Light Condition	Nominal	6 levels	Daylight Dark, Lighted Dusk Dark, Not Lighted Dawn Dark, Unknown Lighting
Road Class	Nominal	4 levels	Interstate US & State Highways Farm to Market City Street
Surface Width	Numeric	24-168	Low Width, High Width
Weather Condition	Nominal	7 levels	Clear Cloudy Rain Fog Sleet/Hail Snow Blowing Sand/Snow

Table 2. Distribution of injury severity by key variables

Crash conditioning variables	Severity frequency	
	Serious Injuries or Killed 2147 (39%)	Light Injuries 3410 (61%)
Adjusted Average Daily Traffic Amount		
Low Volume Traffic	1247 (40%)	1866 (60%)
High Volume Traffic	900 (37%)	1544 (63%)
Crash Time		
Off Peak Hour	1353 (41%)	1936 (59%)
Peak Hour	794 (35%)	1474 (65%)
Light Condition		
Daylight	1334 (34%)	2597 (66%)
Dark, Lighted	565 (49%)	580 (51%)
Dusk	27 (47%)	30 (53%)
Dark, Not Lighted	179 (55%)	147 (45%)
Dawn	21 (45%)	26 (55%)
Dark, Unknown Lighting	21 (41%)	30 (59%)
Road Class		
Interstate	663 (39%)	1053 (61%)
US & State Highways	1319 (38%)	2144 (62%)
Farm to Market	157 (43%)	204 (57%)
City Street	8 (47%)	9 (53%)
Surface Width		
Low Width	1100 (39%)	1705 (61%)
High Width	1047 (38%)	1705 (61%)
Weather Condition		
Clear	1539 (39%)	2380 (61%)
Cloudy	323 (37%)	543 (63%)
Rain	271 (37%)	466 (63%)
Fog	5 (42%)	7 (58%)
Sleet/Hail	8 (38%)	13 (62%)
Snow	1(100%)	0 (0%)
Blowing Sand/Snow	0 (0%)	1 (100%)

5 Results

Using RStudio, the CART method was employed to classify the crash severity. To recognize the crucial factors of injury severity, six independent variables were used. Additionally, The Gini index was used as the CART's default splitting criterion.

Fig. 1 shows the classification tree. As it can be easily distinguished, this tree has five terminal nodes and the Light Condition, Crash Time and Weather Condition are the basic splitters. This implies that the crucial factors in crash severity in Beaumont's crashes are these three variables. The first split in node 1 is based on the most important factor light condition, which points out the most appropriate variable to classify the crash severity base on the dataset. CART splits the light conditions into dark-lighted, dark-not lighted, dawn or dusk in the left node and dark-unknown lighting and daylight in the right node. In fact, the tree predicts that if a crash happens in the light condition of dark with unknown lighting or daylight, 34% of the crashes will cause serious injuries or killed and 66% will cause light injuries (Terminal node 5).

