 2008), negative binomial (NB) regression (Chang, 2005; Chang & Chen, 2005; Ladron de Guevara, Washington, & Oh, 2004; Lord & Mannering, 2010), probit models (Kockelman & Kweon, 2002; Ye & Lord, 2011), Sensitivity Analysis (Geurts, Wets, Brijs, & Vanhoof, 2004; J.-j. Wang & Chen, 2012b), Analysis of Variance, Zero-inflated Poisson (Aguero-Valverde, 2013; Anastasopoulos, 2016; Dong, Clarke, Yan, Khattak, & Huang, 2014; Lord, Washington, et al., 2005) and negative binomial (Chang, 2005; Chang & Chen, 2005; Dong et al., 2014; Ladron de Guevara et al., 2004).These models were superior over one another but the problem was with multiple inputs and multiple outputs i.e. in case of multiple dependent and independent variables it was difficult to rank and compare the performance of decision making units(DMUs).For example during the analysis of a Highway, it was difficult to select the worst and best segment in the presence of large data set. However, after advancement in safety research, during the European Safety Net Project (2005) (SafetyNet, 2005) on road safety performance indicators, researchers discussed the concept of benchmarking and have identified different domains and indicators of road safety. The concept of benchmarking was focused on a composite road safety index to identify the one value based road safety performance evaluation (Al Haji, 2005). For benchmarking and analysis, following five weighted method techniques have been applied and received a lot of importance (Hermans, Brijs, Wets, & Vanhoof, 2009; Hermans, Van den Bossche, & Wets, 2008) i.e. factor analysis, analytic hierarchy process, budget allocation, data envelopment analysis and equal weighting (Hermans et al., 2008). Out of these five considered methods, data envelopment analysis is the most different technique and provided a better understanding for the road safety performance index for ranking purpose (Hermans et al., 2009; Hermans et al., 2008).

2.4 DEA for Accident Data Analysis

As per latest European Commission report of 2016, road safety performance analysis of motorways is necessary for the safety of travelers. In last decade, studies have revolved around the application of DEA for benchmarking and safety analysis. Road safety performance measurement is an important part of safety analysis and operations research. "DEA is a data-oriented method for measuring and benchmarking the relative performance of peer decision-making units (DMUs) with multiple inputs and multiple outputs" (Alinezhad, 2016). DEA was initiated in 1978 when Charnes et al. (Charnes, Cooper, & Rhodes, 1978) established how to convert a fractional linear measure of efficiency(or risk) into a linear programming format. This non-parametric approach solves an linear programming (LP) origination per DMU and the weights allocated to each DMU are the results of the corresponding LP (Alinezhad, 2016; Amirteimoori & Kordrostami, 2012). DEA has been successfully applied in road safety research as for the analysis of European countries, concept of composite performance indicator was used and DEA was used for weights determination and ranking. Road fatalities were tested on the basis of fatalities per million inhabitants (Hermans et al., 2008). To enhance the output of risk value, inputs like population, passenger kilometers and a number of registered cars were used against the output of a number of fatalities but cluster analysis technique was also added to analyze the performance of similar European countries (Shen, Hermans, Brijs, Wets, & Vanhoof, 2012). At municipality level in Israel, for the allocation of annual budget used DEA without puts like a number of drivers involved in crashes, number of seatbelt tickets issued per year, the percentage of cars and the average age of private cars were included (Alper, Sinuany-Stern, & Shinar, 2015). Similarly, for road safety performance analysis, member states of Brazil by using DEA, mortality rate and fatality rate (fatality per vehicle & fatality per vehicle kilometer traveled) were used as output variables (Bastos et al., 2015). For road safety analysis of Serbia, DEA method was applied to data for 27 police departments for traffic and public risk evaluation (Rosić, Pešić, Kukić, Antić, & Božović, 2017) and road safety risk for motorways in Belgium (Shah et al., 2017). So from the discussion about DEA application in the field of Transportation Engineering, it confirms that DEA is one of the established techniques to evaluate the severity level of road safety (Shah et al., 2017).

2.5. DT for Accident Data Analysis

The Decision Trees as a data mining technique allows DMUs to discover significant information that had previously been hidden in large databases. It has been developed in 1984 by Breiman (Breiman, Friedman, Stone, & Olshen, 1984) and improved in 1996 by Ripely (Miller & Ripley, 1996). The problem is illustrated by a decision making tree so that each non-leaf node is associated with one of the decisions making variables, each branch of a non-leaf node is associated with a subset of the decision making variables values, and each leaf node is linked to a target variable (the dependent variable) value(Rahimi, Behmanesh, & Yusuff, 2013). Each leaf is associated with a target variable's mean value; therefore, this tree can be an alternative to continuous linear models for solving the problems of regression analyses of classified data (Clark, 1992). Researchers have vastly used this reliable technique in road safety field as found relationship significance of crashes and factors like geometric pattern (Karlaftis & Golias, 2002), collision order in Korea (Sohn & Lee, 2003), geometric factors (Chang & Chen, 2005) causality class (Kashani & Mohaymany, 2011) crash seriousness (Abellán, López, & De OñA, 2013), geometric factors along with number of vehicles (Chang & Chien, 2013) and over speeding (Olutayo & Eludire, 2014). From the analysis, we can summarize that DT was previously used as an road safet analysis tool, which is a useful technique to study the impact of road related features, geometry, and other contributing factors on road safety.

2.6. DEA-DT Approach for Accident Data Analysis

Combination of DEA and DT has not been used in road safety field; however, it is popular in other fields like banks and corporate sectors. From the previous studies it is concluded that DEA is suitable tool for risk calculation but for decision-making process regarding contributing factors, decision tree is ahead, so a discussion started after (Athanassopoulos & Curram, 1996) why not to combine DEA with a data mining technique to get best possible outputs, i.e., risk evaluation for elaborating levels of severity and then risk prediction for factor impact analysis purpose (Shah et al., 2017). Researchers are using this combination of DEA and DT to evaluate risk and efficiency in different fields like banking sectors (Alinezhad, 2016), technology commercialization projects (Sohn & Moon, 2004) service units of firm(Seol, Choi, Park, & Park, 2007) Arabic banks (Samoilenko & Osei-Bryson, 2013) hybrid supplier evaluation (Wu, 2009) telecommunication service (Samoilenko & Osei-Bryson, 2008) financial firms (S. Lee, 2010) and Insurance branches (Shafaghizadeh & Attari, 2014).The successful applications of this DEA-DT joint approach imply that it can be a useful technique for road safety data analysis.

3. Materials and Methods

3.1 Study Area and Data Description

Belgium is located in the heart of European Schengen areas. From European Commission reports (CARE, 2016; EU, 2016), the problematic sector identified for analysis is based on these reports as motorways are most problematic areas in Belgium with respect to safety issues. The study area selected for this study is two Motorways in Belgium named E-313(57 Km) and E-314(50 Km) (Limburg Province Sections). For the selected Motorways (E-313 & 314), 67 segments were having at least one crashes in three years. So data has been derived from the Flemish Ministry of Mobility and Public Works and FEATHERS model (Janssens, Wets, Timmermans, & Arentze, 2007) for these motorways as shown in Fig.3.

