A Time of Day Analysis of Pedestrian-Involved Crashes in California:

Investigation of Injury Severity, a Logistic Regression and Machine Learning Approach Using HSIS Data

CKA / Shutterstock

By Seyedmirsajad Mokhtarimousavi (S)

According to the National Highway Traffic Safety Association (NHTSA), in 2016, 5,987 pedestrians died in the United States, or 16 people every day, for one year. More specifically, California, USA stands among the top five states for motor vehicle collisions and was ranked first with respect to pedestrian traffic fatalities, with 867 in 2016.[1] Compared to vehicle-to-vehicle collisions, pedestrian-involved crashes typically result in more severe injuries and fatalities. In fact, pedestrians are threatened by a higher risk of injuries and death, especially in poorly designed roadways with less consideration for pedestrian safety. Although incorporating safety policies into traffic operations, such as law enforcement strategies, has resulted in some safety improvements, the trendline of pedestrian fatalities still illustrates a steady increase over the past 10 years (from 4,699 fatalities in 2007 to 5,987 fatalities in 2016).[1,2]

HSIS First Place Safety Data Award

First place winning paper of the Federal Highway Administration's (FHWA) 2019 Excellence in Highway Safety Data Award, which is designed to encourage university students to use Highway Safety Information System (HSIS) data to investigate a topic that advances highway safety and to develop a paper to document the original research. The HSIS Highway Safety Data Awards Program is jointly administered by FHWA and the Institute of Transportation Engineers.

HSIS
Highway Safety Information System

Considering the importance of pedestrian safety to society, the study of pedestrian crashes is essential to mitigate the risk of injuries, loss of life, and societal costs. It also has a considerable impact on transportation networks capacity reduction caused by congestion.[3] Since the circumstances around the crashes are not unique, the study of pedestrian safety calls for a detailed investigation. Recent studies have addressed pedestrian-involved crashes with a specific focus on injury severity.[4-8] This research has disaggregated crash data by the number of vehicles and pedestrians involved, and demographic and socioeconomic characteristics, age, weather conditions, location of interest, etc. However, the relationship between crash severities and contributing factors, along with time of the day, are not clearly understood.

To provide some insight, the present study seeks to identify the contributing factors of pedestrian-involved crash injury severity by time of day in California (daytime and nighttime crashes) through a multinomial logit-modeling framework. In order to determine if such crashes need to be considered independently for safety analyses, a parameter transferability test will be conducted. In the course of parameter transferability, the null hypothesis will test whether parameter estimates by time-of-week are statistically different (i.e., daytime and nighttime should be modeled holistically). Therefore, this study aims to determine if the null hypothesis is rejected (i.e., injury severity models by time of day need to be considered independently) while identifying injury severity contributing factors by time of day.

The application of machine learning (ML) techniques in data mining has seen explosive growth in recent years and has garnered interest from a broadening variety of research domains such as safety studies. Support vector machine (SVM) is one of the most popular models, which has shown better prediction outcomes for injury severity analysis compared to conventional statistical models.[9,10] In this paper, SVM is employed to predict injury severity outcomes of pedestrian crashes for compression purposes.

Data Description

Data used for the analysis consisted of pedestrian-involved crashes that occurred in California from 2010 to 2014. Two crash files (accident and vehicle) were merged based on the variable "caseno," and the final dataset was filtered based on the variable "severity," which represents the most severe injuries in daytime and nighttime pedestrian-involved crashes.[11] According to the average times for sunset and sunrise conditions for the state of California,[12] two time periods, from 6:00 a.m. to 7:59 p.m. and 8:00 p.m. to 5:59 a.m., were considered for daytime and nighttime conditions. After identifying these crashes and performing data cleaning, 8,573 crashes remained for model development, which consisted of 4,910 and 3,663 daytime and nighttime crash records, respectively.

The Highway Safety Information System (HSIS) data classify injury severity into five distinct categories: 1) fatal; 2) severe or incapacitating injury; 3) visible or non-incapacitating injury; 4) complaint of pain; and, 5) property damage only (PDO) or no injury. The obtained crashes were composed of a total number of 1,585 (18.5 percent) fatal, 1,697 (19.8 percent) incapacitating injury, 2,512 (29.3 percent) non-incapacitating injury, 2,325 (27.1 percent) complaint of pain, and 454 (5.3 percent) PDO crashes. Ascertaining enough observations in each severity category of the same severity levels were considered in the modeling procedure.

