International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 10 Issue 6 Series I || June 2021 || PP 30-38

In-Service Performance Comparison of Roadside Safety Devices in Rural and Metro Counties in Texas

Hanzhen Wang[1], Fengxiang Qiao[1], Jianbang Du[1], Yunpeng Zhang[2]
[1](*Innovative Transportation Research Institute, Texas Southern University, USA*)
[2](*Department of Information and Logistics Technology, University of Houston, USA*)

ABSTRACT: *The In-Service Performance Evaluation (ISPE) satisfies transportation safety requirements and is an integral part of the process of traffic safety policy making. The purpose of this paper is to evaluate the performance of safety devices installed in Texas rural and metro areas, so as to enhance the equity of safety devices in the decision-making process, and thus prevent a higher proportion of serious traffic casualties in both rural and metro areas. Ten-year crash data from 2010 to 2019 in Texas was gathered from the Texas Crash Records Information System. XGBoost Machine Learning based analysis shows that, the roadside safety devices in rural areas and their safety related performances are relatively limited. A higher percentage of incapacity and fatal crashes were counted in rural areas with an area specified inequality of transportation safety risk. Addressing the transportation safety equity is recommended when deploying and maintaining safety devices in all regions.*

KEYWORDS – *Crash Data, ISPE, Machine Learning, Roadside Safety Devices, Transportation Equity*

Date of Submission: 02-06-2021 Date of Acceptance: 16-06-2021

I. INTRODUCTION

In the United States, there are about 38,000 people lost their lives each year due to traffic crashes, with a traffic fatality rate being 12.4 deaths for every 100,000 residents, plus 4.4 million people injured with medical attention [1]. In 2018, the statewide traffic related deaths were 3,641 in State of Texas, which implies that 10 Texans would die of traffic crashes in each single day [2]. In order to deal with such a critical situation, the Texas Department of Transportation (TxDOT) has targeted on significantly reducing crashes by 50 percent till year 2035 and eliminating deaths of traffic to ZERO by year 2050 [3]. To enhance the roadway safety, there is a chain of countermeasures proposed to enhance the road safety management system. These countermeasures are traditionally falling into the fields of engineering, enforcement, and education. While enforcement and education can be called the “soft” treatments, engineering related prevention activities are more like a kind of “hardware” related ones [4].

The roadside safety device is such an engineering hardware that is designed and operated with the intention to significantly reduce roadway related traffic crashes [5]. The common road traffic safety devices include: (1) longitudinal barriers (guardrails), (2) barrier terminals (guardrail end treatments), (3) crash cushions, and (4) breakaway hardware (signs, luminaires, etc.). The term “barrier” is defined by the American Association of State Highway Transportation Officials (AASHTO) Roadside Design Guide (RDG) as: (1) roadside barrier, (2) median barrier, and (3) bridge railing [6]. The AASHTO guide refers the Strategic Plan for Highway Safety and the NCHRP Report 500 Series of guidance reports to assistall States to design their ways for reducing injuries and fatalities of traffic crashes.

The United States Fixing America's Surface Transportation Act or "FAST Act" required to identify methods to collect data about roadside highway safety hardware for evaluating the in-service performance [7]. The In-service Performance Evaluation (ISPE) of roadside safety devices can be specified into four levels [8, 9]. Level I is to continuously monitor the comprehensive database system including roadway design information, safety device information, traffic, incidents, and crash. Level II is to collect supplemental data for ad hoc study and review to obtain additional information. Level III is to conduct an in-depth investigation of selected studies with details of high levels, which are normally associated with high cost of crashes. Level IV is to integrate the evaluation of new products with problems occurred from the maintenanceand operationof roadside safety devices [10].

While many studies and tests having been conducted related to the ISPE of roadside safety devices, there is a lack of study on imbanded crash rates in rural and metro areas that are related to roadside devices. This motivates the research of this paper.

II. DATA COLLECTION AND METHODOLOGY

2.1 Data Collection

In this research, two main parts of data were collected. The first part is Texas statewide crash data, which provides the crash information related to each type of roadside safety device. In the meantime, information on all Texas counties is demanded to integrate with the on-road crashes during the analysis of this research. Thus, the second part is the data set on the basic property of each Texas county including population, area, and designation types (rural or metro county).