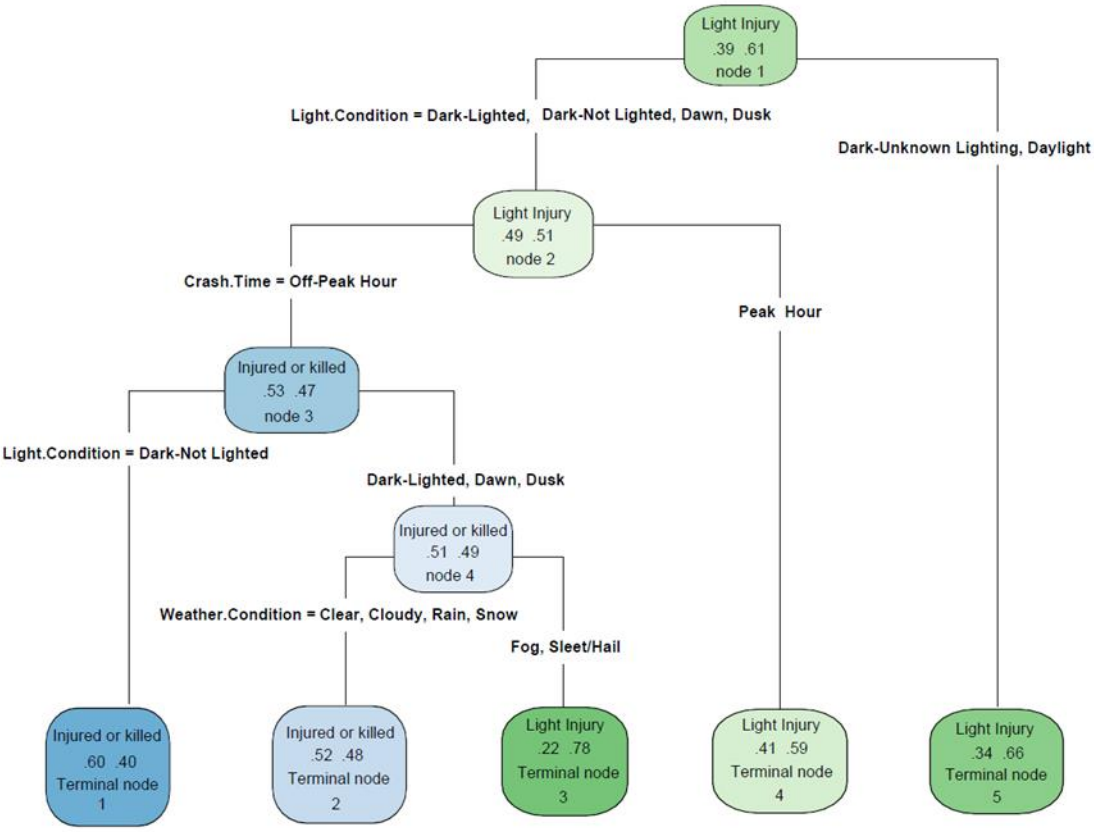

Fig. 1. The output of the CART tree

Yet, in the left branch of the tree, CART continues to split node 2 based on the variable of crash time and divide the crash time into peak hour and off-peak hour. It forecasts that in the light condition of dark-lighted, dark-not lighted, dawn or dusk and the crash time of peak hour, 41% of the crashes will cause serious injuries or killed and 59% will lead to light injuries (Terminal node 4). Regarding to the crash time of off-peak hour and light condition node 3 is formed. CART splits node 3 based on light condition and sends the dark-lighted, dawn and dusk to the right branch which forms node 4 and the rest of light conditions to the left branch which makes terminal node 1. Therefore, as indicated in terminal node 1 if the light condition is dark-not lighted and the crash time is not in peak hours (off-peak hours), the tree predicts that the injury severity is more likely to be severe and will cause severe injuries or people killed (60%). Moreover, the CART divides nod 4 based on weather condition into two parts. At terminal node 2, tree predicts that 52% of the crashes in clear, cloudy, rain or snow weather condition have severe injuries while in weather conditions of fog, sleet or hail (Terminal 3) 78% of crashes will cause light injuries. Besides, applying the CART method, table 3 represents the accuracy of the prediction which is made by this decision tree for testing models. As the table 3 indicates the model prediction accuracy for the testing data is 62% which means the CART model with the probability of 62% predicts the future data properly. As, Xu et al. [25] achieved the prediction accuracy of 55.2% for testing data and Wang et al. [21] got 62% overall model prediction accuracy in the testing phase, this model accuracy, is acceptable. Table 3 gives general information about the accuracy and performance of the model but should not be considered as the singular measure for assessing model performance.

Table 3. Prediction accuracy of the CART model for two classes of severity injuries

	Testing data	
	Correctly predicted	Observed severity
Serious Injuries or Killed	16 (64%)	25
Light Injuries	453 (61%)	733
Overall	469 (62%)	758

6 Discussion

CART results show that light condition, crash time and weather condition are the three most influential factors to crash severity. The output of terminal node 1 shows that in off-peak hours and dark-not lighted conditions the probability of a severe crash is bigger than a light one. Chen and Fan [27] implied in their research that crashes in daytime and off-peak were not severe. These findings show that low visibility in dark conditions is highly likely to increase the severity of crashes.