Several indicator variables were created for each disaggregated dataset of daytime and nighttime crashes. Of the created indicator variables, 48 and 58 variables were found to be significant contributing factors for daytime and nighttime crashes, respectively.

Modeling Framework

Multinomial Logit Model (MNL). The multinomial logit (MNL) model is the most widely applied discrete-outcome modeling approach in safety analysis. In general, in MNL models, crash is addressed in terms of injury severity outcomes in the sense that the propensity of crash i towards severity category k is represented by severity propensity function, T_{ki}, as shown in Eq. (1).[12]

$$T_{ki} = \alpha_k + \beta_k X_{ki} + \varepsilon_{ki} \tag{1}$$

In the above equation, α_k is a constant parameter for crash severity category k; β_k is a vector of the estimable parameters for crash severity category k; $k = 1, \ldots,$ ($k = 5$ in this paper) representing severity levels. X_{ki} represents a vector of explanatory variables affecting the crash severity for i at severity category k (driver characteristics, environmental conditions, temporal variables, etc.); ε_{ki} is a random error term following the Type I generalized extreme value (i.e., Gumbel) distribution; and $i = 1, \ldots, n$ where n is the total number of crash events included in the model.

Eq. (2) shows how to calculate the probability for each crash severity category. Let $P_i(k)$ be the probability of accident i ending in crash severity category k, such that:

$$P_i(k) = \frac{\exp(\alpha_k + \beta_k X_{ki})}{\Sigma_{\forall k} \exp(\alpha_k + \beta_k X_{ki})} \tag{2}$$

The coefficients β_k, can be estimated by the maximum likelihood method.

In order to determine the effect of explanatory variable X on the outcome probability $P_i(k)$ of injury severity i, marginal effects are computed. The marginal effect is the difference in probabilities of a severity outcome for a one-unit change in an explanatory variable (i.e., from zero to one), while all others remain constant.[14, 15]

Parameter Transferability. Following the procedure found in (*14*) and (*16*), a log-likelihood ratio test is conducted to evaluate model transferability by time of day through the Eq. (3).

$$x^2 = -2[LL(\beta_{MX1_{MX2}}) - LL(\beta_{MX1})] \quad (3)$$

where $LL(\beta_{MX1_{MX2}})$ is the log-likelihood at convergence of model MX_1 using data for model MX_2 and $LL(\beta_{MX1})$ is the log-likelihood at convergence of model MX_1. Supposing the model for daytime crashes is fit using data from nighttime crashes and vice versa, then the original log-likelihood values are used to calculate the chi-square statistic. Finally, by considering the degree of freedom (which is the number of estimated parameters in the model using the other model's data), the significance is determined.

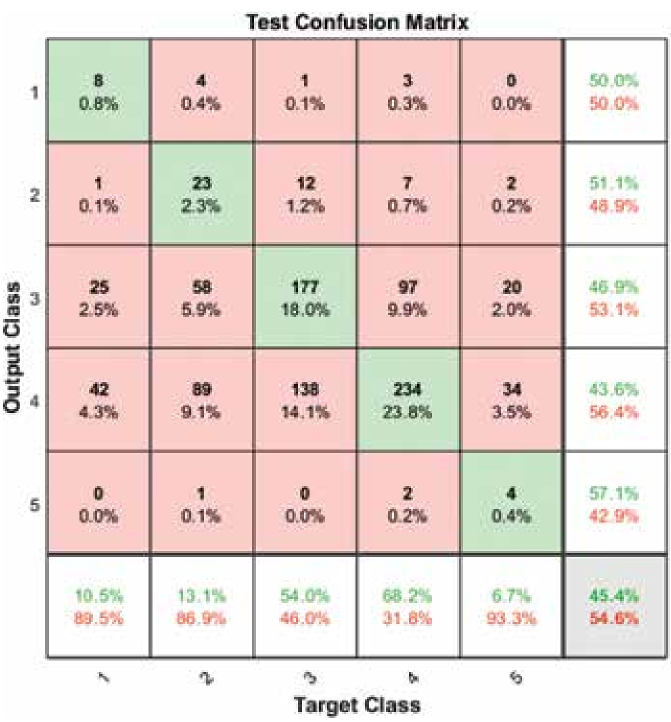
Support Vector Machine (SVM). SVM is a supervised learning approach that represents the instances as a set of points in N-dimensional (i.e., severity levels) space. It then generates a (N-1) dimensional hyperplane to separate those points into groups. The ultimate goal is finding the line (i.e., hyperplane of pattern *x*) illustrated in Eq. (4), while simultaneously maximizing the margin between the linear decision boundaries. Hyperplane $y(x) = 0$ defines a decision boundary in the feature space, while the parameters of a normal vector (w) and bias (*b*) are determined through the learning procedure on a training set (x_1, y_1) to (x_n, y_n). In this study, the training input includes all crash-related explanatory variables (x_n), while the training output represents injury severity outcomes (y_n).

