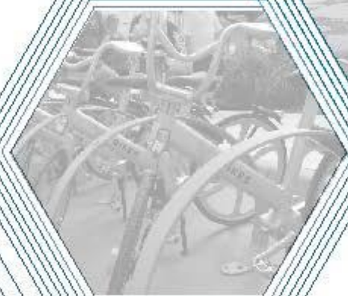

Development of Automated Roadway Lighting Diagnosis Tools for Nighttime Traffic Safety Improvement

Zhenyu Wang, Ph.D.
Pei-Sung Lin, Ph.D.
Srinivas Katkoori, Ph.D.
Mingchen Li
Abhijit Vasili
Runan Yang

FINAL REPORT

Center for Transportation Equity, Decisions & Dollars

DEVELOPMENT OF AUTOMATED ROADWAY LIGHTING DIAGNOSIS TOOLS FOR NIGHTTIME TRAFFIC SAFETY IMPROVEMENT

FINAL REPORT

By:

Zhenyu Wang, Ph.D.
Pei-Sung Lin, Ph.D.
Srinivas Katkoori, Ph.D.
Mingchen Li
Abhijit Vasili
Runan Yang
UNIVERSITY OF SOUTH FLORIDA

Sponsorship:

CTEDD

For:

Center for Transportation, Equity, Decisions and Dollars (CTEDD)
USDOT University Transportation Center
The University of Texas at Arlington
601 W. Nedderman Drive, Suite 103
Arlington, TX 76019-0108
817-272-5138 | C-Tedd@uta.edu

In cooperation with US Department of Transportation
Research and Innovative Technology Administration (RITA)

Acknowledgments

This work was supported by a grant from the Center for Transportation Equity, Decisions, and Dollars (CTEDD) funded by the U.S. Department of Transportation Research and Innovative Technology Administration (OST-R) and housed at The University of Texas at Arlington.

The authors would like to express their deepest appreciation to Edith Wong from the Florida Department of Transportation (FDOT) and Sara Beresheim from Johnson, Mirmiran & Thompson, Inc. As stakeholders, they provide great help in data collection and valuable comments to improve the tools. Special gratitude authors give to the FDOT Research Center for their support in local project matching.

Furthermore, the authors would also like to acknowledge Garrett Speed, Pete Reehling, and Ben Mittler from the University of South Florida (USF) Libraries for their technical support in building the ArcGIS web-GIS platform. Special thanks go to a USF Graduate Research Assistant, Ms. Qianwen Li, who made noble contributions in developing Crash Modification Factors using a case-control study. The authors must thank CUTR's faculties, staff, and students who have willingly helped in this project.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Center for Transportation, Equity, Decisions and Dollars (CTEDD), the U.S. Government and matching sponsor assume no liability for the contents or use thereof.