The ten-years statewide crash data from 1/1/2010 to 12/31/2019 were collected from the Crash Record Information System (CRIS), which is a statewide database for reportable motor vehicle traffic crashes operated by TxDOT [11]. The contributor to the crash records is mainly Texas Peace officer's reports. The raw data collected from the CRIS database includes 172 integrated information of 5,629,779 uniquely numbered crash records. The information contains the features of each crash such as location, date, weather, road type, vehicle involved, personal, injury severity, and related safety devices, *etc.* The selected features of each crash record in the raw data collected are presented in their original indexes, which were abbreviated in this paper as shown in Table 1 for the purpose of analyzing the accuracy and efficiency.

Table 1. The Selected Features in The Collected Database

Selected Features	Abbreviation
Roadway System ID	RRSI
Road Part ID	RRPI
CrashSpeedLimit	SPL
Weather Condition ID	WCI
Light Condition ID	LCI
Surface Condition ID	SCI
Object Struck ID	OSI
Road Relation ID	RRI
Base Type ID	BTI
MedianTypeID	MTI
$1000 Property Damage	TDF
Crash Severity ID	CSI

The information of 254 Texas counties was collected based on the population and area of each county. The population data were released from the U.S. Census Bureau on the American FactFinder (AFF) platform, which provides the annual estimates of the resident population [12]. The area data were collected from the National Association of Counties (NACo) that lists the area for each county in Texas [13]. The population and area data were integrated with crash data in rural and metro areas for further analysis and comparison.

2.2 Data Processing

After collecting the ten-years statewide crash data in Texas, crashes that hit roadside safety devices were filtered out based on the OSI feature for further processing. By selecting six struck object that categorized as roadside safety devices (guardrail, work zone barricade, concrete and cable median barrier, abutment or rail end, bridge rail, and concrete traffic barrier), the raw data were filtered into 275,333 crash records with 14 selected features, including individual crash ID and county ID that are related to roadside safety devices. The TDF feature was used as an output to evaluate property damage in crashes, and the CSI feature was applied as another output indicating crash severity ending in different injuries. To gain better performance in the machine learning process, values of CSI feature were replaced by the Equivalent Property Damage Only (EPDO) weights based on crash costs analysis conducted by FHWA [14]. EPDO weights wereassigned to five crash severity types: (1) 1 for no injury, (2) 6 for possible injury, (3) 11 for non-incapacity injury, (4) 30 for incapacity injury, and (5) 568 for fatal injury. Higher EPDO weight represents a more severe loss during crashes.

According to the definition assigned by the Office of Management and Budget (OMB) of the United State, some counties are designated as Metropolitan counties, which is consisted of a core urban area with more than 50,000 population. The rest counties that do not belong to metro counties are considered as rural counties [15]. Currently, there are 177 counties designated as rural counties while 77 counties as metro counties in Texas as presented by Texas Health Professions Resource Center (HPRC) [16]. The county type distribution map is shown in Fig. 1.

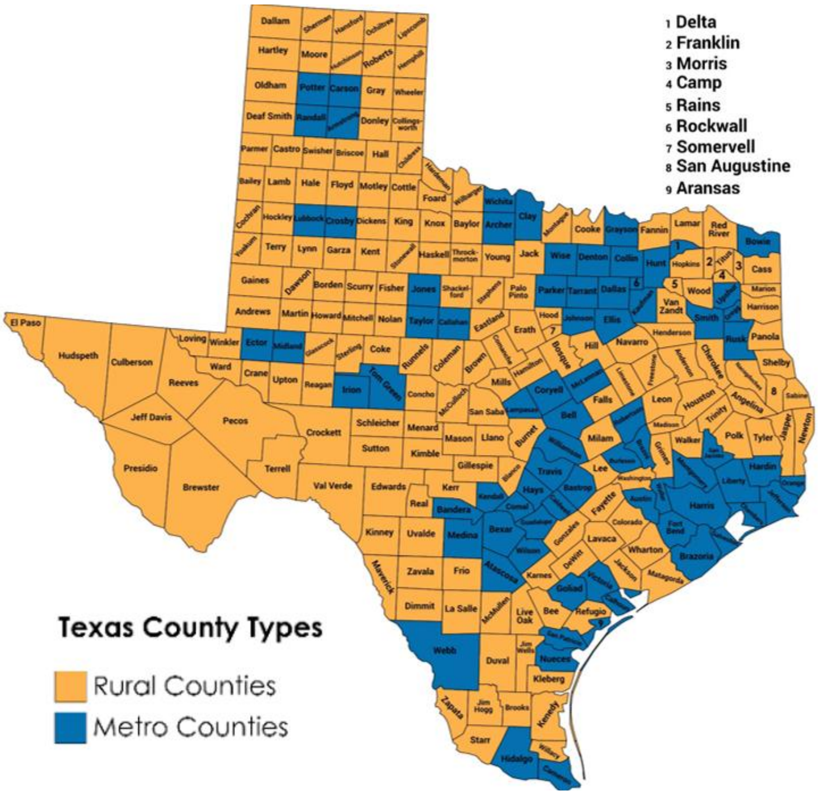


Figure 1. Texas metro and rural counties map [Source of data: 24]

