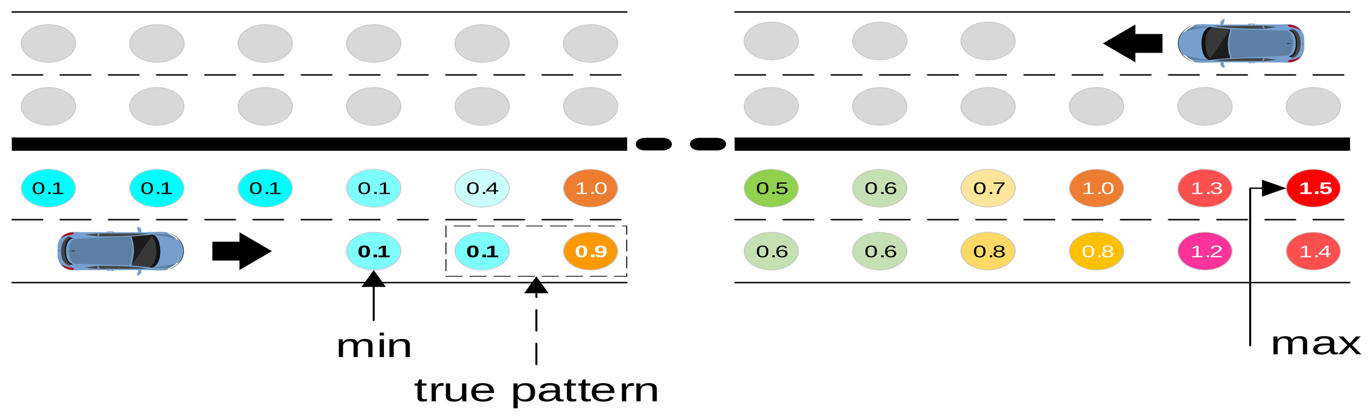
A Machine Learning-based Study of Lighting Patterns Contributing to Nighttime Crash Severity on Roadway Corridors

Zhenyu Wang[1,*], Mingchen Li[2], Pei-Sung Lin[1], Runan Yang[1], Abhijit Vasili[1], and Edith Wong[3]
[1]Center for Urban Transportation Research, University of South Florida, [2]Department of Electrical Engineering, University of South Florida, [3]Florida Department of Transportation
* Corresponding author – zwang9@cutr.usf.edu

Introduction

Roadway lighting pattern is a significant factor that influences nighttime safety. Since the illumination distribution along roadway corridors presents an intricate pattern, the traditional photometric measure, such as average horizontal illuminance and ratio-based uniformity, cannot capture the patterns that truly contribute to the injury severity of a nighttime crash on roadway corridors. This study aimed to develop a machine learning model for accurately predicting the crash injury severity based on horizontal illuminance patterns and other features, such as geometric, traffic control, and environmental factors.



Example of Illuminance Patterns Influencing Drivers' Vision

Data Preparation

Lighting data:

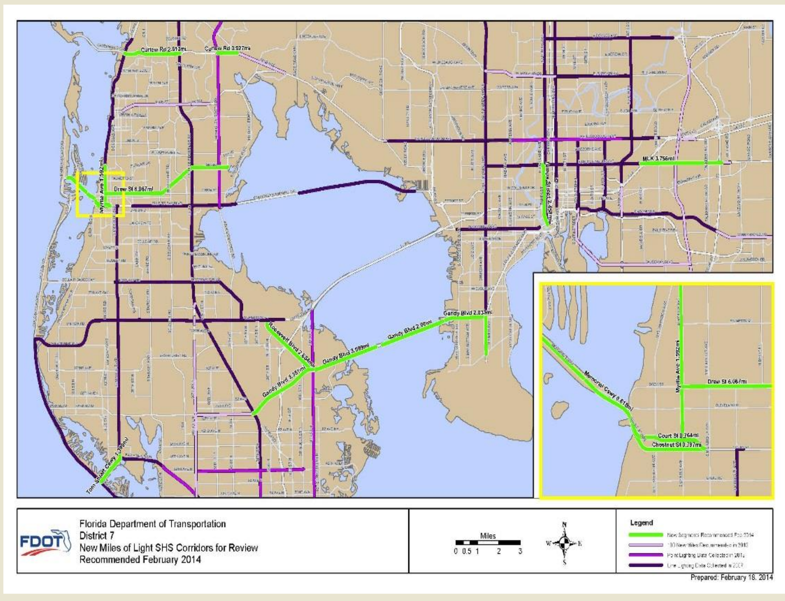
"Big" Lighting data were collected using the Advanced Lighting Measurement System (ALMS) :