$$y(x) = w^T x + b = 0 \quad (4)$$

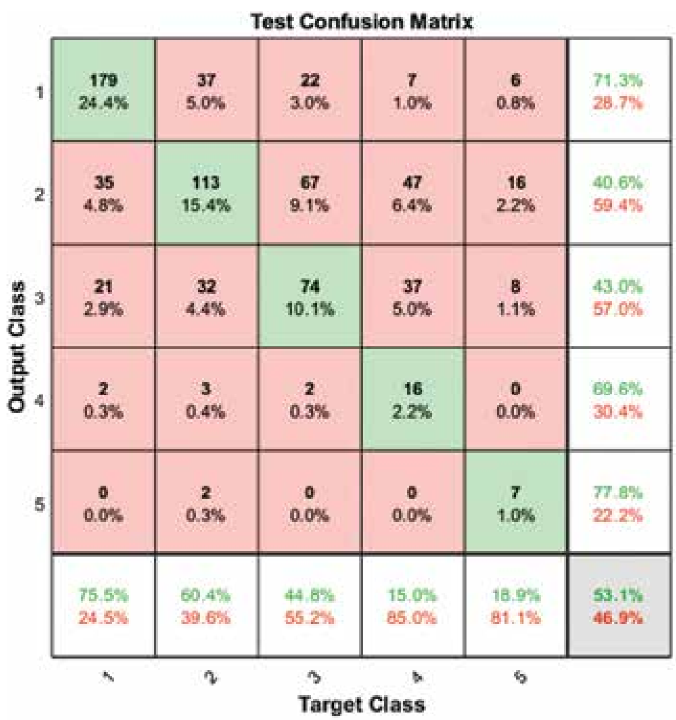
Construction of the higher dimensional space by the SVM model is based on the concept of a kernel function. The kernel function is applied to data in the original space and are defined as $K(x_i, x_j) \equiv \Phi(x_i)^T \Phi(x_j)$. Different kernel functions have been proposed in the domain of SVM; however, one of the most commonly applied, the Gaussian Radial Basis (RBF), which demonstrated better results in related works, is used in this study.[10,17]

Model Estimation Results. The MNL model was estimated using the statistically significant independent variables. The best fit model specifications and corresponding marginal effects for the three most impactful variables in each severity level are summarized in Table 1 and Table 2 for daytime and nighttime models, respectively. The estimated coefficients demonstrate the independent effects of contributing factors on severity levels; however, to further make sense of the results, we may rely on marginal effects.

The same datasets are used in SVM models, and the prediction results are illustrated in the confusion matrix format in Figure 1. A confusion matrix is a tabular layout that describes the performance of supervised learning classification models. In the matrix, a column represents the true class (target class) instances, whereas the row represents a predicted class (output class) instances. In this study, the dataset was split randomly into sub-datasets, as 80 percent for training and 20 percent for testing. As shown in Figure 1, the developed models could test data with the overall classification accuracy of 45.4 percent for daytime and 53.1 percent for nighttime crashes.



Daytime Model



Nighttime Model

Figure 1. Results of SVM Confusion Matrices.