Based on Fig. 1, the crash data in Texas were divided into metro and rural crashes based on their respective county IDs. Since the raw crash data contains null values for invalid data like missing or unknown values, the original data sets were preprocessed by replacing the null values with 0 for further analysis. A statistical analysis was then conducted to simply count and compare the proportion of crashes based on severity types and property damage scale in rural and metro areas. From gathered county information data, the total population and area in rural and metro counties were added up to count the distribution of crash severity and property damage based on population and area. The total population of 77 metro counties is 25,096,592 and the total area is 176,765 square kilometers (km^2), while 177 rural counties consist of a total of 3,236,708 population and a total 498,619 km^2 area [12, 13].

The crash severity distribution in two areas is shown in Table 2. Compared to metro areas, the total number of crashes happened in rural areas was much lower. However, the proportion of crashes by crash amount resulted in fatal and incapacity injury in rural counties is higher than that in metro counties. Since rural counties tend to be with more area and less population, the distributions of fatal and incapacity crashes are higher by population than metro counties in rural counties, but are lower by total area.

Table 2. Crash Severity Distribution in Texas Rural and Metro Counties (2010-2019)

	Unknown	**Incapacity Injury**	**Non-Incapacity Injury**	**Possible Injury**	**Fatal Injury**	**No Injury**
Rural Crash Amount	686	1,172	3,402	3,356	459	22,538
Percentage by Amount	2.170%	3.707%	10.761%	10.616%	1.452%	71.293%
Percentage by Population	0.021%	0.036%	0.105%	0.104%	0.014%	0.696%
Percentage by Area	0.138%	0.235%	0.682%	0.673%	0.092%	4.520%
Metro Crash Amount	14,337	7,791	31,247	42,089	2,138	146,118
Percentage by Amount	5.883%	3.197%	12.821%	17.269%	0.877%	59.953%
Percentage by Population	0.057%	0.031%	0.125%	0.168%	0.009%	0.582%
Percentage by Area	8.111%	4.408%	1.769%	23.811%	1.210%	82.6662%

Similarly, the property damage distributions by three elements in rural and metro counties are shown in Table 3. The proportion of crashes result in more than $1,000 damage on the property is higher in rural counties integrated by the number of crashes and total population. The percentage by area with severe property damage in metro counties isrelatively higher due to the limited area compared to rural counties.

Table 3. Property Damage Distribution in Texas Rural and Metro Counties (2010-2019)

	More than $1,000 Property Damage	**Less than $1,000 Property Damage**
Rural Crash Amount	31,241	372
Percentage by Amount	98.823%	1.177%
Percentage by Population	0.965%	0.011%
Percentage by Area	6.266%	0.075%
Metro Crash Amount	235,439	8,281
Percentage by Amount	96.602%	3.398%
Percentage by Population	0.938%	0.033%
Percentage by Area	133.193%	4.685%

2.3 Methodology

2.3.1 Feature Selection

To conduct machine learning based data analysis, feature selection, also known as attribute selection, is a vital part of the machine learning process. The step of feature selection before actual modeling can reduce the training time and improve the efficiency and performance of selected models by picking up the most suitable features for further analysis [17].

Some of the multiple input attributes contribute mostly to the specific output attribute, which can be selected during the feature selection process. Popular machine learning models that can be implemented for feature selection include: (1) Classification and Regression Tree (CART) or Decision Tree, (2) XGBoost, and (3) Random Forest.

2.3.2 Candidate Machine Learning Models

The data analysis is conducted by applying three candidate machine learning models on collected data to select features with higher importance scores. The model with the highest accuracy through model evaluation is chosen to continue selecting values under selected features in the last step.

Decision Tree model is widely used in machine learning modeling to evaluate output features based on several inputs. Classification and Regression Trees (CART) often refers to Decision Tree modeling that uses classification or regression to predict or select important features based on the output [18]. The model consists of a binary tree, which is a tree-like model with various nodes and paths to perform evaluation based on the decision rules learned in original data sets. The Decision Tree Classifier shows potential in recognizing feature patterns and can be implemented in feature selection practices [19].