In terminal nodes 2 and 3 which are divided regarding weather condition, clear, cloudy, rainy and snowy weather are associated with severe injuries and other weather conditions are in association with light injuries. While some approaches confirm this study's result, others have conflicts with these outputs. Likely, regarding each investigation characteristic, many factors can play roles and could affect the results. For example, the study of Tavakoli Kashani and Shariat Mohaymany [8] confirmed that weather condition of clear, snowy and foggy is associated with serious injuries. Also, Nilsson et al. [49] analysis mentioned that adverse weather conditions enhance the risk of fatal run-off-road crashes. However, Dissanayake [30] argued that severe weather and physical disabilities do not significantly affect single vehicle crashes which is on contrary with other researches. Since the percentage of crashes in terminal node 3 is only 0.17% of the whole data, it can be concluded that clear, cloudy, rainy and snowy weather can increase the severity of injuries in this research.

According to terminal node 4, crashes in peak hours, and dark light conditions can cause less severe injuries. Chu [50] found similar results in his investigation that crashes during peak time are less severe in comparison with those during the off-peak time. This can be associated with the fact that drivers may drive at a higher speed in off-peak hours when the traffic volume is smaller than off-peak hours.

Furthermore, terminal node 5 reveals that driving in daylight time can cause more light injuries. This may be because, lighting increases drivers' ability to see the scenes properly and respond rapidly and in an appropriate manner if they detect any danger. Driving in poor light conditions may cause drivers to ignore the presence of traffic signs or pedestrians which is stated by a study of Li [51].

7 Conclusion

The purpose of this investigation is to identify the major factors of crash severity in Beaumont, Texas, and give some evidence-based recommendations to policymakers to alleviate the effects of crashes. This study showed that “light condition”, “crash time” and “weather condition” are the most crucial factors influencing the injury severity of crashes in Beaumont. The output indicates Beaumont roads have inadequate or insufficient lightning condition. Crash time is another main recognized variable. To be more precise, most probably drivers in peak hours, drive more carefully and as a result, the injuries are less severe in Beaumont. Also, the weather condition is found to be another important factor that causes severe injuries or fatality. The analysis revealed that in clear, cloudy and rainy weather, which is the dominant weather condition in Beaumont, more fatality and serious injuries will occur.

However, this research is implemented in Beaumont which is a port with large number of heavy vehicles on its roads. The result of this study can be taken into consideration to reduce the crash severity in new humid subtropical climate port and coastal urban areas. Additionally, regarding the complexity of transportation and traffic crashes, some future investigations are suggested. First, using other data mining methods, may help to extract additional risk factors and information. Combining human related with road-based factors can be the next step, as well. It can also help researchers and policy makers to achieve better understanding of traffic crashes and as a result assist decision-makers to make more efficient and cost-effective decisions.

Acknowledgments
This study was partially supported by the Natural Science Foundation (1726500) and the Center for Advances in Port Management (CAPM). The findings and conclusions of this paper are those of the authors and do not necessarily represent the official position of NSF and CAPM.

References:

[1] World Health Organization. Global Health Observatory data repository. *Road traffic deaths data by country*, 2019. Available from: https://apps.who.int/gho/data/node.main.A997

[2] Texas Department of Transportation. Crash Record Information System. *TxDOT Crash Query Tool*, 2019. Available from: https://cris.dot.state.tx.us/public/Query/app/welcome

[3] Fanny M, Norros I, Innamaa S. Accident risk of road and weather conditions on different road types. *Accident Analysis & Prevention*. 2019;122: 181–188.

[4] Taamneh M, Alkheder S, Taamneh S. Data mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *Journal of Transportation Safety & Security*. 2017; 9(2):146–166.

[5] Pakgohar A, Sigari Tabrizi R, Khalil M, Esmaeili A. The role of human factor in incidence and severity of road crashes. *Procedia Computer Science*, 2011. P. 764–769.

[6] Ossenbruggen P, Pendharkar J, Ivan J. Roadway safety in rural and small urbanized areas. *Accident Analysis and Prevention*. 2001;33: 485–498.

[7] Chang L-Y., Wang H-W. Analysis of traffic injury severity: An application of non-parametric classification tree techniques . *Accident Analysis and Prevention*. 2006;38: 1019–1027.