Table 1. Best Fit Multinomial Logit Estimations for Daytime Crashes

Variable	Coefficient	t-statistic	Marginal Effects				
			Fatal	Severe Injury	Visible Injury	Complaint of Pain	PDO
Fatal							
Constant	-2.82	-10.21					
Driver Sobriety (1 If Had Been Drinking, Not Under Influence, 0 Otherwise)	0.88	1.80	0.048	-0.005	-0.008	-0.017	-0.018
Weather Condition (1 if Clear, 0 Otherwise)	0.62	3.20	0.034	-0.003	-0.006	-0.012	-0.013
Vehicles Involved (1 if Parked Motor Vehicle, 0 Otherwise)	-1.35	-2.62	-0.073	0.007	0.012	0.026	0.028
Severe Injury							
Constant	-1.62	-13.18					
Roadway Type (1 if Rural Freeways, 0 Otherwise)	1.43	5.97	-0.007	0.110	-0.021	-0.042	-0.040
Pedestrian Action (1 if Crossing in Crosswalk not at Intersection, Otherwise)	-0.72	-1.66	0.004	-0.056	0.010	0.021	0.020
Accident Type (1 if Sideswipe, 0 Otherwise)	-1.02	-3.04	0.005	-0.078	0.015	0.030	0.029
Visible Injury							
Constant	-0.62	-4.41					
Weather Condition (1 if Raining, 0 Otherwise)	0.30	1.65	-0.003	-0.004	0.039	-0.016	-0.016
Driver Sex (1 if Male, 0 Otherwise)	0.29	3.17	-0.003	-0.004	0.038	-0.015	-0.016
Type of vehicle at fault (1 if Trk Trac W/1 Trlr, 0 Otherwise)	-1.23	-2.31	0.011	0.018	-0.162	0.066	0.067
Complaint of Pain							
Constant	-0.18	-2.80					
Driver Sex (1 if Male, 0 Otherwise)	0.24	3.36	-0.005	-0.007	-0.013	0.053	-0.028
Light Condition (1 if Dusk-Dawn, 0 Otherwise)	0.24	1.68	-0.005	-0.007	-0.013	0.053	-0.029
Roadway Type (1 if Urban Freeways, 0 Otherwise)	-0.17	-2.30	0.003	0.005	0.009	-0.038	0.020
Property Damage Only							
Location Type (1 if Intersection. 0 Otherwise)	0.23	2.67	-0.005	-0.007	-0.013	-0.028	0.051
Driver Sobriety (1 if Had Been Drinking, Impairment Unknown, 0 Otherwise)	-0.80	-2.98	0.016	0.023	0.044	0.095	-0.178
Light Condition (1 if Dark-No Street Lights, 0 Otherwise)	-0.69	-4.97	0.014	0.019	0.038	0.082	-0.153
Model Statistics							
Number of Observations	4,910						
Log Likelihood Function	-6,690.088						
Constants Only Model	-6,902.219						
Overall Prediction Accuracy	37.82%						

Table 2. Best Fit Multinomial Logit Estimations for Nighttime Crashes

Variable	Coefficient	t-statistic	Marginal Effects				
			Fatal	Severe Injury	Visible Injury	Complaint of Pain	PDO
Fatal							
Constant	-0.76	-2.26					
Roadway Type (1 if Rural Multilane Divided Non-Freeways, 0 Otherwise)	1.90	5.20	0.081	-0.025	-0.021	-0.021	-0.014
Weather Condition (1 if Raining, 0 Otherwise)	-1.37	-2.83	-0.058	0.018	0.015	0.015	0.010
Accident Type (1 if Head-On, 0 Otherwise)	-3.11	-3.02	-0.132	0.041	0.033	0.035	0.023
Severe Injury							
Constant	0.45	1.97					
Accident Type (1 if Auto-Pedestrian, 0 Otherwise)	2.64	6.75	-0.035	0.502	-0.199	-0.159	-0.109
Accident Type (1 if Head-On, 0 Otherwise)	-1.08	-2.45	0.014	-0.205	0.081	0.065	0.045
Weather Condition(1 if Raining, 0 Otherwise)	-1.85	-6.21	0.024	-0.352	0.139	0.112	0.077
Visible Injury							
Constant	0.71	2.98					
Driver Sobriety (1 if Had Been Drinking, Impairment Unknown, 0 Otherwise)	0.72	4.51	-0.008	-0.055	0.131	-0.042	-0.027
Location Type (1 if Intersection, 0 Otherwise)	-0.66	-3.31	0.007	0.050	-0.119	0.038	0.024
Weather Condition (1 if Raining, 0 Otherwise)	-0.65	-3.05	0.007	0.049	-0.118	0.038	0.024
Complaint of Pain							
Constant	1.54	6.66					
Pedestrian Action in the Crash (1 if Crossing in Crosswalk at Intersection, Otherwise)	0.38	3.38	-0.004	-0.023	-0.022	0.064	-0.014
Weather Condition (1 if Raining, 0 Otherwise)	-0.96	-4.36	0.011	0.058	0.056	-0.161	0.036
Accident Type (1 if Hit Object, 0 Otherwise)	-0.69	-2.31	0.008	0.042	0.040	-0.116	0.026
Property Damage Only							
Roadway Type (1 if Urban Multilane Undivided Non-Freeways, 0 Otherwise)	1.24	5.41	-0.009	-0.051	-0.046	-0.046	0.152
Accident Type (1 if Sideswipe, 0 Otherwise)	0.82	3.72	-0.006	-0.034	-0.030	-0.031	0.101
Vehicles Involved (1 if Parked Motor Vehicle, 0 Otherwise)	-0.93	-2.47	0.007	0.039	0.035	0.035	-0.115
Model Statistics							
Number of Observations	3,663						
Log Likelihood Function	-4,997.495						
Constants Only Model	-5,426.538						
Overall Prediction Accuracy	41.35%						