XGBoost is a newly developed machine learning model that implements the gradient boosting technique based on the basic Decision Tree algorithm. It is featured as a scalable tree boosting system with higher running speed compared to other basic applications on the individual machine [20]. Apart from the improved construction speed in tree modeling, XGBoost also provides a developed algorithm in tree searching [21].

Random Forest is used to fit a number of Decision Tree models and estimates the average result by randomly assigning features on each tree model. The selected feature is determined based on the results from a set of CART models [22]. Random Forest belongs to the bagging method in the assembly algorithm. The importance of features is determined by Gini index in equation (1) [23].

$$G = 1 - \sum_{i=1}^{n} f^2 \quad (1)$$

where, G is the Gini index, *n* is the number of features, and *f* is the frequency of the feature.

2.3.3 Encoding Method and Model Performance Measure

Since the original input data may contain string type data that cannot be processed by machine learning models as numeric attributes, the training data are commonly labeled by a Label Encoding. The characters as values under features in the database are converted to numeric forms to be operable by machine learning models.

The performance measure for chosen machine learning models is an important step in finalizing the feature selection model between candidate models. The *k*-fold cross-validation is widely used for this purpose, in which *k* represents the folding times and ten is the most popular one. It estimates the prediction error and accuracy for selected machine learning models by randomly assigning data sets into subsets with the number of *k*[25].

III. MACHINE LEARNING RESULTS AND SELECTED FEATURES

3.1 Accuracy of Machine Learning Models

To proceed machine learning process on CRIS crash data, a final machine learning model between three candidate machine learning models need to be determined by measuring their performance through ten-fold cross validation. The feature selection process was performed twelve times on three candidate models. The crash data were split into two groups (rural and metro). Each group of crash data were examined with two outputs (property damage and crash severity) by three models. The result of accuracy is shown in Table 4.

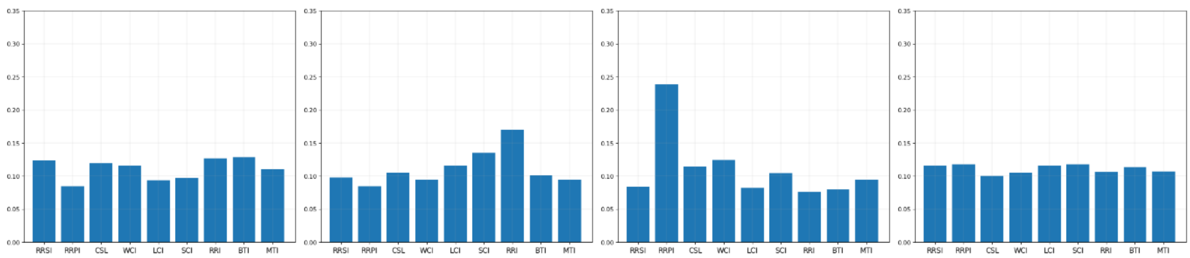
Table 4. Accuracy Result of Ten-fold Cross Validation for Candidate Models

Area	Property			Severity		
	Decision Tree	**XGBoost**	**Random Forest**	**Decision Tree**	**XGBoost**	**Random Forest**
Rural	0.978	0.987	0.987	0.638	0.709	0.682
Metro	0.952	0.966	0.963	0.521	0.599	0.561

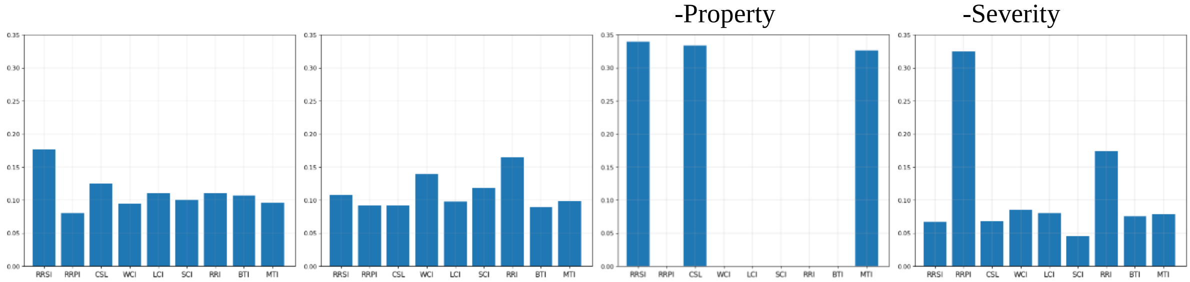
From the results shown in Table 4, higher model accuracies (the red ones) mean more suitable for feature selection on crash data. The XGBoost (the red columns) provides the highest accuracies in both data sets with both outputs, while Decision Tree provides the lowest accuracies relatively. Thus, the XGBoost model was chosen for further analysis in this study.