[8] Tavakoli Kashani A, Shariat Mohaymany A. Analysis of the traffic injury severity on two-lane, two-way rural roads based. *Safety Science*, 2011;49(10): 1314–1320.

[9] Shirali G, Valipour Noroozi M, Saki Malehi A. The outcome of occupational accidents by CART and CHAID. *Journal of Public Health Research*. 2018; 7(1361): 74–80.

[10] Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression. Montery: Wadsworth and Brooks/Cole; 1984.

[11] American Association of Port Authorities (APPA). 2013. *U.S. Port Ranking by Cargo Volume 2013*. United States of America.

[12] Hegar G. Port of entery, Port of Beaumont, *Texas Comptroller of Public Accounts*. 2018. Available from: https://comptroller.texas.gov/economy/economic-data/ports/snap-beaumont.php

[13] Dick J. Port of Beaumont plans new truck-queuing station, *Beaumont Enterprise*. 2020. Available from: https://www.beaumontenterprise.com/news/article/Port-of-Beaumont-plans-new-truck-queuing-station-15372117.php

[14] Brooks S. TxDOT awards Port of Beaumont $1.57 million grant, *Beaumont Business Journal*. 2020. Available from: https://www.beaumontbusinessjournal.com/news/txdot-awards-port beaumont-157-million-grant

[15] Freedman A, Samenow J. Flooded again: Climate change is making flooding more frequent in Southeast Texas. *The Washington Post*, 2019. Available from: https://www.washingtonpost.com/weather/2019/09/20/flooded-again-climate-change-is-making-flooding-more-frequent-southeast-texasthanks-part-climate-change/

[16] Washington S P, Karlaftis M G, Mannering F. Statistical and Econometric Methods for Transportation Data Analysis (2nd Edition ed.). Chapman and Hall/CRC; 2010.

[17] Chang L, Chen, W. Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*. 2005;36: 365–375.

[18] Huang H, Peng Y, Wang J, Luo Q, Li X. Interactive risk analysis on crash injury severity at a mountainous freeway with tunnel groups in China. *Accident Analysis and Prevention*. 2018;111: 56–62.

[19] Rakotonirainy A, Steinhardt D, Delhomme P, Darvell M, Schramm A. Older drivers' crashes in Queensland, Australia. *Accident Analysis and Prevention*. 2012;48: 423–429.

[20] Mergia W Y, Eustace D, Chimba D, Qumsiyeh M. Exploring factors contributing to injury severity at freeway merging and diverging locations in Ohio. *Accident Analysis and Prevention*. 2013; 55: 202–210.

[21] Wang J, Zheng Y, Li X, Yu C, Kodaka K, Li K. Driving risk assessment using near-crash database through data mining of tree-based model. *Accident Analysis and Prevention*. 2015; 84:54–64.