Technical Report Documentation Page			
1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**	
4. Title and Subtitle Development of Automated Roadway Lighting Diagnosis Tools for Nighttime Traffic Safety Improvement		**5. Report Date** August 31, 2020	
		6. Performing Organization Code	
7. Author(s) Zhenyu Wang, Pei-Sung Lin, Srinivas Katkoori, Mingchen Li, Abhijit Vasili, Runan Yang		**8. Performing Organization Report No.**	
9. Performing Organization Name and Address Center for Transportation, Equity, Decisions and Dollars (CTEDD) USDOT University Transportation Center The University of Texas at Arlington 601 W. Nedderman Drive, Suite 103 Arlington, TX 76019-0108		**10. Work Unit No. (TRAIS)**	
		11. Contract or Grant No.	
12. Sponsoring Organization Name and Address U.S. Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590		**13. Type of Report and Period Covered**	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract Roadway lighting is a conventional roadway infrastructure to ensure nighttime safety and security for multimodal road users (motorists, pedestrians, cyclists, and transit passengers). To cost-effectively maintain a roadway lighting system, key tasks in infrastructure management include periodically measuring roadway lighting levels, diagnosing lighting performance based on collected data, and providing decision-making support for maintenance and improvement. The Advanced Lighting Measurement System (ALMS) developed by CUTR provides a low-cost and time-effective solution to collect high resolution lighting data for a big-scale roadway network. This project aimed to develop innovative methods and tools to effectively analyze the "big" lighting data. A computer tool developed on the ArcGIS web-GIS platform in this study integrates three core modules and provides core functions for lighting pattern diagnosis, nighttime crash risk prediction, and data visualization. The hierarchical clustering algorithm is used to recognize lighting patterns based on the similarity of a photometric measure (average illuminance or uniformity). Safety performance function and Empirical Bayesian model were adopted to predict nighttime crash frequency with given lighting conditions. The study developed Crash Modification Factors to assess crash reductions due to lighting pattern improvement. The developed methods and tools were applied in CUTR's new lighting data collection and analysis tasks from the Florida Department of Transportation (FDOT) and Johnson, Mirmiran & Thompson (JMT), Inc. Two case studies demonstrated the performance of the computer tool in various application scenarios, such as lighting pattern recognition, lighting system upgrading validation, and nighttime crash risk analysis. To implement the developed tools in an operational environment, the research team will upgrade the tools, including developing Standard Operation Procedures, improving the lighting pattern diagnosis model, and providing planning-level lighting management for decision-makers.			
17. Key Words Street lighting, horizontal illuminance, nighttime safety, machine learning, web-GIS		**18. Distribution Statement**	
19. Security Classification (of this report) Unclassified.	**20. Security Classification (of this page)** Unclassified.	**21. No. of Pages** 59	**22. Price**

Table of Contents

List of Tables

List of Figures

CENTER FOR TRANSPORTATION, EQUITY, DECISIONS AND DOLLARS (CTEDD)
University of Texas at Arlington | 601 W Nedderman Dr #103, Arlington, TX 76019
C-tedd@uta.edu 817 272 5138

Abstract

Roadway lighting is a conventional roadway infrastructure to ensure nighttime safety and security for multimodal road users (motorists, pedestrians, cyclists, and transit passengers). To cost-effectively maintain a roadway lighting system, key tasks in infrastructure management include periodically measuring roadway lighting levels, diagnosing lighting performance based on collected data, and providing decision-making support for maintenance and improvement. The Advanced Lighting Measurement System (ALMS) developed by CUTR provides a low-cost and time-effective solution to collect high resolution lighting data for a big-scale roadway network. This project aimed to develop innovative methods and tools to effectively analyze "big" lighting data.

A computer tool was developed on the ArcGIS web-GIS platform in this study that integrates three core modules and provides core functions for lighting pattern diagnosis, nighttime crash risk prediction, and data visualization. The hierarchical clustering algorithm is used to recognize lighting patterns based the similarity of a photometric measure (average illuminance or uniformity). Safety performance function and Empirical Bayesian model were adopted to predict nighttime crash frequency with given lighting conditions. The study developed Crash Modification Factors to assess crash reduction due to lighting pattern improvement.

The developed methods and tools were applied in CUTR's new lighting data collection and analysis tasks from the Florida Department of Transportation (FDOT) and Johnson, Mirmiran & Thompson (JMT), Inc. Two case studies demonstrated the performance of the computer tool in various application scenarios, such as lighting pattern recognition, lighting system upgrading validation, and nighttime crash risk analysis. To implement the developed tools in an operational environment, the research team will continue to upgrade the tools, including developing Standard Operation Procedures (SOPs), improving lighting pattern diagnosis model, and providing planning-level lighting management for decision-makers.

Chapter 1: Introduction

1.1 Background

Nighttime crashes, particularly those that result in fatalities and injuries, are overrepresented on the US highway system. About 51% of fatal crashes and 30% of injury crashes occur at night, although only 21–23% of vehicle miles traveled (VMT) are at night (Monsere and Fischer, 2008). This issue is even more serious for vulnerable road users; nighttime crashes account for almost 70% of pedestrian fatalities (NHTSA, 2015). Reduced visibility in darkness, accompanied by drowsy and impaired driving, are the primary contributing factors to nighttime crash occurrence and injuries (National Safety Council, 2018). Roadway lighting, which "significantly improves the visibility of the roadway, increases sight distance, and makes roadside obstacles more noticeable to the driver, and therefore more avoidable" (FHWA, 2012), has been recognized as a vital countermeasure to prevent nighttime crashes. Additionally, roadway lighting provides clear benefits of personal security for pedestrians, bicyclists, and transit users during nighttime (FHWA, 2012). Thus, ensuring that a roadway lighting system provides adequate illumination is critical to improve nighttime safety and security for all road users.