3.2 Selected Features

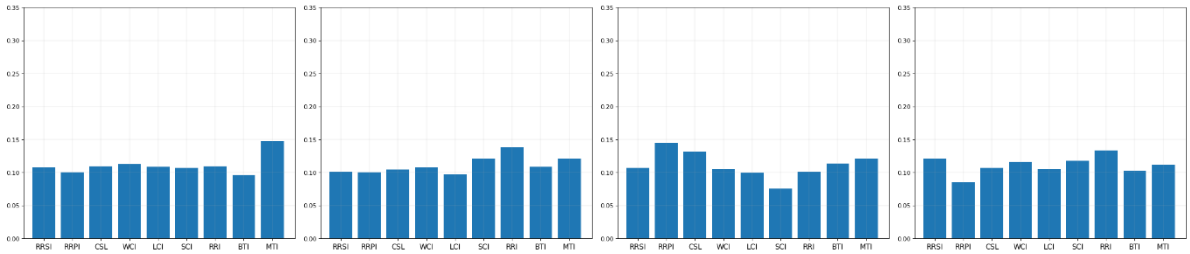
Based on the feature selection conducted by the XGBoost model, the ranking results of importance scores for selected features in the crash database are presented in two groups (rural and metro). Crash records related to six roadside safety devices are individually processed in each group with two outputs (property damage and crash severity). The feature selection results of crashes involving rural roadside safety devices are shown in Fig. 2.



(a) Guardrail-Property (b) Guardrail-Severity (c) Work Zone Barricade -Property (d) Work Zone Barricade -Severity



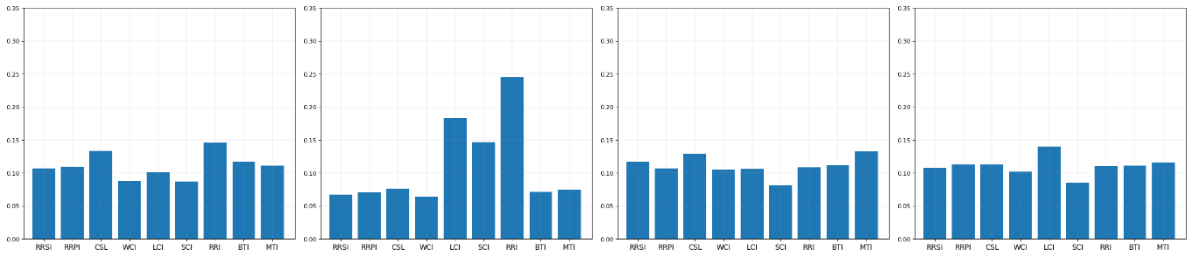
(e) Median Barrier-Property (f) Median Barrier-Severity (g) Abutment or Rail End -Property (h) Abutment or Rail End -Severity



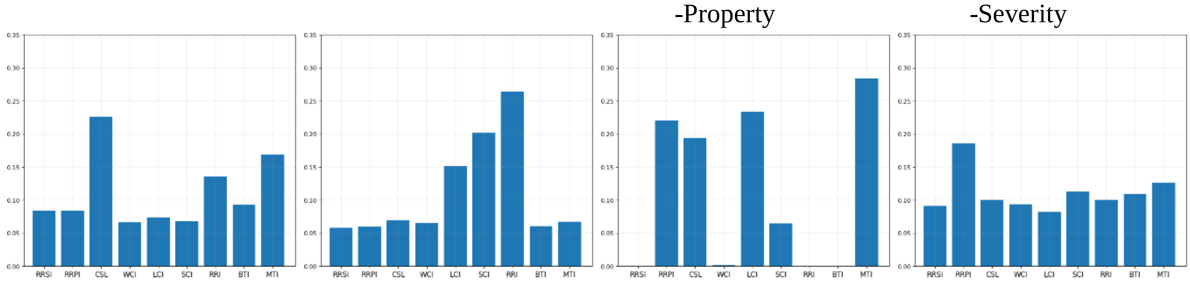
(i) Bridge Rail-Property (j) Bridge Rail-Severity (k) Concrete Traffic Barrier (l) Concrete Traffic
Barrier -Property -Severity