- Horizontal illuminance (foot-candles)
- Two measurement points every 10ft per lane
- Completed 400 center-miles measurement in Florida

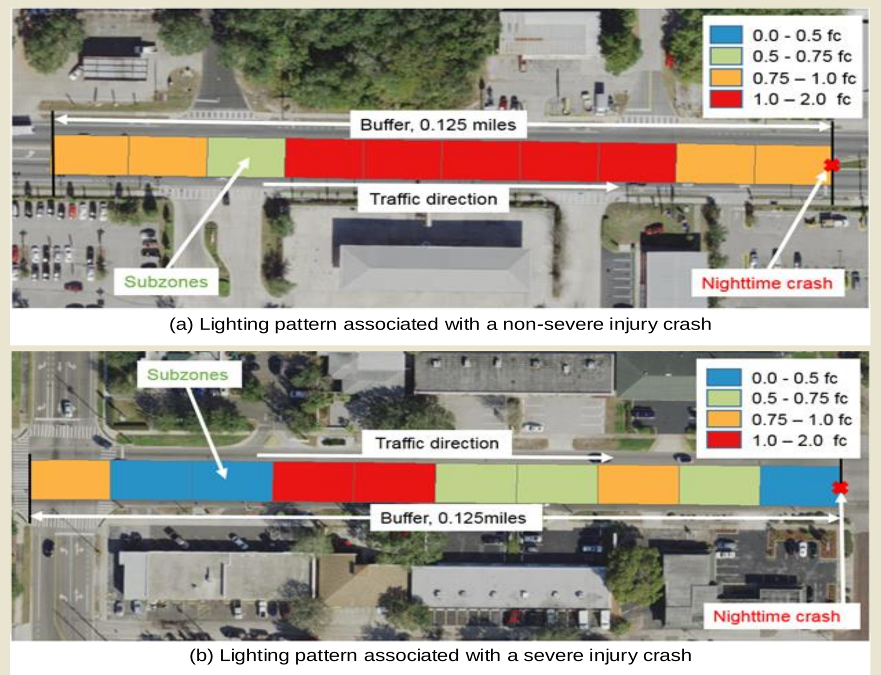
Crash data:

2,129 nighttime crashes were collected and matched to measured lighting data for 2011-2014:

- Create an upstream buffer of 1200 ft. for each crash
- Divide the buffer into 20 sub-zones
- Calculate lighting statistic for each sub-zone to describe lighting patterns associating with each crash
    - Mean & Standard Deviation



Roadway Corridors with Street Lighting Measurement



Lighting Buffers and Sub-zones Associated with Nighttime Crashes

Technical Approach

Identify significant non-lighting features with Probit model → Re-sampling → SVM Training with Parameter Grid Search → "Best" Model with highest F1 Score → SVM Testing → Final Model