The *AASHTO Roadway Lighting Design Guide* (AASHTO, 2011) provides roadway lighting design standards in terms of average horizontal/vertical illuminance, average luminance, and ratio-based uniformity metrics. The standards have been adopted by state departments of transportation (DOTs) for new street lighting design and existing system maintenance. Table 1 presents the lighting level criteria adopted in Florida.

Table 1. FDOT Lighting Level Criteria

Roadway Classification	**Average Illuminance Level (foot-candle)**		**Illumination Uniformity Ratios**		**Veiling Luminance Ratio**
	Horizontal	***Vertical***	***Avg./Min.***	***Max./Min.***	$L_{v(max)}/L_{avg}$
Conventional Roadway Lighting					
Freeway	1.5	N/A	≤ 4:1	≤ 10:1	≤ 0.3:1
Major Arterials	1.5				
Other	1.0				
High Mast Roadway Lighting					
All Roads	0.8 – 1.0	N/A	≤ 3:1	≤ 10:1	N/A
Signalized Intersection Lighting					
New Construction	3.0	2.3	≤ 4:1	≤ 10:1	N/A
Lighting Retrofit	1.5 (std.) 1.0 (min.)	1.5 (std.) 1.0 (min.)			
Midblock Crosswalk Lighting					
Low Ambient Luminance	N/A	2.3	N/A	N/A	N/A
Medium & High Ambient Luminance		3.0			

Source: FDOT Design Manual, Table 231.2.1

However, because of natural bulb degradation, obstacles including trees and animals, external lighting resources, and other unexpected factors, roadway lighting illumination performance may vary over time and cannot meet roadway lighting standards. Transportation agencies must obtain answers for two questions in their decision-making for lighting infrastructure management: Which segments/zones do not meet lighting maintenance standards, and which segments/zones have significant nighttime safety risks caused by poor lighting patterns and what is the risk degree? To answer the questions, periodic lighting level monitoring on a large-scale roadway network is needed.

1.2 Photometric Measures

Multiple photometric measures are adopted to measure street lighting levels in street lighting design and maintenance. The most common measure is illuminance that is the amount of light falling onto a surface. Illuminance could be measured as lumens per unit area either in foot-candles (lumens/ft^2) or in lux (lumens/m^2) (FHWA, 2012). As illuminance is independent of the retroflection characteristics of roadway surface, it is simple to measure and calculate. Two major illuminance metrics are used:

- *Horizontal illuminance* is measured at a horizontal surface six inches above the ground. This metric is related to the visibility of general objects, such as vehicles, obstacles, etc.
- *Vertical illuminance* is measured at a vertical surface height of five feet above the ground, which is a commonly criteria for the height of pedestrians' face. This metric represents the visibility of pedestrians.

Many factors influence driver vision when driving at night. From the photometric perspective, contrast and glare are two major factors. The contrast of an object is the luminance and color difference of the object from its background. High contrast of an object means that drivers can more easily detect the object. The function of street lighting is to produce sufficient luminance contrast of an object on a road that exceeds the requirement by drivers for detecting it. Glare is the eye adaption level affected by nonuniform lighting sources, especially bright ones. A long adaption procedure caused by nonuniformities in the visual field reduces driver vision for detecting objects and may result in collisions. A good design of a street lighting system can provide a uniform visual environment for drivers and reduce the risk of glare.