Figure 2. Feature selection results in rural counties

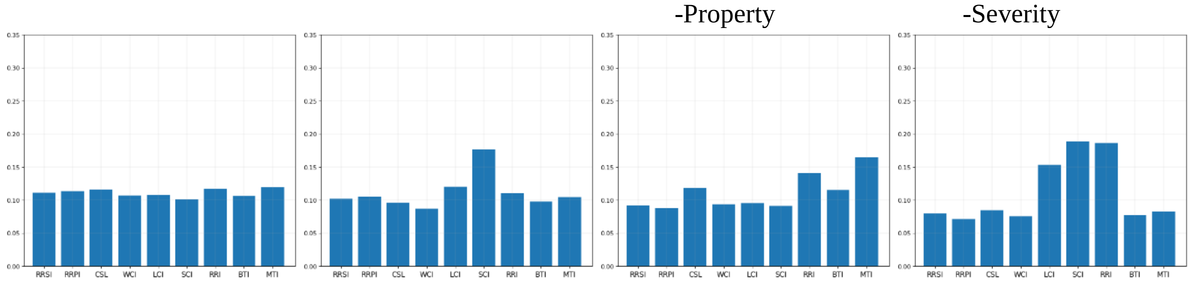
According to the importance ranking for selected features, the feature selection results vary based on types and outputs of safety devices. A general conclusion from Figure 2 is that, when the output is crash severity, the RRI tends to have more impact that emphasizes the crash location in relation to the roadway (whether on or off roadway, on shoulder or median), which contributes higher to the crash severity in rural counties. As for the property damage, the results are different depending on the device types. However, the CSI always has higher ranks in every case, which demonstrates that the speed limit in rural area has vital impact on property damage in a crash. Another group has the same trend during feature selection process, and the results of metro counties are shown in Fig. 3.



(a) Guardrail-Property (b) Guardrail-Severity (c) Work Zone Barricade -Property (d) Work Zone Barricade -Severity



(e) Median Barrier-Property (f) Median Barrier-Severity (g) Abutment or Rail End -Property (h) Abutment or Rail End -Severity



(i) Bridge Rail-Property (j) Bridge Rail-Severity (k) Concrete Traffic Barrier -Property (l) Concrete Traffic Barrier -Severity

Figure 3. Feature selection results in metro counties

Fig. 3 shows the results similarly conducted in metro counties, where the distributions of feature importance are quite different from rural counties. When considering crash severity as an output, the SCI feature is generally ranking higher compared to the RRI, as well as the LCI. It means that in metro counties, surface condition and light condition can be important elements in causing crashes with serious injuries. The CSL feature also has a higher proportion when the output is property damage, which is followed by the MTI feature generally. This result suggests that apart from speed limit, median type also has vital influence over the property damage that exceeds $1,000 in metro counties.

3.3 Detailed Attributes Selected through Advanced Feature Selection

At this stage, advanced feature selection was conducted in order to indicate which detailed attributes under selected features contribute more to the crash severity or property damage. After plotting the importance scores for each feature under various conditions, up to three features that have the importance score higher than or close to 1.0 were considered as highly correlated with corresponding output and selected for advanced feature selection.

The selected features were converted to binary format before processing advanced feature selection. The XGBoost machine learning model was implemented to perform advanced feature selection continuously with inputs replaced with binary attributes under selected features. After two groups of data were similarly trained by the XGBoost model again, up to three detailed attributes with an importance score higher than or close to 1.0 were selected. The advanced feature selection results are shown in Table 5.

Table 5. Advanced Feature Selection Results

	Guardrail	**Work Zone Barricade**	**Median Barrier**	**Abutment or Rail End**	**Bridge Rail**	**Concrete Traffic Barrier**
Rural (Property)	Ranch Road	Speed Limit 55	State Loop System	State Highway	Positive Barrier	One-way Pair Median
	Roadbed Soil	Cloudy	Ranch Road	Unprotected Median	Curbed Median	Speed Limit 40
	State Highway	Speed Limit 65			Fog Weather	
Rural (Severity)	On Roadway	County Road	Sand, Mud, Dirt Surface	Main/Proper Lane	Dry Surface	Dry Surface
	Dry Surface	Business US System		Median Road Part	On Roadway	
		Standing Water Surface				
Metro (Property)	Speed Limit 75	Ranch Road	Speed Limit 60	Unprotected Median	Speed Limit 10	Speed Limit 70
	Speed Limit 70	Speed Limit 80	Cloudy	Daylight		
			Positive Barrier	Main/Proper Lane		
Metro (Severity)	On Roadway	Daylight	Sand, Mud, Dirt Surface	Connector/Flyover	Dry Surface	Dry Surface
	Daylight		Off Roadway	Curbed Median	Daylight	Daylight
	Dry Surface		Dawn Light			On Roadway

IV. COUNTERMEASURES TO ENHANCE THE ROADSIDE DEVICES RELATED SAFETY IN RURAL AND METRO COUNTIES

4.1 Countermeasures to Reduce Property Loss in Rural Counties

Comparing to metro counties, rural counties have less population and larger area, resulting in fewer traffic flows. However, the property loss during crashes that happened in rural counties cannot be ignored as the proportion of higher property loss is bigger than in metro counties. According to this study, several median types of rural roadways have influences over the property loss. Since most positive barriers for the median part of roadways are installed in metropolitan areas [26], most rural median parts were left unprotected or without proper infrastructure. Therefore, improving the deployment and maintenance in rural counties on roadway hardware is suggested especially for medians to reduce overall property loss [27].