In order to simplify the discussion, the significant contributing factors found in MNL models (Tables 1 and 2) that influenced weekday and weekend injury severity will be discussed separately. This is followed by a prediction accuracy comparison of the MNL and SVM models and results from the parameter transferability test.

Discussion

Daytime Injury Severity Model. Factors with the largest impact (positive or negative) on fatal, severe injury, visible injury, complaint of pain, and PDO crashes include parked motor vehicle in crash, rural freeways roadway type, truck at fault, dusk-down weather condition, and driver sobriety, respectively.

In regard to the fatal category, crashes in which a parked motor vehicle were involved have a 0.073 lower probability according to marginal effects. A possible explanation for this finding may be attributed to lower speed limits in urban areas and consequently, a lower chance of fatality. In addition, pedestrian actions such as stepping from in between parked cars and running into the street are recognized as the most common contributions to crash occurrence when a parked vehicle is involved.[18] As for crashes that occurred on rural freeways, the marginal effects show a 0.110 higher probability of sustaining a severe injury. These findings are reasonable due to higher speed limits on rural roadways and less traffic volume, which tempts drivers to speed on rural roadways. This is consistent with the results found in previous injury severity studies which implied that higher/lower speed impacts lead to more/less severe injuries.[4,14,19.]

In terms of vehicle type, it has been demonstrated that single trailer trucks have a 0.162 lower probability of visible injuries. For complaint of pain, the higher and lower impacts were found by dusk-down light conditions and urban freeway roadway type. While dusk-down variables increase the probability of complaint of pain by 0.053, crashes that occur on urban freeways are associated with the probability of -0.038. Transitions to and from darkness are the most critical times of day for pedestrian-involved crashes as a driver's view and vision are compromised.[20] Although the urban freeway variable shows the lower probability of 0.038 for complaint of pain, traveling at a higher speed on urban freeways increases the severe injury probability, which agrees with the findings of studies by Uttley et al and Das et al.[19,21] It has been also shown that speed change is a significant factor for accident detection.[22] Drivers who had been drinking (impairment unknown) and dark with no street lights conditions are the most significant impactful variables for PDO crashes which show negative impacts of PDO crashes. Drinking and driving has been recognized as a primary reason for different types of crashes including pedestrian-involved crashes[23] which results in more severe crashes. The results of a study of pedestrian crashes in North Carolina that found that drinking and driving and darkness with no streetlights more than doubles the risk of fatal injury.[24]

Nighttime Injury Severity Model. As shown in Table 2, head-on accident for fatal crashes, rainy weather condition for severe injury, intersection location for visible injuries, pedestrian crossing in crosswalk at intersection for compliant of pain, and urban multilane undivided non-freeway roadway type in PDO crashes are the most impactful and significant contributing factors.

Head-on crashes significantly decrease the probability of fatal crashes by 0.132. Considering that 87.2 percent of pedestrian crashes investigated in this paper occurred on urban roadways, a possible explanation for this finding may be related to lower posted speed limits, as mentioned earlier. In addition, the protective effect of safety equipment like seatbelts and airbags may be another possible reason for saving lives, but can be a cause of injury as well.[25] Rainy weather conditions had a decreased impact on severe injury. Results showed that the rain condition leads to a decreased likelihood of 0.352 in a severe injury crash. This might be attributed to the fact that drivers are more cautious when experiencing a slippery road surface and is consistent with the results of a recent study that showed that serious crashes rarely occurred during the rainy season.[26] However, the weather issue may remain controversial since it depends on wide range of crash, environmental and driver characteristics.