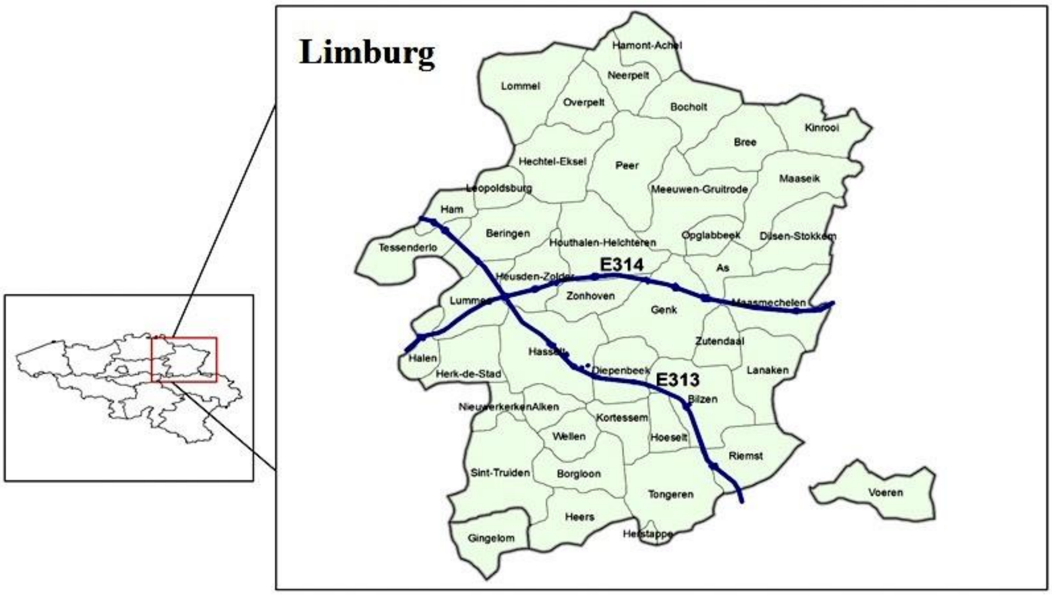

Figure 3. Study Area of Limburg

Table 1. Data Description for Motorways of Belgium Limburg Section

Road	Variable	Units	N	Mean	SD	Min.	Med.	Max.
	Traffic Flow	Veh/hr	36	1116.4	414.2	44.8	1272.7	1483.4
	Travel Speed	Km/hr	36	107.82	7.34	99.51	104.86	120
	Horz_Curve	1-Yes,0-No	36	0.7778	0.4216	0	1	1
	Vert_Curve	1-Up,2-Down,3-No	36	1.917	0.732	1	2	3
	V/C	Volume/Capacity	36	0.5131	0.1442	0.139	0.581	0.62
E-314	VMT	Veh.Mile Travel	36	1718	1154	77	1567	4366
	VHT	Veh.Hr Travel	36	1061	737	38	1014	2986
	Segment Length	Km	36	1.068	0.736	0.09	0.955	3.35
	NoA	Numb. of Accidents	36	10.92	11.71	1	5	49
	NoAP	Numb. of Affected Persons (Inj/Killed)	36	16.56	18.09	1	8	76
	Traffic Flow	Veh/hr	31	795.9	433.1	31.4	794.4	1330
	Travel Speed	Km/hr	31	114.66	7.76	96.89	119.44	120
	Horz_Curve	1-Yes,0-No	31	0.8387	0.3739	0	1	1
	Vert_Curve	1-Up,2-Down,3-No	31	2.032	0.706	1	2	3
	V/C	Vol/Capacity	31	0.3562	0.1842	0.08	0.327	0.643
E-313	VMT	Veh.Mile Travel	31	1955	1630	83	1668	5186
	VHT	Veh.Hr Travel	31	1129	1031	41	954	3616
	Segment Length	Km	31	1.518	0.965	0.22	1.63	3.77
	NoA	Numb. of Accidents	31	8.03	14.64	1	3	74
	NoAP	Numb. of Affected Persons (Inj/Killed)	31	11.81	21.13	1	6	105

3.2 Framework for Decision Making Process

To evaluate the road safety condition of roads a two-stage performance analysis concept has been introduced as shown in Fig.4. This method consists of two sections; first DEA model portion to evaluate the severity level of accident locations and the second the decision tree mechanism to evaluate the impact of the factors on the severity. It also helps decision makers (engineers) to locate the location and identify attribute to be improved during decision-making and evaluation. Process of application of DEA in collaboration with DT model to calculate the Risk level of road segments can be explained as.

Step.1: Selection of the Problematic Motorways or Highways (Limburg-Motorways E-313 & E314)

Step.2: Segmentation of the Selected Motorways according to the data network

Step.3: Selection of the Problematic Section having at least one Accident in three years

Step.4: Selection of Parameters from Accident data (No. of accidents & No. of persons injured or killed), traffic features data (VMT, VHT,V/C, Speed and Flow) and Road Feature data (Horizontal Curve, Vertical Curve) with reference to domain of Transportation Engineers.

Step.5: Data was distributed and applied into Two Section:

Section. I: DEA Models to identify risk level and rank priority wise most risky road segments with reference to road safety. Following the basic concept of risk

$$Risk = \frac{Road\ Safety\ Outcome}{Exposure} \tag{1}$$

While calculating road safety risk, road safety outcome is number of accidents and number of persons injured or killed (NoA and NoAP) while exposure variables are related to traffic (i.e. VMT, V/C and VHT).

Step.6: After calculating risk value using linear program with the help of Lingo software, some values were same i.e. 1 and it was difficult to rank them. Therefore, a cross-efficient approach was applied which follows the concept of using the weights of each DMU by multiplying it the values variables of each DMU to calculate a unique value called as cross efficiency method and value obtained is named as Risk CE. Ranking helps in prioritizing the risky segments to set the priority of risky segments.

Step.7: Development of GIS maps to identify and facilitate the Transportation Engineers for decision-making procedure.

Step.8: DT Models to identify the factors influencing Risk level (identified by DEA Model) of each segments to mitigate and control contributing factors to control and improve the road safety situation.

Step.9: Rearrangement of Data for application of DT model (i.e.; To normalize the data of Risk value nature log has been applied so named as NLog Risk and it was considered as dependent variables while Speed, Flow, Horizontal Curve and Vertical Curve were considered as independent variables)

Step.10: A comparative performance of DT in comparison with MLR (Multiple Linear Regression) was analyzed by comparing R^2 (Coefficient of Determination) and RMSE (Root mean squared Error) value.

Step.9: Analysis of factors influencing the Risk value of road segments as a single unit and in combination of other factors.

Step.10: Discussion on obtained results as policy makers and transportation engineers.

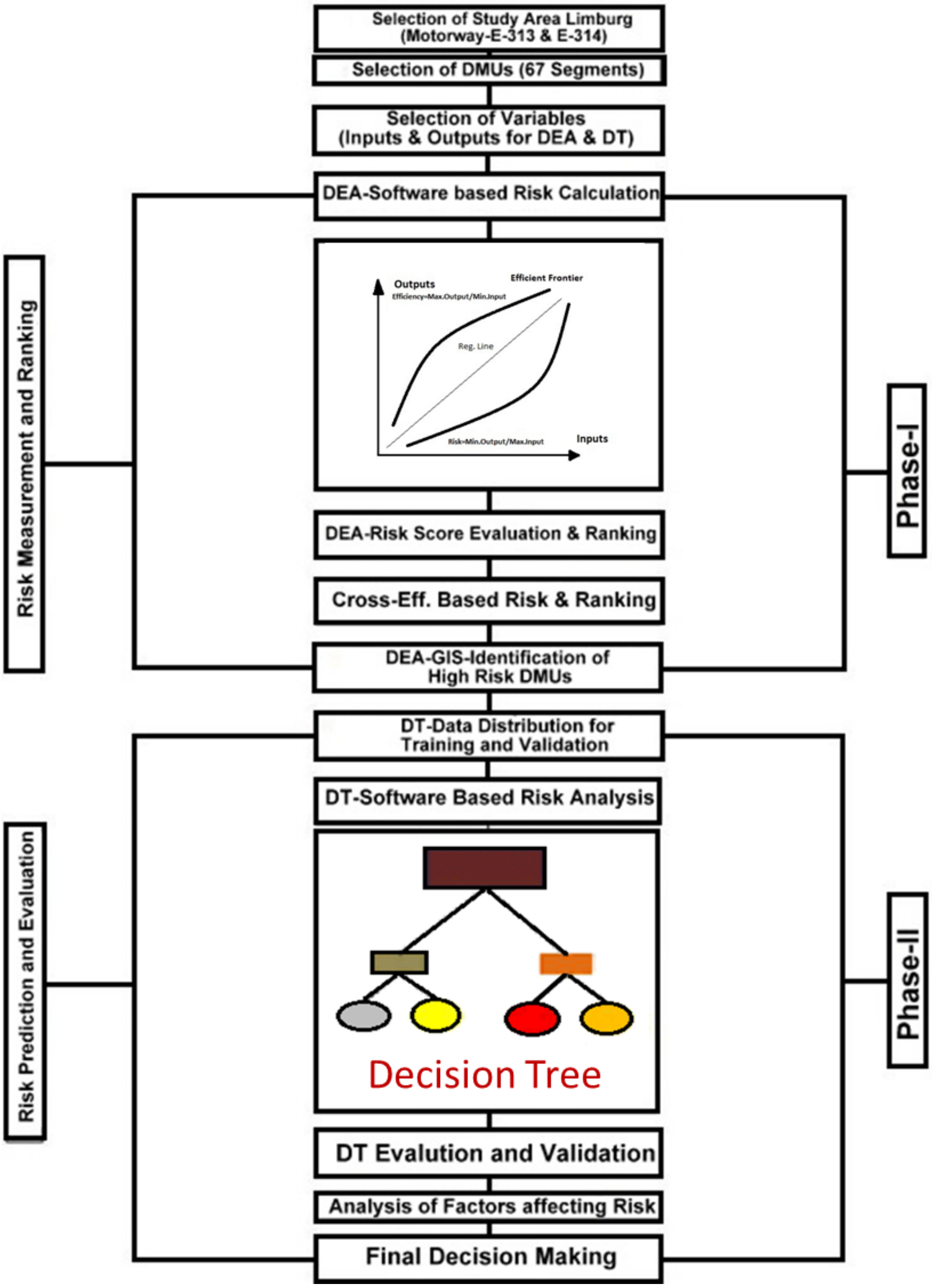

Figure 4. DEA-DT Framework for Accident Risk Evaluation

4. Analysis and Discussion

4.1 Section-I: Decision Making regarding Road Sections.

During the analysis phase-I, four major steps are conducted to evaluate road safety risk value i.e.(i) Selection of Input and Output variables (ii) Application of DEA model for calculation of calculation of Risk (iii) Calculation of risk using cross efficient matrix (called Risk CE) for ranking of risk value for prioritization of risky segments and (iv) Normalization of risk value by taking Natural Log (called NLog Risk) to be used in nest phase. So, road traffic crashes and fatalities have been taken as output while exposure variables have been taken as input (Shen et al., 2012). Researchers have explained DEA as "Consider an n-DMUs set, each consuming m different inputs to produce s different outputs. The relative efficiency of a DMU is defined as the ratio of its total weighted output to its total weighted input, subjected to lie between zero and the unity. Mathematically, the efficiency score of a particular DMU0, i.e., E0, is obtained by solving the following constrained optimization problem(Charnes et al., 1978; Shen et al., 2012):

$$\begin{aligned} E_0 &= \max \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \\ \textit{subject to} \quad & \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad j = 1, \cdots, n \\ & u_r, v_i \ge 0, \quad r = 1, \cdots, s, \quad i = 1, \cdots, m \end{aligned} \tag{1}$$

where yrj and xij are the rth output and ith input respectively of the jth country, ur is the weight given to output r, and vi is the weight given to input i."