4.2 Countermeasures to Reduce Severe Crashes in Rural Counties

Receiving the results from advanced feature selection modeling, the surface of roadway including dry, sand, mud, dirt, and standing water condition would likely to impact severity level of crashes in rural counties. Besides, those frequently used roadways also contribute to crash severity in rural areas. With more human activities and density of structures in metro counties, roadways in metro areas have more chances for maintenance and re-construction. However, due to the limited accessibility and complicated traffic environment in rural counties, it is hard to improve and maintainroadway conditions, which may leave relatively worse surface condition on their main roads. It is recommended allocating enough funding for transportation projects in improving the performance of basic traffic infrastructure in rural areas with reduced severe crashes and property loss.

V. CONCLUSIONS AND RECOMMENDATIONS

This paper analyzed the safety in Texas rural and metro areas that are related to roadside safety devices. The XGBoost machine learning model was selected for advanced feature selection and modeling. Countermeasures on the operation and maintenance of roadside safety devices were proposed to reduce the property loss and severe crashes in rural areas. To achieve the HALF crash goal by 2035 and ZERO death goal by 2050 in Texas [3], it is recommended considering crashes in rural and metro areas equivalently. Transportation equity in the safety of roadside devices should be regarded as a necessary part in the decision-making process. High-level policy and strategy on the development and maintenance of transportation infrastructures in both rural and metro areas are vital in transportation planning and operations. Incorporating ISPE on roadside safety devices distributed in both rural and metro counties and examining the equity are highly recommended to enhance safety in all regions.

ACKNOWLEDGEMENTS

The authors acknowledge that the crash data used in this paper were provided by Texas Department of Transportation (TxDOT), United States. The findings,opinions, conclusions, and recommendations presented in this studyare those of the authors and do not necessarily reflect the views of TxDOT.

REFERENCES

[1] ASIRT (Association for Safe International Road Travel). (2020). *Annual Global Road Crash Statistics*. <https://www.asirt.org/safe-travel/road-safety-facts/#:~:text=More%20than%2038%2C000%20people%20die,for%20people%20aged%201%2D54>.

[2] Downey, R. (2019). "TxDOT goal: no traffic deaths by 2050." *Stateman*, <https://www.statesman.com/news/20190530/txdot-goal-no-traffic-deaths-by-2050#:~:text=Texas%20Department%20of%20Transportation%20officials%20set%20the%20ambitious%20goal%20Thursday,annually%20to%20meet%20those%20goals> (May 30, 2019).

[3] TxDOT (Texas Department of Transportation). (2019). *TxDOT Embraces Goal to End Deaths on Texas Roads By 2050*, <https://www.txdot.gov/inside-txdot/media-center/statewide-news/012-2019.html> (May 30, 2019).

[4] Penmetsa, P. (2017). *Modeling Crash Risk Due to Traffic Rule Violations for Education, Enforcement and Engineering Countermeasures*. Doctoral dissertation, The University of North Carolina at Charlotte, <https://ninercommons.uncc.edu/islandora/object/etd%3A1101/datastream/PDF/download/citation.pdf> (November 26, 2020).

[5] Cantisani, G., Di Mascio, P., & Polidori, C. (2017). "New research opportunities for roadside safety barriers improvement." *Building up Efficient and Sustainable Transport Infrastructure*, 21-22.

[6] AASHTO (American Association of State Highway Transportation Officials), (2012). *AASHTO Roadside Design Guide*, 4th Edition. <https://safety.fhwa.dot.gov/roadway_dept/countermeasures/reduce_crash_severity/policy_memo/aashto_rdg_120626/aashto_rdg_062612.pdf>

[7] FHWA (Federal Highway Administration. (2020). "Roadside Safety Hardware Identification Methods." *Safety*, <https://safety.fhwa.dot.gov/roadway_dept/countermeasures/reduce_crash_severity/id_methods/ch1.cfm> (August 11, 2020).