Identified Non-lighting Features with Probit Model

Feature	Codes	Coef.	Z-test
SEVERITY	1 - severe injury (fatal or incapacitating injury) 0 – others		
HFC MEAN	Average horizontal illuminance (fc) of sub-zones		
HFC SD	Average horizontal (fc) standard deviation of sub-zones		
MULTIPLE LANE	1 - number of lanes > 4, 0 – others	-2.18329	-12.45
HIGH SPEEDDLMT	1 - speed limit >= 40 mph , 0 - others	0.18082	2.18
TWLTL	1 - Two-way-left-turn-lane median, 0 - others	0.40805	2.93
WEEKEND	1 – weekend, 0 - weekdays	0.32431	1.86
ALCHOL/DRUG	1 - alcohol or drug involved, 0 – none	-0.20208	-2.3
AGGRESSIVE	1 - aggressive driving involved, 0 – none	0.38913	4.07
TEEN-DRIVER	1 - teenage driver involved, 0 - none	0.54863	1.95
NO SAFETY BELT	1 - no belt used by one more person, 0 - belt used	0.25048	1.78
SPEEDING	1 – speeding, 0 - no speeding	0.41887	4
HIT AND RUN	1 - hit and run, 0 - no hit and run	0.41168	1.82
YOUNG-INJURED	1 - the highest-injured person is younger than 20, 0 – others	-0.43402	-3.36
PEDD-BIKE	1 - pedestrian or bicycle involved, 0 - others	-0.58025	-4.64
MOTORCYCLE	1 - motorcycle involved, 0 - others	1.71331	11.63
REAREND	1 - rear-end crash, 0 - others	1.15367	9.28
ANGLE	1 - angle crash, 0 - others	0.2325	1.78
HEADON	1 - head on crash, 0 - others	0.5134	3.99
HITFIXED	1 - hit on roadside fixed objects, 0 - others	0.82825	3.38

Feature Selection:

- Non-lighting features that significantly contribute to crash injury

Re-sampling:

- Crash data are unbalanced (non-severe data is predominant, 648: 1481)
- Avoid under-fitting issues
- Randomly match **one** severe-injury with **two** non-severe injury data

Model Training with Grid Search:

- 80% samples
- SVM kernel parameter (*g*), and SVM regularization parameter (*c*):

Model Selection with Highest F-1 Score:

- Consider both true and false positive/negative
- More realistic measurement of models

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$Precision\ (p) = \frac{True\ positive}{True\ positive + False\ positive} \qquad Recall\ (r) = \frac{True\ positive}{True\ positive + False\ Negative}$$

Testing Result

(accuracy)	Trained SVM Models	
Experiment	**Lighting statistics only**	**Lighting + Non-lighting Features**
1	66.66	74.4
2	66.66	69
3	66.92	69.5
4	66.66	74.7
5	66.66	73.1
6	66.66	72.6
7	66.66	74.7
8	66.66	70.8
9	66.66	73.6
10	66.66	75
Average	66.686	**72.74**

- Each experiment represents an SVM model with random training samples
    - Re-sampling on small dataset
- Multiple experiments provide more convincing prediction results

Case Study

	Item	Crash I	Crash II
Crash Features	True severity	Possible Injury	Fatal
SVM Model	Buffer length	0.225 miles	
	Number of sub-zones	20	
	Predicted probability of "1"	5.6%	73.3%
	Predicted severity	0 - Non-severe	1 - Severe
Lighting Statistic (whole segment)	Average horizontal illuminance (fc)	1.298	0.705
	Ave/Min	11.4	43.5
	Max/Min	19.5	134.5

Conclusion

- The developed SVM model can effectively predict crash injury severity with limited features
    - Crash data absent during-crash many features that directly contribute to injury severity
- Re-sampling and voting prediction are beneficial to improve prediction performance for small and unbalanced crash data
- Lighting statistics (mean and SD) of 20 sub-zones (30 feet zone length) have limitations:
    - Correlated to other non-lighting features
    - Still to rough to depict lighting patterns

Acknowledgement

The work reported in this paper is supported by **University of Texas at Arlington C-TEDD Funding: 019-13 Development of Automated Street Lighting Diagnosis Tools for Nighttime Traffic Safety Improvement**