As for the visible injury condition, crashes that occurred at intersections showed a negative effect on visible injury with a probability of 0.119, which might be reflective of lower speed limits at intersections. This is also consistent with the results found in studies by Savolainen et al and Pahukula et al.[25,27] The variable indicates pedestrians crossing a crosswalk at an intersection is associated with a higher probability of being a complaint of pain by 0.064. However, this shows a decrease in impacts for all other severity levels. The results of recent studies show that speed variation at intersection locations has a significantly impact on crash frequency and injury severity.[28,29] Driver awareness (or lack thereof) of the presence of pedestrians at a crosswalk can also be a possible explanation for crashes that occur at crosswalk locations, which tend to be less severe compared to roadways. The same results were obtained in previous works.[4,30] In terms of PDO crashes, urban multilane undivided non-freeway roadways increase the probability of PDO crashes by 0.152. Although the lower posted speed limits on this type of roadway reduce the probability of fatal and severe injury by 0.009 and 0.051, careless driving behaviors as a results of in-vehicle distraction potentials (such as adjusting audio system, vehicle controls, and navigation system, or texting while driving and talking on the phone) may be possible reasons. A recent study conducted by Vasebi et al.,[31] shed light on the impacts of in-vehicle automated technologies on accident reduction and the corresponding traffic consequences.

Prediction Comparison: SVM vs. MNL. This study investigated the superior prediction performance of MNL and SVM models based on the measure of overall models accuracy (i.e., the percentage of correctly predicted instances over the total number of observations). Results from SVM models indicate that it can provide a higher prediction accuracy of 20.00 percent for daytime and 28.41 percent for nighttime crashes. These accuracy levels can be further improved through incorporating Artificial Intelligence (AI) techniques and evolutionary strategies for parameter tuning of SVM.[10]

In the context of machine learning, in multi-class classification problems with imbalanced data, in addition to the overall model accuracy, AUC (the area under the Receiver Operating Characteristic (ROC) curve) is the most widely used metric for evaluating the quality of model's prediction.[32] The AUC results of the ROC analysis, 0.7542 for daytime and 0.7937 for nighttime models (the value varies between 0.5 and 1, where 1 belongs to a perfect classifier), provides additional insights of SVM prediction performance. In addition to the overall model prediction accuracy and AUC results, other prediction criterion including average accuracy (i.e., the accuracy per severity levels), sensitivity, and specificity of SVM models are presented in Table 3.

Table 3. Prediction Metrics of SVM Models

SVM Models	Accuracy*	Sensitivity*	Specificity*
Daytime	70.73%	49.76%	75.79%
Nighttime	75.29%	60.47%	81.58%

** Higher values indicate better prediction performance.*

Possible reasons for the better performance of SVM compared to MNL can be identified based on the mechanism of the two methods. Even though conventional statistical models are the primary method used in crash severity analyses, they suffer from the pre-assumption of data distribution and model structure.[19] These assumptions may not necessarily exist in crash data and may produce wrong estimations together with incorrect inferences. The ML models, on the other hand, do not make any assumptions about the underlying data distribution. Without the assumptions of statistical models, machine learning techniques are perhaps the best tools for modeling the potentially nonlinear relationship between crash severity outcomes and the related factors.

Parameter Transferability. Once the models were developed, the log-likelihood ratio tests were conducted to determine if parameter estimates were statistically different from daytime to nighttime. From Eq. (3), the x^2 statistic for the daytime ($x^2 = 2752.6704$), and the nighttime models ($x^2 = 4018.5725$) with corresponding degrees of freedom equal to the summation of the number of estimated parameters ($d.f = 48$) and ($d.f = 58$), provides a confidence level well over 99 percent. That is to say, we can reject the null hypothesis which states that there is no difference between the model parameters, and both need to be modeled holistically.

Summary and Conclusion

This study attempted to provide insight into factors affecting injury severity in daytime and nighttime pedestrian-involved crashes. To accomplish this, a large number of crash records from the HSIS dataset for California were used, and a multinomial logit modeling framework was applied. Furthermore, a parameter transferability test was conducted to determine if pedestrian-involved injury severity analyses need to be conducted by time of day. As an example of machine learning techniques, which has been increasingly implemented in transportation research as well as safety research, SVM models were also employed for comparison purposes.