[22] Karlaftis M G, Golias I. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis and Prevention*. 2002;34: 357–365.
[23] Öström M, Eriksson A. Pedestrian fatalities and alcohol. *Accident Analysis and Prevention*. 2001;33(2): 173–180.
[24] Castro Y, Kim Y J. Data mining on road safety: factor assessment on vehicle accidents using classification models. *International Journal of Crashworthiness*. 2015; 21(2): 1–7.
[25] Xu X, Šaric Ž, Kouhpanejade A. Freeway incident frequency analysis based on CART method. *Promet – Traffic & Transportation*. 2014;26: 191–199.
[26] Ma Z, Chien S I-J, Dong C, Hu D, Xu T. Exploring factors affecting injury severity of crashes in freeway tunnels. *Tunnelling and Underground Space Technology*. 2016;59: 100–104.
[27] Chen Z, Fan W D. A multinomial logit model of pedestrian-vehicle crash severity. *International Journal of Transportation*. 2019;8: 43–52.
[28] Prati G, Pietrantoni L, Fraboni F. Using data mining techniques to predict the severity of bicycle crashes. *Accident Analysis and Prevention*. 2017;101: 44–54.
[29] Carlin J B, Taylor P, Nolan T. School based bicycle safety education and bicycle injuries in children: a case-control study. *Injury Prevention*. 1998;4: 22–27.
[30] Dissanayake S. Young Drivers and Run-Off-the-Road Crashes. *Proceedings of the 2003 Mid-Continent Transportation Research Symposium;* 2003. P. 1–6.
[31] Qiong W, Guohui Z, Yusheng C, Lina, W, Rafiqul, A T, Adélamar A. Exploratory multinomial logit model–based driver injury severity analyses for teenage and adult drivers in intersection-related crashes. *Traffic Injury*. 2016;17(4):1–9.
[32] Koetse M J, Rietveld P. The impact of climate change and weather on transport: An overview of empirical findings. *Transportation Research Part D*. 2009;14: 205–221.
[33] Çelik A K, Oktay E. A multinomial logit analysis of risk factors influencing road trafficinjury severities in the Erzurum and Kars Provinces of Turkey. *Accident Analysis and Prevention*. 2014;72: 66–77.
[34] Fan W D, Kane M R, Haile E. Analyzing severity of vehicle crashes at highway-rail grade crossings: multinomial logit modeling. *Journal of the Transportation Research Forum*. 2015;54(2): 39–56.
[35] Moore D N, Schneider IV W H, Savolainen P T, Farzaneh M. Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations. *Accident Analysis and Prevention*. 2011;43: 621–630.
[36] Tay R, Choi J, Kattan L, Khan A. A multinomial logit model of pedestrian–vehicle crash severity. *International Journal of Sustainable Transportation*. 2011;5: 233–249.
[37] Li Z, Ci Y, Chen C, Zhang G, Wu Q, Qian Z. Investigation of driver injury severities in rural single-vehicle crashes under conditions using mixed logit and latent class models. *Accident Analysis and Prevention*. 2019;124: 219–229.
[38] Wu Q, Zhang G, Ci Y, Wu L, Tarefder R A. Exploratory multinomial logit model–based driver injury severity analyses for adult drivers in intersection-related crashes. *Traffic Injury Prevention*. 2016;4(17): 413–422.
[39] Iranitalab A, Khattakb A. Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis and Prevention*. 2017;108: 27–36.
[40] Zeng Q, Huang H, Pei X, Wong S C. Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. *Analytic Methods in Accident Research*. 2016;10: 12–25.
[41] Tang J, Liang J, Han C, Li Z, Huang H. Crash injury severity analysis using a two-layer stacking framework." *Accident Analysis and Prevention*. 2019; 122: 226–238.
[42] Abellán J, López G, de Ona J. Analysis of traffic accident severity using Decision Rules via Decision. *Expert Systems with Applications*. 2013;40: 6047–6054.
[43] Kuhnert P M, Do K-A, McClure R. Combining non-parametric models with logistic regression: an application to motor vehicle injury data. *Computational Statistics & Data Analysis*. 2000;34: 371–386.
[44] Young Sohn S, Shin H. Pattern recognition for road traffic accident severity Korea. *Ergonomics*. 2001;44(1): 107–117.
[45] Wikipedia, Cimate of Beaumont, Texas, 2021. Available from: https://en.wikipedia.org/wiki/Climate_of_Beaumont,_Texas
[46] Rovšek V, Batista M, Bogunović B. Identifying the key risk factors of traffic accident injury severity on slovenian roads using a non-

parametric classification tree. *Transport*. 2017; 32(3): 272–281.

[47] Stewart J R. Applications of classification and regression tree method in roadway safety study. *Transportation Research Record.* 1996;1542(1): 1–5.

[48] Iacobucci D, Posavac S S, Kardes F R, Schneider M J, Popovich D L. The median split: Robust, refined, and revived. *Journal of Consumer Psychology*. 2015;25(4): 690–704.

[49] Nilsson D, Lindman M, Victor T, Dozza M. Definition of run-off-road crash clusters—For safety benefit estimation and driver assistance development. *Accident Analysis and Prevention.* 2018;113: 97–105.

[50] Chu H-C. Assessing factors causing severe injuries in crashes of high-deck busesin long-distance driving on freeways. *Accident Analysis and Prevention*, 2014;92: 130–136.

[51] Li G. Big data based exploration of risk factors to traffic crashes in southeast Texas and an experimental validation. *Doctoal thesis*. Lamar University Beaumont, Texas; 2019.