The concept of DEA model has been shown in Figure 5. "It is a linear programming technique for measuring the relative performance of entities or units of a similar pattern. Highway sections are considered to be Decision-Making Units (DMUs) for the application of the DEA model, and the risk level is calculated by applying the road safety outcome and exposure variables"(Shah et al., 2018).

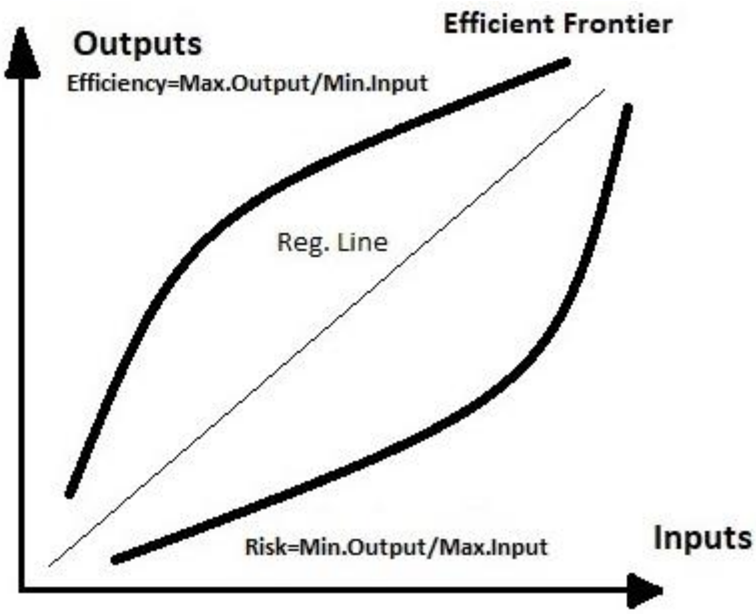

Figure 5. DEA Conceptual Model

During the severity level risk analysis, it can be assumed that "the lowest level has been considered as the frontier of safety. In the case of efficiency evaluation, Efficiency is calculated by maximizing output and minimizing input, while, for calculating risk, we minimize the output and maximize the input(Shah et al., 2018). The simplest form of calculating Efficiency by DEA is as follows"(Shah et al., 2018).

Efficiency: The basic concept of the DEA Efficiency calculation is as in Equation (3)(Shah et al., 2018).

$$Efficiency = \frac{Weighted\ Sum\ of\ Output}{Weighted\ Sum\ of\ Input} = \frac{Maximize\ Output}{Minimize\ Input} \quad (2)$$

Following the above-explained concept, Risk is calculated as in Equation (4):

$$\text{Risk} = \frac{\text{Weighted Sum of Output}}{\text{Weighted Sum of Input}} = \frac{\text{Minimize Output}}{\text{Maximize Input}} = \frac{\text{Road Safety Outcome}}{\text{Exposure}}. \quad (3)$$

"DEA model is applied through Lingo software (programming-based), which provides the severity values for each segment of road (i.e.; Decision-Making Unit-DMU). Researchers in the field of road safety introduced DEA by assigning weights for the construction of composite performance indicators; then, for the evaluation of road safety rankings, a risk value was calculated"(Shah et al., 2018).

DEA holds an smart feature that each DMU is permitted to select its own most promising input and output weights, or multipliers for evaluating its best efficiency, rather than the same weights for all the DMUs (Dyson & Thanassoulis, 1988; Wong & Beasley, 1990). Sometime due to similar efficiency or risk value, ranking procedure is difficult to set priorities of DMUs. To overcome these difficulties, a cross-efficiency method (Sexton, Silkman, & Hogan, 1986) was developed as a DEA extension tool that can be used to identify the best overall performers and to effectively rank all DMUs. The main theme of this concept is use weights of each DMU to evaluate efficiency and a sum is used to get the best possible value. The conceptual demonstration can be seen in cross-efficiency matrix (CEM) as shown in Table 3. In the CEM, the element in the ith row and jth column represents the efficiency scores of DMU j using the optimal weights of DMU i. The basic DEA efficiencies are thus located in the leading diagonal. Each column of the CEM is then averaged to obtain a mean cross-efficiency score for each DMU. This process can help in selecting highest to lowest efficiency for ranking purpose(Boussofiane, Dyson, & Thanassoulis, 1991). Therefore, this process can be considered as a kind of sensitivity analysis since different sets of weights are applied to each unit, and they are all internally derived rather than externally imposed. This concept has been explained as it is by(Shen et al., 2012).

Table 3 A generalized cross-efficiency matrix (CEM)

Rating DMU	Rated DMU				
	1	2	3	…	n
1	E_{11}	E_{12}	E_{13}	…	E_{1n}
2	E_{21}	E_{22}	E_{23}	…	E_{2n}
3	E_{31}	E_{32}	E_{33}	…	E_{3n}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	E_{n1}	E_{n2}	E_{n3}	…	E_{nn}
Mean	$\overline{E}_1$	$\overline{E}_2$	$\overline{E}_3$	…	$\overline{E}_n$

In this study, the concept of identification of risk was elaborated as road safety outcome which was basically associated with road accidents and causalities. However, the exposure was defined as Volume/Capacity, Vehicle Miles Travelled and Vehicle Hourly Travelled

(Shah et al., 2017) as shown in Table.1. It is one of the familiar concepts by the researchers based on Model (1), a segment with a value of up to 1 is considered as the safest one while the Highway segment with the maximum value is considered as the most dangerous one (Shah et al., 2017). Moreover, by using the weights calculated for each factor are used to calculate risk by the cross risk method (Shen et al., 2012) so that all the DMUs can be made comparable.

Table 2. Application of DEA Model for Accident Analysis of Motorway Segments

DMU	Input 1	Input 2	Input 3	Output 1	Output 2	DEA	DEA CE	NLog_Risk	RANK
Seg_ID	V/C	VMT	VHT	RA	NoAP	Risk	Risk		
1	0.369	3039.22	1541.61	74	105	60.79	91.07	4.51	1
29	0.139	109.17	54.58	6	8	25.72	71.73	4.27	2
19	0.603	183.73	118.16	12	20	11.95	69.92	4.25	3
2	0.384	2494.33	1268.38	49	76	51.90	65.10	4.18	4
34	0.080	82.51	41.26	3	6	22.62	62.90	4.14	5
5	0.278	2190.90	1096.68	38	50	41.09	62.28	4.13	6
25	0.139	76.74	38.37	3	6	13.02	58.11	4.06	7
26	0.237	202.31	101.24	9	11	22.58	57.15	4.05	8
3	0.361	2683.94	1361.27	40	61	39.16	53.26	3.98	9
21	0.534	594.78	336.96	13	24	14.65	35.40	3.57	10
-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
27	0.499	4147.38	2297.64	4	6	2.50	3.61	1.28	51
33	0.486	3158.98	1742.42	3	7	3.05	3.50	1.25	52
36	0.585	5013.80	3103.21	4	8	2.30	3.19	1.16	53
59	0.337	1882.65	954.05	2	2	2.02	2.92	1.07	54
65	0.439	441.58	226.31	1	1	1.35	2.79	1.03	55
55	0.602	1711.50	1152.94	2	3	1.91	2.34	0.85	56
52	0.595	3400.89	2256.01	3	3	1.46	2.33	0.85	57
47	0.603	1435.26	944.89	2	2	1.94	2.26	0.82	58
51	0.465	618.77	328.33	1	1	1.27	2.17	0.77	59
61	0.575	808.56	517.91	1	2	1.05	2.16	0.77	60
60	0.555	2423.61	1428.66	2	2	1.52	1.90	0.64	61
48	0.486	3956.90	2182.53	2	2	1.00	1.73	0.55	62
46	0.603	1159.82	763.55	1	2	1.00	1.66	0.51	63
53	0.631	4275.09	3046.69	2	3	1.00	1.47	0.39	64
67	0.592	1093.97	734.83	1	1	1.00	1.31	0.27	65
49	0.499	3214.27	1780.70	1	2	1.00	1.08	0.08	66
66	0.575	1714.22	1098.00	1	1	1.00	1.07	0.07	67

Note: DEA Risk: Risk value calculated by Data Envelopment Analysis(DEA) Method

DEA Risk CE: Cross Efficient Risk calculated after application of DEA to get the unique Risk value for ranking

Ln(Risk)-Natural Log of Risk value-To Normalize the Data

GIS mapping is one of the advanced tools to use to visualize the concept of the analysis and it provides an opportunity to evaluate and analyze the big data in layers. The joining of risk value and severity calculation by DEA model can be visualized by GIS(Shah et al., 2017). Previously for the same section analysis has been done to visualize the risk as shown in Fig.6. Risk value evaluated through DEA has been visualized in combination with GIS in previous researches (Shah et al., 2017) a step further , the decision tree mechanism has been employed to get a view that these results can be focused on the basis of tree mechanism and a chain of steps can be decided to evaluate the severity and risk level for certain section of highways.