To assess the photometric performance of street lighting, two major metrics are used. Average illuminance is defined as the arithmetic mean of illuminance on a roadway facility. This metric represents the average lighting level of the roadway facility; a high average illuminance value indicates that the visual condition of the roadway facility is brighter, and the contrast of objects is sufficient. Meanwhile, max-min ratio (maximum illuminance ÷ minimum illuminance) or avg-min ratio (average illuminance ÷ minimum illuminance) are used to scale the uniformity of a street lighting system. Ratio metrics closing to 1 designates that the luminance distribution on a roadway facility is uniform and the risk of glare is small. As shown in Table 1, average illuminance (for both horizontal and vertical) and illuminance ratios have different requirements by various roadway functional classifications in street lighting design.

1.3 Advanced Lighting Measurement System (ALMS)

The conventional method for measuring roadway illumination involves spot-checking along a grid set up on a selected road. Measurements are taken using a light meter placed six inches above the ground. The operator must set up the light meter, then stand back and trigger a measurement. The light meter is then moved to the next point on the grid, and the process is repeated.

This method is both time-consuming and dangerous. It is a lengthy process for the operator to set up the collection device in the roadway, step back, and record a reading. This also makes the operator and the testing equipment vulnerable to the dangers of the road for prolonged periods of time. This danger can be avoided with lane closures and police presence but doing so adds cost to the measurement process; it is estimated that a one-mile measurement costs around $5,000. In addition, human error may reduce the accuracy and reliability of lighting data. Due to these factors, extended roadway illumination studies are rarely implemented, leaving questions on whether roadways meet sufficient lighting standards.

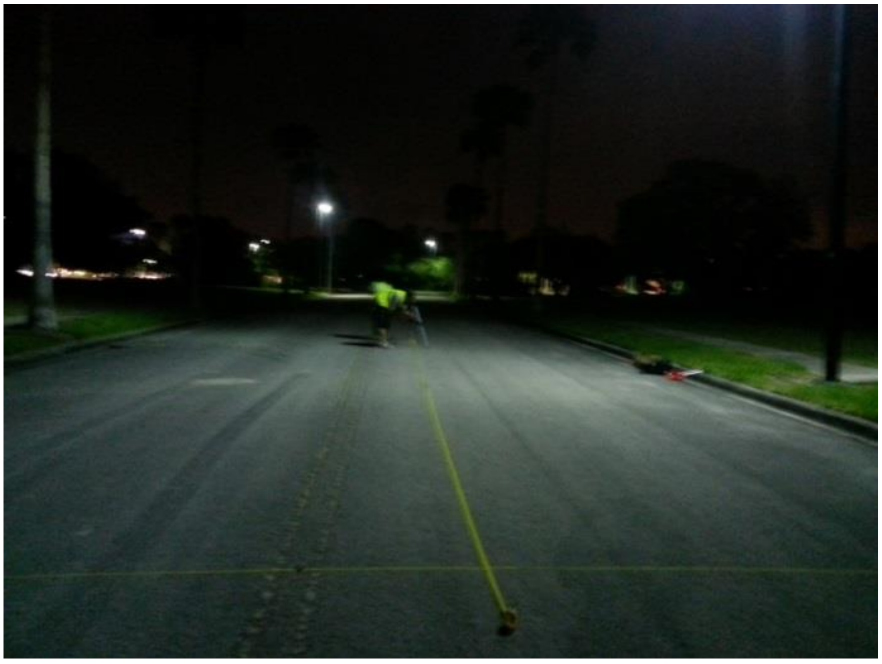

Figure 1. Conventional Lighting Data Measurement

Adequate and accurate lighting data from periodical lighting monitoring are the basis for supporting such decision-making. To overcome the defects of conventional lighting measurement methods, an innovative lighting measurement technology, the Advanced Lighting Measurement System (ALMS), was developed the University of South Florida (Johnson et al., 2014). The ALMS, as shown in Figure 2, is powered by a microcontroller that connects to two lighting meters (Konica Minolta T-10A). A Distance Measurement Instrument (DMI) reads accurate distance information from the vehicle's on-board diagnostics (OBD) bus and triggers the microcontroller reading lighting meters at a given distance interval. Compared to the conventional method, the ALMS has the following advantages:

- Can read two illuminance points per 10 ft per lane, and the high-resolution lighting data can describe the lighting pattern of a roadway facility more accurately. Figure 3 presents an example of horizontal illuminance data in foot-candles at a signalized intersection using the ALMS.
- Can be operated at a high speed (≥ 30 mph) in fully automatic method, operating cost is much lower than the conventional method ($300 vs $5,000). The high operating speed and low cost allow periodic lighting data collection in a big-scale roadway network.
- Fully eliminates operator exposure in traffic and human error in operations; provides highly accurate and reliable lighting data.