For factors that were found to be significant in both daytime and nighttime models, variable impacts on severity outcomes were obtained. There are many instances of this, which needs to be investigated in depth. For example, rainy weather condition results in a 0.039 increase in the likelihood of visible injury in the daytime model, whereas such weather conditions in the nighttime model result in a 0.118 decrease in the likelihood of a visible injury. One possible explanation could be the combination of the light condition, in which the sight distance is restricted, as well as the weather conditions, which affects the friction between the tire and the roadway's surface. These may lead drivers to take more precautions under dark lighting conditions, especially when it is raining.

In addition to the difference in contributing factors between severity models with a high level of confidence (99 percent), the null hypothesis indicating that daytime and nighttime crashes need to be modeled holistically was rejected. A certain number of statistically significant variables were found to be exclusive in each classification, such as type of vehicle involved in crash in fatality or weather condition in severe injuries. This also implies that injury severity factors differ by time of day, and two severity models need to be considered separately for safety analysis.

Aside from the investigation of contributing factors, the performance of the MNL models was compared with the SVM models in this study. It was found that SVM models produced better prediction performance for crash injury severity than the MNL model. The percent of correct prediction for the SVM model was 45.4 percent and 53.1 percent for daytime and nighttime models, which was higher than that produced by the MNL model (37.82 percent and 41.35 percent). These results suggest that the SVM model is a promising tool for crash injury severity analysis and can be employed in future research. **itej**

Acknowledgment

The authors would like to acknowledge Anusha Patel Nujjetty, manager at the Highway Safety Information System Laboratory and all of the staff involved in providing accident data used in this study. In addition, support provided by ABC-UTC at Florida International University (FIU) is appreciated.