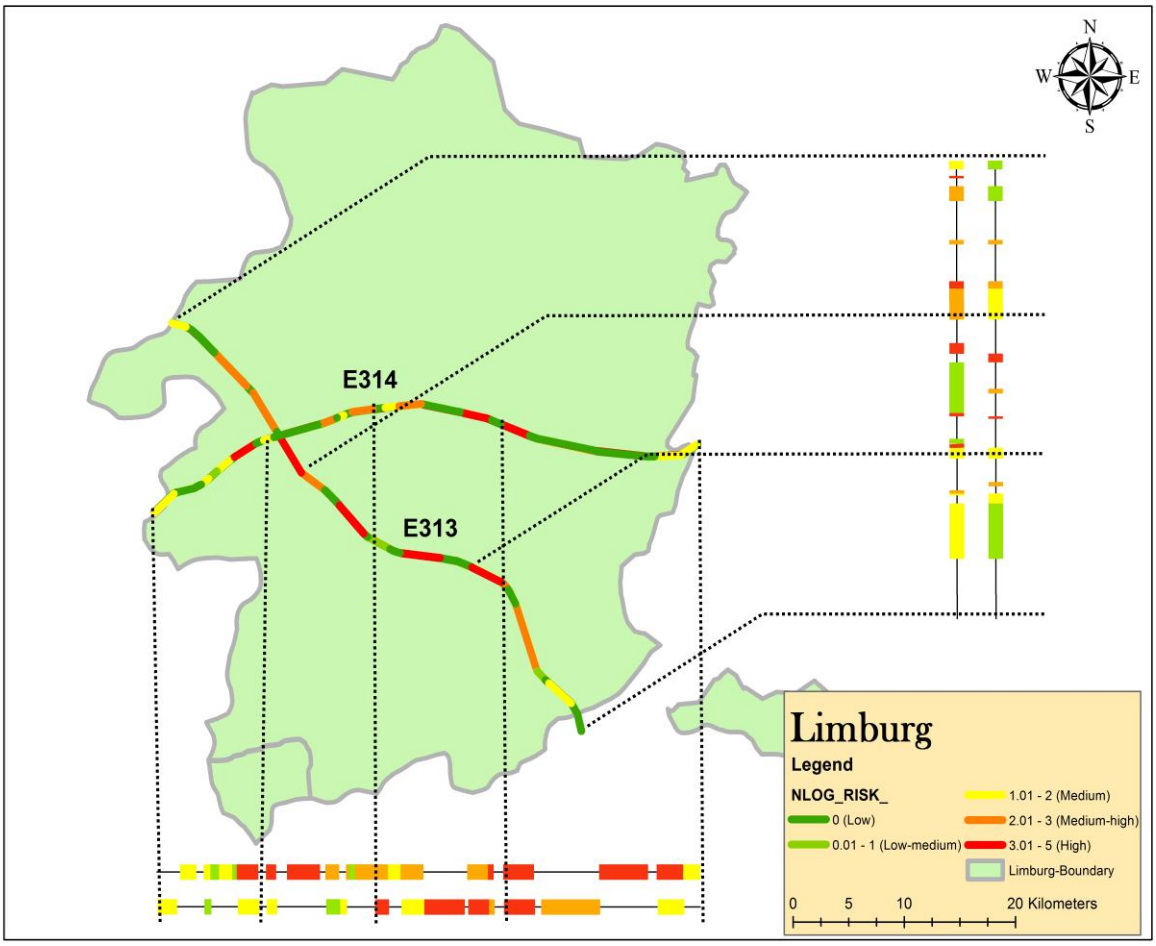

Figure 6. Accident Severity Risk Map for Motorway Segments

4.2 Section-2: Decision Making regarding Factors

Although the identification of worst section on the basis of safety issues is one of the key tasks during safety audit the safety can be ensured by improving the factors contributing to a safety problem. Decision Makers/Transportation Engineers need to know the basic contributing factors. Decision tree technique has been applied to analyze the impact of contributing factors (Description of factors can be observed in Table 2.).

Table 3. Descriptive Statistics for Phase-II Analysis

Variables	Description	Mean	S.D	Min.	Max.
Ln(Risk)	Risk-Severity Value of the section(DEA based)	2.212	1.220	0.068	4.512
Flow	Average Annual Daily Traffic on each Segment-vph	968.1	449.6	31.5	1483.4
Speed	Average Travel Speed for each Segment-kph	110.99	8.23	96.89	120
Horz_Curve	1-Presence of Curve,0-Tangent			0	1
Vert_Curve	1-Curve Upward 2-Curve Downward 3-Flat			1	3

Note: Ln(Risk)-Natural Log of Risk value-To Normalize the Data

Traffic Flow characteristics and geometric design feature are of major factors to be dealt by engineers to improve the safety condition of a highways or road. So these four factors have been considered during the decision making process. These factors can be improved and budget can be allocated for the improvement of these factors during the budget distribution of Highway agency and Federal government. In choice tree displaying, an exact tree speaks to a division of the information that is made by applying a progression of straightforward guidelines. In the creation of a decision tree, data set is usually divided into two parts: the training data set and the test dataset (Hand, Mannila, & Smyth, 2001; Jiawei & Kamber, 2001).

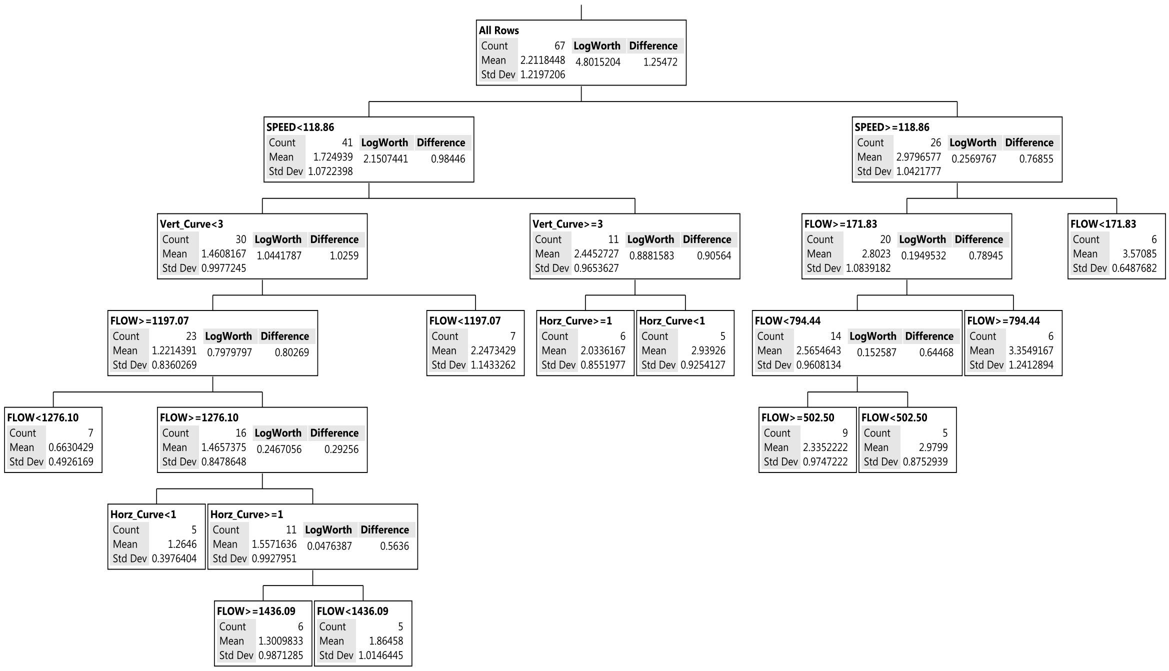

Figure 7. Road Safety Risk Prediction through Decision Tree Model

Then they undergo two main processing phases of growth and pruning. In the development stage, a decision tree is constructed from a set of training data. In this phase, each leaf node is associated with a class. A noteworthy favorable position of the choice tree over other displaying systems is that it delivers a model, which may speak to interpretable principles or rational explanations. After calculation of Risk evaluation, DT technique was applied to check which are the factors influencing the risk factor of DMUs as shown in Fig.7. The performance of a model is predicted by using the basic paramters of statistices.These parmeters are root mean squared error (RMSE) which is difference between the actual and the predicted values and coefficient of determination ($R^{2)}$(Siddique, Aggarwal, & Aggarwal, 2011).