[8] Mak, K. & Sicking, D. L. (2002). *Continuous Evaluation of In-service Highway Safety Feature Performance*. Arizona Department of Transportation in cooperation with U.S. Department of Transportation Federal Highway Administration, <https://rosap.ntl.bts.gov/view/dot/15037>

[9] Zhang, Y., Cheng, L.-C., Qiao, F. and Patel, A. (2019). "Enhancement of In-Service Performance Evaluation (ISPE) Process for Roadside Safety Devices: A Survey." *IEEE/ACM International Conference on & International Conference on Cyber, Physical and Social Computing (CPSCom) Green Computing and Communications (GreenCom)*. < https://ieeexplore.ieee.org/document/8875426>

[10] Qiao, F., Wang, H., Du, J., Zhang, J., Cheng, V., and Odell, W. (2020). *Development and Enhancement of In Service Performance Evaluation (ISPE) Process for Roadside Safety Devices*. Tech Memo 3, Texas Department of Transportation (TxDOT) Project 0-7018.

[11] TxDOT (Texas Department of Transportation). (2020). *Crash Data Analysis and Statistics.*<https://www.txdot.gov/government/enforcement/crash-statistics.html>.

[12] U. S. Census Bureau (2018). *American FactFinder - Results.* <https://archive.vn/20200213192117/https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2016/PEPANNRES/0400000US48.05000#selection-203.19-203.38> (Nov. 2020)

[13] NACo (National Association of Counties). (2007). *NACo Find a County*. <https://web.archive.org/web/20070213225518/http://www.naco.org/Template.cfm?Section=Find_a_County&Template=%2Fcffiles%2Fcounties%2Fstate.cfm&state.cfm&statecode=tx> (Nov. 2020)

[14] Harmon T., Bahar, G. B., and Gross, F. B. (2018). *Crash Costs for Highway Safety An*alysis. Federal Highway Administration. Office of Safety, No. FHWA-SA-17-071.

[15] HRSA (Health Resources & Services Administration). (2020). *Defining Rural Population*. <https://www.hrsa.gov/rural-health/about-us/definition/index.html> (Nov. 2020)

[16] THHS (Texas Health and Human Services). (2020). *Definitions of County Designations*. <https://www.dshs.texas.gov/chs/hprc/counties.shtm> (Nov. 2020)

[17] Guyon, I. and Elisseeff, A. (2003). "An introduction to variable and feature selection." *Journal of Machine Learning Research*, 3(Mar), 1157-1182.

[18] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. CRC press.

[19] Safavian, S. R., & Landgrebe, D. (1991). "A survey of decision tree classifier methodology." *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3), 660-674.

[20] Chen, T., & Guestrin, C. (2016, August). "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 785-794.

[21] Torlay, L., Perrone-Bertolotti, M., Thomas, E., & Baciu, M. (2017). "Machine learning–XGBoost analysis of language networks to classify patients with epilepsy." *Brain informatics*, 4(3), 159-169.
[22] Belgiu, M., & Drăguţ, L. (2016). "Random forest in remote sensing: A review of applications and future directions." *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
[23] Schott, M. (2019). "Random Forest Algorithm for Machine Learning." *Medium, Capital One Tech*, 25.
[24] SORH (State Office of Rural Health). (2012, April). Texas County Designations.<https://www.google.com/url?q=https://www.texasagriculture.gov/portals/0/forms/er/rural-metro%2520counties.pdf&sa=U&ved=2ahUKEwiGzeWer5rxAhWCWM0KHWDCAZEQFjAAegQICRAB&usg=AOvVaw0pnwPcnGxeGaWtozCXwzjt>
[25] Fushiki, T. (2011). "Estimation of prediction error by using K-fold cross-validation." *Statistics and Computing*, 21(2), 137-146.
[26] TxDOT (Texas Department of Transportation). (2020). *TxDOT Median*. <https://gis-txdot.opendata.arcgis.com/datasets/txdot-median> (Nov. 2020)
[27] Du, J., Qiao, F., Wang, H., Zhang, Y. and Yu, L. (2021) "Frequent Pattern Analysis of the Roadside Safety Devices Related On-road Crashes." *International Journal of Engineering Science Invention (IJESI)*, Vol. 10(05), 2021, PP 35-46. Journal DOI10.35629/6734

Hanzhen Wang, et. al, "In-Service Performance Comparison of Roadside Safety Devices in Rural and Metro Counties in Texas." *International Journal of Engineering Science Invention (IJESI)*, Vol. 10(06), 2021, PP 30-38. Journal DOI- 10.35629/6734