References

1. NHTSA. 2016 Pedestrians Traffic Safety Fact Sheet. *National Highway Traffic Safety Administration, U.S. Department of Transportation*, No. DOT HS 812 493, 2018.
2. Retting, R., H. Rothenberg, and S. Schwartz. Pedestrian traffic fatalities by state: 2017 preliminary data. Governors Highway Safety Association, Washington, DC, 2018.
3. Almotahari, A., M. A. Yazici, S. Mudigonda, and C. Kamga. Analysis of Incident-Induced Capacity Reductions for Improved Delay Estimation. *Journal of Transportation Engineering, Part A: Systems,* Vol. 145, No. 2, 2018, p. 04018083.
4. Pour-Rouholamin, M. and H. Zhou. Investigating the risk factors associated with pedestrian injury severity in Illinois. *Journal of Safety Research,* Vol. 57, 2016, pp. 9–17.
5. Kim, M., S.-Y. Kho, and D.-K. Kim. Hierarchical ordered model for injury severity of pedestrian crashes in South Korea. *Journal of Safety Research,* Vol. 61, 2017, pp. 33–40.
6. Koopmans, J. M., L. Friedman, S. Kwon, and K. Sheehan. Urban crash-related child pedestrian injury incidence and characteristics associated with injury severity. *Accident Analysis & Prevention,* Vol. 77, 2015, pp. 127–136.
7. Li, D., P. Ranjitkar, Y. Zhao, H. Yi, and S. Rashidi. Analyzing pedestrian crash injury severity under different weather conditions. *Traffic Injury Prevention,* Vol. 18, No. 4, 2017, pp. 427–430.
8. Haleem, K., P. Alluri, and A. Gan. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accident Analysis & Prevention,* Vol. 81, 2015, pp. 14–23.
9. Li, Z., P. Liu, W. Wang, and C. Xu. Using support vector machine models for crash injury severity analysis. *Accident Analysis & Prevention,* Vol. 45, 2012, pp. 478–486.
10. Mokhtarimousavi, S., J. C. Anderson, A. Azizinamini, and M. Hadi. Improved Support Vector Machine Models for Work Zone Crash Injury Severity Prediction and Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2019.
11. Nujjetty, A., S. Sharma, and F. Council. *Guidebook for Data Files: Minnesota.* Washington, DC, 2014.
12. Timeanddate. *Sunrise, Sunset, and Daylength*. https://www.timeanddate.com/sun/usa/california-city?month=7&year=2018. Accessed Jannuary 10, 2019.
13. Kim, J.-K., G. F. Ulfarsson, V. N. Shankar, and S. Kim. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention,* Vol. 40, No. 5, 2008, pp. 1695–1702.
14. Anderson, J. C. and S. Dong. Heavy-vehicle driver injury severity analysis by time of week: a mixed logit approach using HSIS crash data. Institute of Transportation Engineers. *ITE Journal,* Vol. 87, No. 9, 2017, p. 41.
15. Washington, S. P., M. G. Karlaftis, and F. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall/CRC, 2010.
16. Anderson, J. and S. Hernandez. Roadway classifications and the accident injury severities of heavy-vehicle drivers. *Analytic Methods in Accident Research,* Vol. 15, 2017, pp. 17–28.
17. Yu, R. and M. Abdel-Aty. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science,* Vol. 63, 2014, pp. 50–56.
18. Ulfarsson, G. F., S. Kim, and K. M. Booth. Analyzing fault in pedestrian–motor vehicle crashes in North Carolina. *Accident Analysis & Prevention,* Vol. 42, No. 6, 2010, pp. 1805–1813.
19. Chen, C., G. Zhang, Z. Qian, R. A. Tarefder, and Z. Tian. Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accident Analysis & Prevention,* Vol. 90, 2016, pp. 128–139.
20. Uttley, J. and S. Fotios. The effect of ambient light condition on road traffic collisions involving pedestrians on pedestrian crossings. *Accident Analysis & Prevention,* Vol. 108, 2017, pp. 189–200.
21. Das, A. and M. Abdel-Aty. A genetic programming approach to explore the crash severity on multi-lane roads. *Accident Analysis & Prevention,* Vol. 42, No. 2, 2010, pp. 548–557.
22. Parsa, A. B., H. Taghipour, S. Derrible, and A. K. Mohammadian. Real-time accident detection: coping with imbalanced data. *Accident Analysis & Prevention,* Vol. 129, 2019, pp. 202–210.
23. Mann, R. E., G. Stoduto, E. Vingilis, M. Asbridge, C. M. Wickens, A. Ialomiteanu, J. Sharpley, and R. G. Smart. Alcohol and driving factors in collision risk. *Accident Analysis & Prevention,* Vol. 42, No. 6, 2010, pp. 1538–1544.
24. Kim, J.-K., G. F. Ulfarsson, V. N. Shankar, and F. L. Mannering. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis & Prevention,* Vol. 42, No. 6, 2010, pp. 1751–1758.
25. Savolainen, P. and F. Mannering. Probabilistic models of motorcyclists' injury severities in single-and multi-vehicle crashes. *Accident Analysis & Prevention,* Vol. 39, No. 5, 2007, pp. 955–963.
26. Huang, H., M. Abdel-Aty, and A. Darwiche. County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2148, 2010, pp. 27–37.
27. Pahukula, J., S. Hernandez, and A. Unnikrishnan. A time of day analysis of crashes involving large trucks in urban areas. *Accident Analysis & Prevention,* Vol. 75, 2015, pp. 155–163.
28. Kamrani, M., R. Arvin, and A. J. Khattak. Extracting useful information from Basic Safety Message Data: an empirical study of driving volatility measures and crash frequency at intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2018, p. 0361198118773869.
29. Arvin, R., M. Kamrani, and A. J. Khattak. How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data. *Accident Analysis & Prevention,* Vol. 127, 2019, pp. 118–133.
30. Kim, J.-K., G. F. Ulfarsson, S. Kim, and V. N. Shankar. Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention,* Vol. 50, 2013, pp. 1073–1081.
31. Vasebi, S., Y. M. Hayeri, C. Samaras, and C. Hendrickson. Low-level automated light-duty vehicle technologies provide opportunities to reduce

fuel consumption. *Transportation Research Record Journal of the Transportation Research Board,* Vol. 2672, No. 24, 2018, pp. 60–74.

32. Tharwat, A. Classification assessment methods. *Applied Computing and Informatics*, 2018.

Seyedmirsajad Mokhtarimousavi (S) *is a Ph.D. candidate in the Civil and Environmental Engineering Department at Florida International University in Miami, FL, USA. He received his masters of science in transportation planning from IAU, Science and Research Branch, Tehran, Iran in 2013. His primary focus is on transportation systems planning, travel behavior analysis, and transportation safety. The focus of his Ph.D. is on econometric modeling, data science, and analytics related to transportation safety topics. He is currently working on a project called "Work Zone Safety Analysis, Investigating Benefits from Accelerated Bridge Construction (ABC) on Roadway Safety" under the supervision of Professors Atorod Azizinamini and Mohammed Hadi.*