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{n=1}}^{\mathrm{N}}(\mathrm{actual} - \mathrm{predicted})^2\right)}\text{-----------------}(1)$$

$$R^2 = 1 - \frac{SSE}{SS_y}\text{---}(2)$$

Where SSE is sum of squared errors of prediction and SSy is total variation. Mean absolute error is similar to root mean square except using absolute difference instead of squared difference. Usually performance of a model is compared by coefficient of determination(R^2). A classic fit would bring about a R2 of 1, and poor fit almost 0 while RMSE should be as minimum as possible (Siddique et al., 2011).

The proposed decision tree in this study includes four main components: The first component is the output (dependent) variable. Based on the independent (predictive) variables, this variable is used to predict. This tree diagram is showing the speed is a major factor affecting Risk evaluation of each DMU as shown in Table 2. Thus as a traffic Engineer one can analyze that which factor is most important and can be controlled by changing speed limit of those segments, one can easily make a safety zone with this low-cost solution.

Table 4. Parametric Estimates of Decision Tree Model

Impact of Factors (Independent Variables)			
Term	**No. of Splits**	**SS**	**Portion**
SPEED	1	25.04	0.4844
FLOW	6	16.33	0.3158
Vert_Curve	1	7.80	0.1508
Horz_Curve	2	2.53	0.0489
Cross-Validation		Value	
R^2	Actual	0.52	
	K-Folded	0.32	
RMSE		0.83	

Model Performance: Comparative Analysis of DT Vs MLR

Model	R^2 (Main) Prd.	R^2(K-Fold) Val.	RMSE	Remarks
DT	0.52	0.31	0.83	Better
MLR	0.27	0.14	1.07	Model

A leaf table report as shown in table.3 has been developed to show the possible combinations are associated which high-risk level as shown in in leaf 5,6,7,8,9,10 and 11 are associated with high-risk levels. With the help of this technique, a targeted treatment can opt for all road segments.

Table 4. Leaf Table showing the relationship between Risk level and Factors

Leaf Label	Leaf	Mean	Coun
SPEED<118.86&Vert_Curve<3&FLOW>=1197.07&FLOW<1276.10	1	0.66304	7
SPEED<118.86&Vert_Curve<3&FLOW>=1197.07&FLOW>=1276.10&Horz_Cur	2	1.2646	5
SPEED<118.86&Vert_Curve<3&FLOW>=1197.07&FLOW>=1276.10&Horz_Cur	3	1.30098	6
SPEED<118.86&Vert_Curve<3&FLOW>=1197.07&FLOW>=1276.10&Horz_Cur	4	1.86458	5
SPEED<118.86&Vert_Curve<3&FLOW<1197.07	5	2.24734	7
SPEED<118.86&Vert_Curve>=3&Horz_Curve>=1	6	2.03361	6
SPEED<118.86&Vert_Curve>=3&Horz_Curve<1	7	2.93926	5
SPEED>=118.86&FLOW>=171.83&FLOW<794.44&FLOW>=502.50	8	2.33522	9
SPEED>=118.86&FLOW>=171.83&FLOW<794.44&FLOW<502.50	9	2.9799	5
SPEED>=118.86&FLOW>=171.83&FLOW>=794.44	10	3.35491	6
SPEED>=118.86&FLOW<171.83	11	3.57085	6

From above-explained table 3 and 4 and figure 7 we can have a conclusion that Decision Tree model is presenting required results as a factor which is influencing are Speed and Flow. Thus in the decision-making process for an Engineer, it is a useful method to identify the major effecting factor (i.e. Speed factor). The model performance and validation (K-folded method) has been tested through the basic tests like R^2(0.5) and RMSE(0.84), which show the predictive strength of applied DT model in comparison to MLR as shown in table.3 and fig 8.

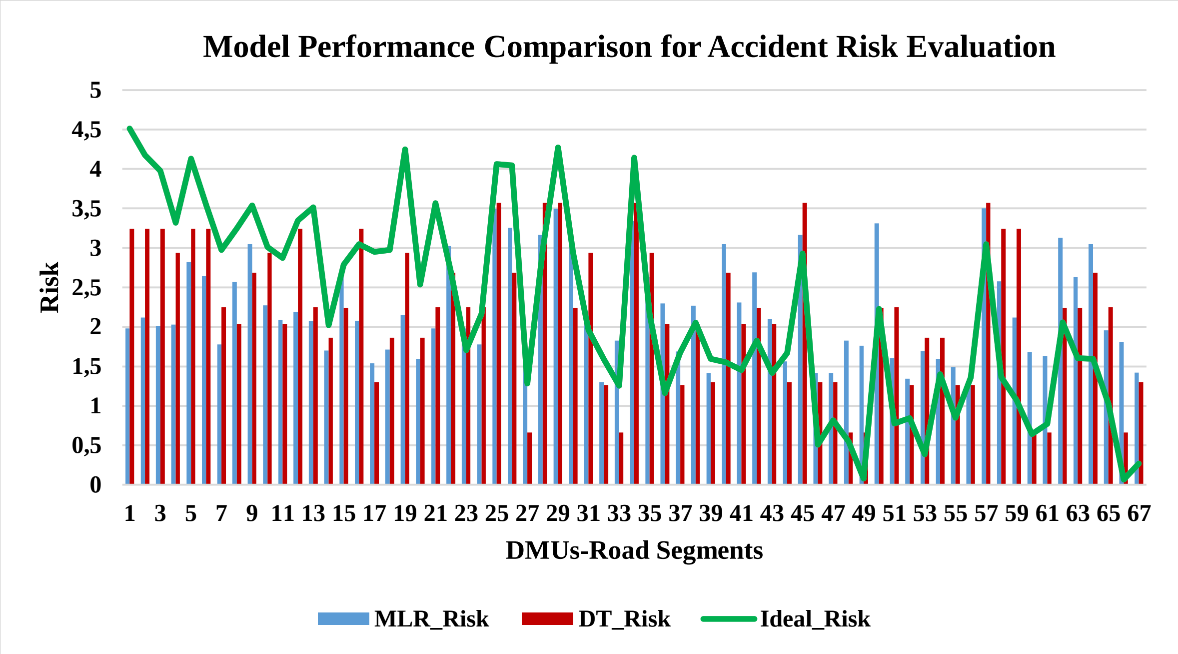

Figure 8. Performance Analysis of Risk Prediction through DT and MLR Model

4.3 Analysis of Factors

GIS-based mapping procedure guides decision makers up to the identification of the problematic areas but still, the improvement factors were of major concern. After effectively applying the DEA-DT approach for motorways safety assessment, we can likewise concentrate on the contributing variables utilized as a part of the hazard expectation. Traffic Engineers first like to concentrate on contributing factors, which are of low cost. In this manner from a graphical examination of the contributing variables, we can see that dominant part of the crashes is on the risky segments of the motorways. Safety Engineers, as a rule, abstain from going for an infrastructural change in light of the fact that updating and recreation is an expensive strategy, so in the event that they concentrate on the ease treatment, can go for Speed and Flow control. Fig. 9 shows the connection between the hazard and the diverse contributing components.

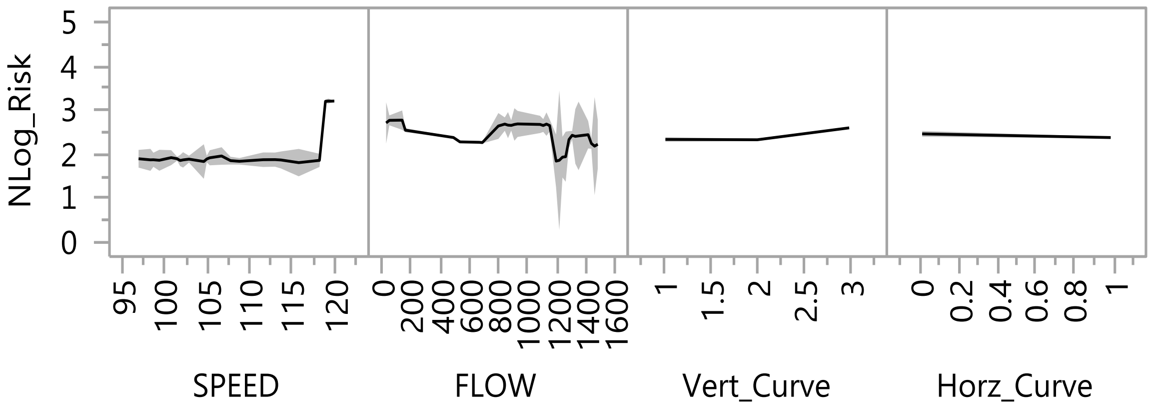

Figure 9. Impact Analysis of Factors on Risk-Severity

There are four major independent variables, which are influencing risk value of motorway segments:

4.3.1 Speed

Speed is considered as one of the most important factors influencing road safety. One can easily understanding even according to the principles of physics that if some object will hit another object with higher speed than impact of that crash/collision will be severe. In the field of Highway Engineering, one of major necessity is to enhance a smooth manuring environment for road users. Especially in case of motorways, designers prefer to maintain a higher speed limit to provide road users a speedy and comfortable environment for time saving and smooth travelling. Some time, it compromises safety, as on high speed there are others factors which combination with this crucial factor causing high number of accidents. In case of Motorways (E-313 & E-314) high speed is a contributing factor, above 115Km/hr is showing a higher risk factor so reduction in Speed limit from 120km/hr to 110 km/hr can contribute in improving the safety of these motorways.