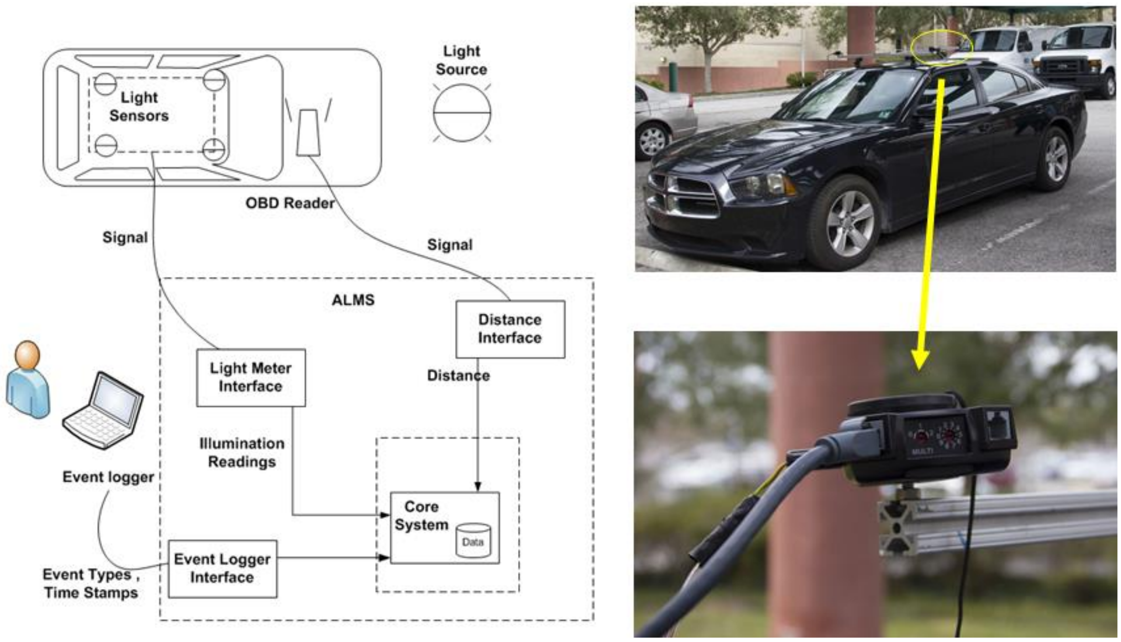

Figure 2. Advanced Lighting Measurement System (ALMS)

Figure 3. Examples of Horizontal Illuminance Data Collected by ALMS

Researchers have used the ALMS to collect horizontal lighting data for 300+ centerline miles in the Tampa Bay areas since 2012, with accumulated lighting points exceeding millions and covering the major corridors in the area, as shown in Figure 4. The big data produced from the ALMS bring opportunities to deeply analyze lighting performance and improve nighttime safety.