4.3.2 Flow

It is a common understanding that increasing number of users on roads will surely contribute increasing number of accidents. Flow is considered as number of vehicles/hr, which is one of the major factors because increase in population also contribute in occupying road space. The behavior variation of vehicles also increases with the increase in vehicles, which contribute in risk increment. Higher Traffic flow creates an environment of congestion, which also contributes in increasing road accidents. Some time it does not singly influence road safety level but it works in combination with other factors like speed and geometric design.

4.3.3 Horizontal Curve

Geometric Design is one of the major components of road design and a true contributor in road safety. As a transportation engineer, horizontal curve design is a key prospect to watch during the safe road design. Researchers explained that accidents on curves are three times higher than that of straight road sections (Torbic et al., 2004). Moreover, there is a serious concern by decision makers for accidents on horizontal curves (Bonneson, Pratt, Miles, & Carlson, 2007). Motor vehicles crashes occur frequently at horizontal curves (Persaud, Retting, & Lyon, 2000). Horizontal curves contribute frequently in accidents because it causes safety hazard to road users with reference to drivers prospective (W. Schneider IV, Zimmerman, Van Boxel, & Vavilikolanu, 2009). Change in design of horizontal curve reduces the sight distance prospective and effects vehicular handling capabilities (W. H. Schneider IV, Savolainen, & Moore, 2010). During the analysis of Motorways of Limburg, there was no problem seen due to horizontal curve design.

4.3.3 Vertical Curve

Second factor of geometric design is vertical curve. Sometime vertical curve also contributes to accidents. However, it is usually occurs in hilly areas and in this section of Limburg motorways there was no issue relevant to this type of design features. So motorways are safe regarding vertical curve perspective.

4.3.5 Combine Effect of Variables

Sometime a single factor do not contributes in affecting safety but a combination of factors effects the safety performance of roads. Therefore, to analyze the risk level, combine effect of different variables have been studied. NLog Risk value is natural log value of risk, which is a composite factor indicating safety level of road. Risk level has been for analysis of highly risky situation under all four contributing factors. During the analysis of Flow and Speed combine impact on road safety performance, we observe that with increase in speed and at higher flow, contribute in risk value increase. Flow in combination with Horizontal curve shows no impact or increase in risk value. Flow in combination with vertical curve shows no trend of increase in risk level. So there is no problem of geometric deign. Tree graph shows the actual location of risk under all contributing factors. We can see from Fig. 10 that the hazard level can be decreased by controlling two elements i.e. Speed and Flow.

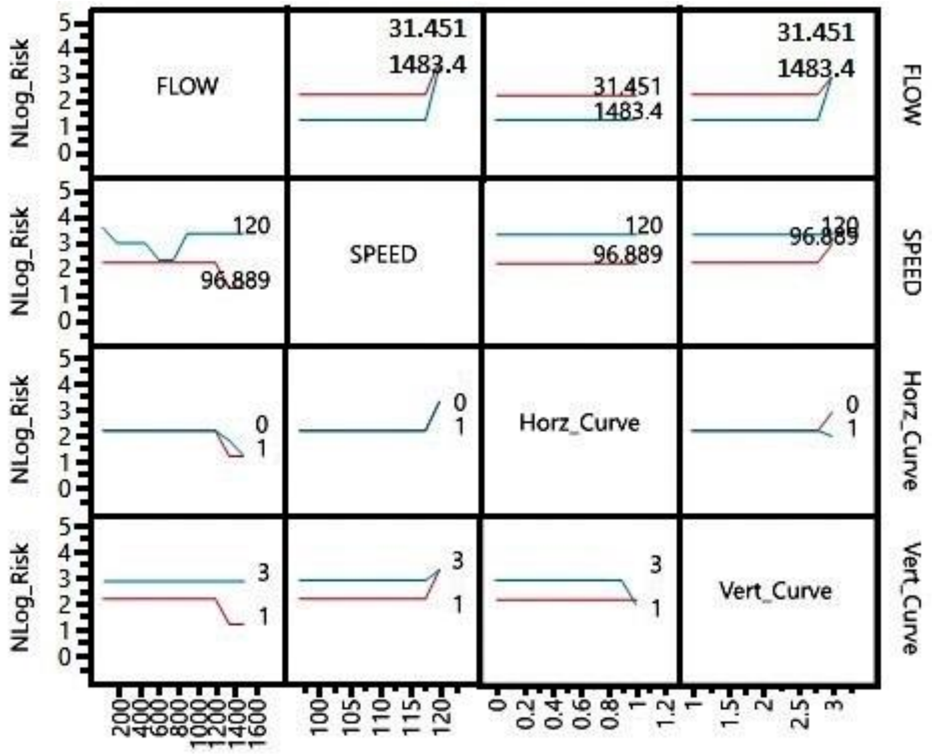

Figure 10. Tree Graph Distribution for Contributing Factor based Risk Analysis

We can see from the information that 35 out of 67 sections with fifty-two percent are over the mean speed constrain 110 km/hr in parallel to traffic flow above mean level of movement stream, i.e. So by controlling these two variables we can enhance an impressive level of safety state of the motorways with low-cost treatment. Furthermore, we can observe that if we target a higher level of safety, the majority of the accidents are on flat vertical curves, so geometric point of view is clear.

5. Conclusions

This study focuses accident analysis of road infrastructure geometry and traffic stream characteristics. To improve the estimation precision, a two-phase framework has been proposed to accomplish the hazard assessment, i.e., a benchmarking instrument of DEA in a blend with a forecast model of DT has been acquainted with the road safety research. A crash dataset separated from the Flemish Road safety is stratified in two elements: a number of accidents and no of influenced people (killed or injured)(Shah et al., 2017), and is used to display the proposed demonstrate development of DEA and DT regression. Decision trees as a technique of controlled data mining are classification method which is convenient to interpret and understand when compared to other performance models. In the first stage, DEA helps to identify the risky segments of the motorways and in second stage contributing factors have been identified by DT. According to the analysis-targeting speed limit is one of the major low-cost treatment and provides that by reducing speed limit we can reduce major risk level furthermore by controlling flow level by controlling flow with controlling access to the motorways. Presently to check the consistency of safety hazard an incentive with the assistance of DT method, following components have been considered: speed, flow, even and vertical curve. They are most imperative variables, which could be affected by the safety engineer. Generally, an infrastructural change like revision in the horizontal and vertical curve can cost much. Yet, this framework can likewise help in better arrangement on the grounds that nobody would want to change structure of the entire interstate (i.e. especially in case of 100 km long roadway), so by choosing the most dangerous area and tackling the safety issue of just those portions will likewise give a minimal effort basic decision makers. Besides, joining DEA with DT, GIS in a road safety investigation framework can additionally incorporate the usefulness of the DEA and, in the meantime, enhance the arrangement of potential utilization of GIS. The framework is amazingly adaptable and self-versatile, equipped for consolidating any change in new informational collection. Therefore, a joint approach of DEA-DT can give a simple as well as the proficient yield for data analysts for road safety accident analysis and decision making for allocating road infrastructure improvement budget.

Acknowledgments: This research is jointly supported by TITE and IMOB for publication. Authors would like to thank HE-Boong Kwon (USA), one of the pioneers of a similar method for his valuable guidance.

References

Abdel-Aty, M., Lee, J., Siddiqui, C., & Choi, K. (2013). Geographical unit based analysis in the context of transportation safety planning. *Transportation Research Part A: Policy and Practice, 49,* 62-75.

Abdel-Aty, M., & Pande, A. (2005). Identifying crash propensity using specific traffic speed conditions. *Journal of Safety Research, 36*(1), 97-108.

Abdel-Aty, M., & Pande, A. (2007). Crash data analysis: Collective vs. individual crash level approach. *Journal of Safety Research, 38*(5), 581-587.

Abellán, J., López, G., & De OñA, J. (2013). Analysis of traffic accident severity using decision rules via decision trees. *Expert Systems with Applications, 40*(15), 6047-6054.

Aguero-Valverde, J. (2013). Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. *Accident Analysis & Prevention, 50,* 289-297.

Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention, 34*(6), 729-741.

Al Haji, G. (2005). *Towards a road safety development index (RSDI): development of an international index to measure road safety performance.* Linköping University Electronic Press,

Alinezhad, A. (2016). An Integrated DEA and Data Mining Approach for Performance Assessment. *Iranian Journal of Optimization, 8*(2), 968-987.

Aljanahi, A., Rhodes, A., & Metcalfe, A. V. (1999). Speed, speed limits and road traffic accidents under free flow conditions. *Accident Analysis & Prevention, 31*(1), 161-168.