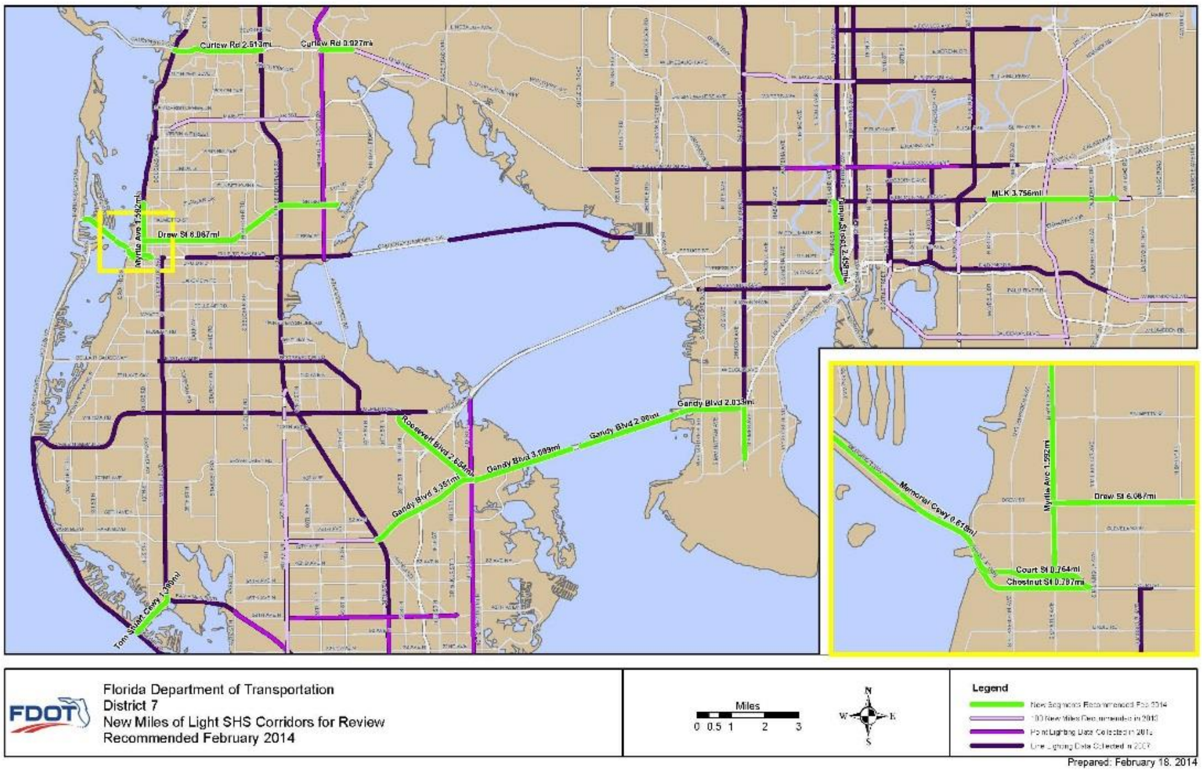

Figure 4. Lighting Data Measurement in Tampa Bay using ALMS

1.4 Safety Performance of Street Lighting

The safety performance of street lighting has been documented in many reports and publications. Most studies considered street lighting as a binary factor (presence or not; before and after improvement) at intersections (Bruneau and Morin, 2005; Bullough et al., 2013; Donnell et al., 2010; Elvik, 1995; Isebrands et al., 2006; Kim and Washington, 2006; Preston and Schoenecker, 1999), roadway segments (Anarkooli and Hadji Hosseinlou, 2016; Wanvik, 2009; Yu et al., 2015; Zhang et al., 2012), or both (Monsere and Fischer, 2008; Sullivan and Flannagan, 2002). These studies all led to the conclusion that the presence of roadway lighting significantly improves nighttime safety in terms of reduction of nighttime crash rate/frequency, injury severity, and night-day crash ratio.

Some efforts, as summarized in Table 2, investigated the relationship between nighttime crash risk and lighting photometric measures (e.g., illuminance, luminance, uniformity) or visibility indicators (e.g., Small Target Visibility) rather than street lighting presence or improvement on roadway segments and at intersections.

Table 2.Summary of Previous Studies on Safety Performance of Street Lighting Photometric Measurers