Alper, D., Sinuany-Stern, Z., & Shinar, D. (2015). Evaluating the efficiency of local municipalities in providing traffic safety using the Data Envelopment Analysis. *Accident Analysis & Prevention, 78,* 39-50.

Amirteimoori, A., & Kordrostami, S. (2012). A distance-based measure of super efficiency in data envelopment analysis: an application to gas companies. *Journal of Global Optimization, 54*(1), 117-128.

Anastasopoulos, P. C. (2016). Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. *Analytic methods in accident research, 11*, 17-32.

Anastasopoulos, P. C., Mannering, F. L., Shankar, V. N., & Haddock, J. E. (2012). A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis & Prevention, 45*, 628-633.

Anastasopoulos, P. C., Shankar, V. N., Haddock, J. E., & Mannering, F. L. (2012). A multivariate tobit analysis of highway accident-injury-severity rates. *Accident Analysis & Prevention, 45*, 110-119.

Anastasopoulos, P. C., Tarko, A. P., & Mannering, F. L. (2008). Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis & Prevention, 40*(2), 768-775.

Athanassopoulos, A. D., & Curram, S. P. (1996). A comparison of data envelopment analysis and artificial neural networks as tools for assessing the efficiency of decision making units. *Journal of the Operational Research Society*, 1000-1016.

Bastos, J. T., Shen, Y., Hermans, E., Brijs, T., Wets, G., & Ferraz, A. C. P. (2015). Traffic fatality indicators in Brazil: State diagnosis based on data envelopment analysis research. *Accident Analysis & Prevention, 81*, 61-73.

Bonneson, J., Pratt, M., Miles, J., & Carlson, P. (2007). Horizontal curve signing handbook. *Texas Transportation Institute. Texas A&M University System, College Station, Texas: Texas Department of Transportation*.

Boussofiane, A., Dyson, R. G., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European journal of operational research, 52*(1), 1-15.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.

CARE. (2016). *Traffic Safety Basic Facts-Main Figures*. Retrieved from Euro Commission: https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs2016_main_figures.pdf

Chang, L.-Y. (2005). Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. *Safety science, 43*(8), 541-557.

Chang, L.-Y., & Chen, W.-C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research, 36*(4), 365-375. doi:http://doi.org/10.1016/j.jsr.2005.06.013

Chang, L.-Y., & Chien, J.-T. (2013). Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. *Safety science, 51*(1), 17-22.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research, 2*(6), 429-444.

Chimba, D., & Kutela, B. (2014). Scanning secondary derived crashes from disabled and abandoned vehicle incidents on uninterrupted flow highways. *Journal of Safety Research, 50*, 109-116. doi:http://dx.doi.org/10.1016/j.jsr.2014.05.004

Clark, L. A. (1992). Tree-based models. *Statistical models*.

Dadashova, B., Arenas-Ramírez, B., Mira-McWilliams, J., & Aparicio-Izquierdo, F. (2016). Methodological development for selection of significant predictors explaining fatal road accidents. *Accident Analysis & Prevention, 90*, 82-94.

Dissanayake, S., & Lu, J. (2002). Analysis of severity of young driver crashes: Sequential binary logistic regression modeling. *Transportation Research Record: Journal of the Transportation Research Board*(1784), 108-114.

Dong, C., Clarke, D. B., Yan, X., Khattak, A., & Huang, B. (2014). Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. *Accident Analysis & Prevention, 70*, 320-329.

Dyson, R. G., & Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society, 39*(6), 563-576.

Eksler, V., & Lassarre, S. (2008). Evolution of road risk disparities at small-scale level: Example of Belgium. *Journal of Safety Research, 39*(4), 417-427.

Elvik, R. (1997). Effects on accidents of automatic speed enforcement in Norway. *Transportation Research Record: Journal of the Transportation Research Board*(1595), 14-19.

Elvik, R., Christensen, P., & Amundsen, A. (2004). Speed and road accidents. *An evaluation of the Power Model. TØI report, 740*, 2004.

ETSC. (2001). Transport Safety Performance Indicators. In: European Transport Safety Council,Brussels.

EU. (2016). *Traffic Safety Basic Facts-Motorways*. Retrieved from https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs2016_motorways.pdf

Frantzeskakis, J. M., & Iordanis, D. I. (1987). Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece. *Transportation Research Record*(1112).

Fu, R., Guo, Y., Yuan, W., Feng, H., & Ma, Y. (2011). The correlation between gradients of descending roads and accident rates. *Safety Science, 49*(3), 416-423. doi:http://dx.doi.org/10.1016/j.ssci.2010.10.006

Garber, N., & Ehrhart, A. (2000). Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. *Transportation Research Record: Journal of the Transportation Research Board*(1717), 76-83.

Geurts, K., Wets, G., Brijs, T., & Vanhoof, K. (2004). Identification and ranking of black spots: Sensitivity analysis. *Transportation Research Record: Journal of the Transportation Research Board*(1897), 34-42.

Givehchi, S., Hemmativaghef, E., & Hoveidi, H. (2017). Association between safety leading indicators and safety climate levels. *Journal of Safety Research, 62*, 23-32.

Golob, T. F., Recker, W., & Pavlis, Y. (2008). Probabilistic models of freeway safety performance using traffic flow data as predictors. *Safety science, 46*(9), 1306-1333.

Golob, T. F., & Recker, W. W. (2003). Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. *Journal of transportation engineering, 129*(4), 342-353.

Golob, T. F., & Recker, W. W. (2004). A method for relating type of crash to traffic flow characteristics on urban freeways. *Transportation Research Part A: Policy and Practice, 38*(1), 53-80.

Golob, T. F., Recker, W. W., & Alvarez, V. M. (2004). Tool to evaluate safety effects of changes in freeway traffic flow. *Journal of transportation engineering, 130*(2), 222-230.

Haddon Jr, W. (1980). Advances in the epidemiology of injuries as a basis for public policy. *Public health reports, 95*(5), 411.

Hakkert, A., Gitelman, V., & Vis, M. (2007). Road safety performance indicators: Theory. *Deliverable D3, 6*.

Hamzeie, R., Savolainen, P. T., & Gates, T. J. (2017). Driver speed selection and crash risk: insights from the naturalistic driving study. *Journal of Safety Research, 63,* 187-194.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining (adaptive computation and machine learning)*: MIT Press Cambridge, MA.

Hermans, E., Brijs, T., Wets, G., & Vanhoof, K. (2009). Benchmarking road safety: lessons to learn from a data envelopment analysis. *Accident Analysis & Prevention, 41*(1), 174-182.

Hermans, E., Van den Bossche, F., & Wets, G. (2008). Combining road safety information in a performance index. *Accident Analysis & Prevention, 40*(4), 1337-1344.

Imprialou, M.-I. M., Quddus, M., Pitfield, D. E., & Lord, D. (2016). Re-visiting crash–speed relationships: A new perspective in crash modelling. *Accident Analysis & Prevention, 86*, 173-185.

Janssens, D., Wets, G., Timmermans, H. J., & Arentze, T. A. (2007). *Modelling short-term dynamics in activity-travel patterns: Conceptual framework of the Feathers model.* Paper presented at the 11th World Conference on Transport Research, Berkeley CA, USA.

Jiawei, H., & Kamber, M. (2001). Data mining: concepts and techniques. *San Francisco, CA, itd: Morgan Kaufmann, 5*.

Joshua, S. C., & Garber, N. J. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation planning and technology, 15*(1), 41-58.

Jovanis, P. P., & Chang, H.-L. (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Record, 1068*, 42-51.

Karlaftis, M. G., & Golias, I. (2002). Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis & Prevention, 34*(3), 357-365.

Kashani, A. T., & Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety science, 49*(10), 1314-1320. doi:http://doi.org/10.1016/j.ssci.2011.04.019

Kaye, S.-A., Lewis, I., & Freeman, J. (2018). Comparison of self-report and objective measures of driving behaviour and road safety: A systematic review. *Journal of Safety Research, 65*, 141-151.

Kim, D.-G., Lee, Y., Washington, S., & Choi, K. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis & Prevention, 39*(1), 125-134.

Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis & Prevention, 34*(3), 313-321.

Kononov, J., Lyon, C., & Allery, B. (2011). Relation of flow, speed, and density of urban freeways to functional form of a safety performance function. *Transportation Research Record: Journal of the Transportation Research Board*(2236), 11-19.

Ladron de Guevara, F., Washington, S., & Oh, J. (2004). Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*(1897), 191-199.

Lee, C., Saccomanno, F., & Hellinga, B. (2002). Analysis of crash precursors on instrumented freeways. *Transportation Research Record: Journal of the Transportation Research Board*(1784), 1-8.

Lee, S. (2010). Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls. *Decision Support Systems, 49*(4), 486-497.

Lord, D., Manar, A., & Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis & Prevention, 37*(1), 185-199.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice, 44*(5), 291-305.

Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention, 37*(1), 35-46.

Luoma, J., & Sivak, M. (2007). Characteristics and availability of fatal road-crash databases in 20 countries worldwide. *Journal of Safety Research, 38*(3), 323-327.

Malyshkina, N., & Mannering, F. (2008). Effect of increases in speed limits on severities of injuries in accidents. *Transportation Research Record: Journal of the Transportation Research Board*(2083), 122-127.

Martin, J.-L. (2002). Relationship between crash rate and hourly traffic flow on interurban motorways. *Accident Analysis & Prevention, 34*(5), 619-629.

Miller, J., & Ripley, E. (1996). Layered intrusions of the Duluth complex, Minnesota, USA. *Developments in Petrology, 15*, 257-301.

Milton, J. C., & Mannering, F. L. (1996). The relationship between highway geometrics, traffic related elements, and motor vehicle accidents.

Mohan, D., Bangdiwala, S. I., & Villaveces, A. (2017). Urban street structure and traffic safety. *Journal of Safety Research, 62*, 63-71.

Mustakim, F., & Busu, A. (2007). *Identifying, prioritizing and treating hazardous location at Federal Route FT023 (Muar-Segamat).* Paper presented at the Malaysian Road Conference, 7th, 2007, Kuala Lumpur, Malaysia.

Mustakim, F., & Fujita, M. (2011). Development of accident predictive model for rural roadway. *World Academy of Science, Engineering and Technology, 58*(2011), 126-131.

Noland, R. B., & Oh, L. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. *Accident Analysis & Prevention, 36*(4), 525-532.

OECD. (2002). *Safety on Roads: What's the Vision?* : The Organisation for Economic Co-operation and Development (OECD),Paris, France.

.

Ogwueleka, F. N., Misra, S., Ogwueleka, T. C., & Fernandez-Sanz, L. (2014). An artificial neural network model for road accident prediction: a case study of a developing country. *Acta Polytechnica Hungarica, 11*(5), 177-197.

Olutayo, V., & Eludire, A. (2014). Traffic accident analysis using decision trees and neural networks. *International Journal of Information Technology and Computer Science (IJITCS), 6*(2), 22.

Pande, A., & Abdel-Aty, M. (2006). Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis & Prevention, 38*(5), 936-948.

Papadimitriou, E., Yannis, G., Bijleveld, F., & Cardoso, J. L. (2013). Exposure data and risk indicators for safety performance assessment in Europe. *Accident Analysis & Prevention, 60,* 371-383.

Persaud, B., Retting, R., & Lyon, C. (2000). Guidelines for identification of hazardous highway curves. *Transportation Research Record: Journal of the Transportation Research Board*(1717), 14-18.

Rahimi, I., Behmanesh, R., & Yusuff, R. M. (2013). A Hybrid Method for Prediction and Assessment Efficiency of Decision Making Units: Real Case Study: Iranian Poultry Farms. *International Journal of Decision Support System Technology (IJDSST), 5*(1), 66-83.

Rosić, M., Pešić, D., Kukić, D., Antić, B., & Božović, M. (2017). Method for selection of optimal road safety composite index with examples from DEA and TOPSIS method. *Accident Analysis & Prevention, 98*, 277-286. doi:http://doi.org/10.1016/j.aap.2016.10.007

Rumar, K. (1999). Transport safety visions, targets and strategies: beyond 2000. *1st European Transport Safety Lecture. European Transport Safety Council, Brussels, Tech. Rep*.

SafetyNet, W. (2005). State of the art report on road safety performance indicators. *EU Integrated Project SafetyNet*.

Samoilenko, S., & Osei-Bryson, K.-M. (2008). Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and decision trees. *Expert Systems with Applications, 34*(2), 1568-1581.

Samoilenko, S., & Osei-Bryson, K.-M. (2013). Using Data Envelopment Analysis (DEA) for monitoring efficiency-based performance of productivity-driven organizations: Design and implementation of a decision support system. *Omega, 41*(1), 131-142.

Schneider IV, W., Zimmerman, K., Van Boxel, D., & Vavilikolanu, S. (2009). Bayesian analysis of the effect of horizontal curvature on truck crashes using training and validation data sets. *Transportation Research Record: Journal of the Transportation Research Board*(2096), 41-46.

Schneider IV, W. H., Savolainen, P. T., & Moore, D. N. (2010). Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways. *Transportation Research Record, 2194*(1), 91-98.

Seol, H., Choi, J., Park, G., & Park, Y. (2007). A framework for benchmarking service process using data envelopment analysis and decision tree. *Expert Systems with Applications, 32*(2), 432-440.

Sexton, T. R., Silkman, R. H., & Hogan, A. J. (1986). Data envelopment analysis: Critique and extensions. *New Directions for Evaluation, 1986*(32), 73-105.

Shafaghizadeh, S., & Attari, M. Y. N. (2014). An integrated fuzzy Data Envelopment Analysis and Data Mining for Performance assessment of Insurance branches. *Recent Developments in Data Envelopment Analysis and its Applications*, 64.

Shah, S. A. R., Ahmad, N., Shen, Y., Pirdavani, A., Basheer, M. A., & Brijs, T. (2018). Road Safety Risk Assessment: An Analysis of Transport Policy and Management for Low-, Middle-, and High-Income Asian Countries. *Sustainability, 10*(389).

Shah, S. A. R., Brijs, T., Ahmad, N., Pirdavani, A., Shen, Y., & Basheer, M. A. (2017). Road Safety Risk Evaluation Using GIS-Based Data Envelopment Analysis—Artificial Neural Networks Approach. *Applied Sciences, 7*(9), 886.

Shankar, V., & Mannering, F. (1996). An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research, 27*(3), 183-194.

Shen, Y., Hermans, E., Brijs, T., Wets, G., & Vanhoof, K. (2012). Road safety risk evaluation and target setting using data envelopment analysis and its extensions. *Accident Analysis & Prevention, 48*, 430-441.

Siddique, R., Aggarwal, P., & Aggarwal, Y. (2011). Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. *Advances in Engineering Software, 42*(10), 780-786.

Sohn, S. Y., & Lee, S. H. (2003). Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea. *Safety science, 41*(1), 1-14.

Sohn, S. Y., & Moon, T. H. (2004). Decision tree based on data envelopment analysis for effective technology commercialization. *Expert Systems with Applications, 26*(2), 279-284.

Sze, N.-N., & Wong, S. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis & Prevention, 39*(6), 1267-1278.

Torbic, D. J., Harwood, D., Gilmore, D., Pfefer, R., Neuman, T., Slack, K., & Hardy, K. (2004). *Guidance for implementation of the AASHTO strategic highway safety plan. Volume 7: A guide for reducing collisions on horizontal curves* (0309087600). Retrieved from

Wang, B., Hallmark, S., Savolainen, P., & Dong, J. (2017). Crashes and near-crashes on horizontal curves along rural two-lane highways: Analysis of naturalistic driving data. *Journal of Safety Research, 63*, 163-169.

Wang, C., Quddus, M. A., & Ison, S. G. (2009). Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accident Analysis & Prevention, 41*(4), 798-808.

Wang, H., Li, J., Chen, Q.-Y., & Ni, D. (2011). Logistic modeling of the equilibrium speed–density relationship. *Transportation Research Part A: Policy and Practice, 45*(6), 554-566.

Wang, J.-j., & Chen, F.-a. (2012a). Risk Evaluation of Expressway Traffic Safety under" Prior Restraint" Idea. In *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient* (pp. 2169-2180).

Wang, J.-j., & Chen, F.-a. (2012b). Sensitivity analysis of expressway traffic safety risk. In *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient* (pp. 2219-2229).

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention, 60*, 71-84.

Wegman, F. (2014). Analyzing road design risk factors for run-off-road crashes in the Netherlands with crash prediction models. *Journal of Safety Research, 49*, 121. e121-127.

Wegman, F., Commandeur, J., Doveh, E., Eksler, V., Gitelman, V., Hakkert, S., . . . Oppe, S. (2008). SUNflowerNext: Towards a composite road safety performance index. *Deliverable D6-SWOV, Leidschendam, The Netherlands., 16*.

WHO. (2004). World report on road traffic injury prevention. In. Switzerland: World Health Organization Geneva.

WHO. (2006). *The world health report 2006: working together for health*: World Health Organization.

WHO. (2015). *Global Status Report on Road Safety 2015*: World Health Organization,Italy.

Wong, Y.-H., & Beasley, J. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society, 41*(9), 829-835.

Wu, D. (2009). Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications, 36*(5), 9105-9112.

Yasin Çodur, M., & Tortum, A. (2015). An Artificial Neural Network Model for Highway Accident Prediction: A Case Study of Erzurum, Turkey. *PROMET-Traffic&Transportation, 27*(3), 217-225.

Ye, F., & Lord, D. (2011). Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit. *Transportation Research Record: Journal of the Transportation Research Board*(2241), 51-58.

Yeo, H., Jang, K., Skabardonis, A., & Kang, S. (2013). Impact of traffic states on freeway crash involvement rates. *Accident Analysis & Prevention, 50*, 713-723.

Zhou, M., & Sisiopiku, V. (1997). Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record: Journal of the Transportation Research Board*(1581), 47-52.