Study	Roadway Facility	Photometric Measures	Safety Measure	Major Conclusions
Box (1971)	Freeway	Horizontal illuminance	Night-to-day crash rate ratio	Freeways with horizontal illumination levels of 0.3–0.6 fc had best night-to-day accident rate ratios.
Box (1976)	Arterial	Horizontal illuminance	Crash frequency	Nighttime crash frequency increased by 10% with decrease in horizontal illuminance from 14 lx to 9 lx.
Janoff et al. (1978)	Urban roadway segments	Horizontal luminance	Crash rate	• Crash rate decreases with increase in visibility level. • Higher illumination levels related to higher crash frequency.
Scott (1980)	Roadway segments	Horizontal luminance, surrounding luminance, overall uniformity, horizontal illuminance, vertical Illuminance	Night-to-day crash ratio	• Clear relationships between nighttime crash risk and measures of lighting quality. • Average horizontal luminance ($\bar{L}$) superior to surround luminance and horizontal/vertical illuminance to fit crash data. • In range 0.5–2.0 cd/m^2, increase of 1 cd/m^2 in luminance associated with 35% lower night-to-day crash ratio ($\frac{N}{D}$). Best-fitting relationship is $\frac{N}{D} = 0.66e^{-0.42\bar{L}}$. • Overall uniformity (max/min) and homogeneity of luminance have no proved relationship with nighttime crash risk.
Mace (1997)	Urban and suburban freeways, arterials, divided roadways	Horizontal illuminance, luminance, small target visibility (STV)	Night-to-day crash and crash cost ratios	• Less uniformity resulted in higher night/day crash rates. • Influence of STV and light level on night/day crash ratios confounded with glare. • Data did not support conclusion that increases in illumination are more likely to reduce crashes than increases in visibility.
Keck (2001)	Urban and suburban freeways, arterials, divided roadways	Horizontal illuminance, luminance, small target visibility (STV)	Night-to-day crash ratio	• Evaluation of visibility of real roadway objects needs include joint effects of both vehicle headlights and fixed street lighting system. • No evidence to support correlation between night-to-day crash ratio and street lighting measures.

Zhao et al. (2015)	Urban arterial	Horizontal illuminance, Uniformity	Night-to-day crash rate difference	• Horizontal illuminance along corridor complies with lognormal distribution. • Mean and standard variance of horizontal illuminance proposed to replace traditional measures for average lighting level and uniformity. • Illuminance parameters significantly related to difference between daytime and nighttime crash rates.
Wang et al. (2017)	Roadway segment	Horizontal illuminance, Uniformity	Expected night-to-day ratio	• Increase in mean horizontal illuminance significantly decreases expected night-to-day crash ratio. • Crash reduction factor for mean of illuminance (*x*) estimated as $(x/\text{baseline})^{-0.0773} \times 100\%$. • Good illuminance uniformity (max/min < 6) significantly reduces expected night-to-day crash ratio by 2.3%.
Bhagavath-ula et al. (2015)	Rural intersection	Horizontal illuminance	Night-to-day crash ratio	• Lighting level significantly impacts night-to-day crash ratio at rural intersections. • Increase of 1 lx in average horizontal illuminance corresponded to 7% decrease in night-to-day crash ratio. • For lighted intersections, night-to-day ratio decrease is 9%; for unlighted intersections, night-to-day ratio decrease is 21%.
Edwards (2015)	Rural intersection	Horizontal illuminance	Night-to-day crash ratio	• 1-lux (≈ 0.1 fc) increase in average lighting (3.91 lux) reduced nighttime crash rate by 9% in Minnesota. • At lighted intersections, a one-lux increase in average lighting (6.41 lux) reduced crashes by 20%. • At unlighted intersections, a one-lux increase in average lighting (0.2 lux) reduced nighttime crash ratios by 94%.
Wei et al.(2016)	Urban intersection	Horizontal illuminance	Expected night-to-day crash ratio	• Increasing intersection illuminance from low (< 0.2 fc) to medium (≥ 0.2 fc and <1.1 fc) can reduce nighttime crash frequency and night-to-day crash ratio by approximately 50%. • Illuminance kept at 0.9 fc or higher, risk of fatality and severe injury significantly decreases, especially for pedestrian/bicycle, head-on, and angle crashes.
Bhagavath-ula et al. (2018)	Rural intersection	Horizontal illuminance		• Visual performance of driver plateaued between 7 and 10 lx of mean intersection illuminance.

Xu et al. (2018)	Access points	Horizontal illuminance	Speed variance, Crash rate	• Improved illuminance can decrease speed variation among vehicles and improve safety levels. • High-grade highways need better illuminance at access points.
Yang et al. (2019)	Roadway segment	Mean and standard deviation of horizontal illuminance	Nighttime crash frequency and daytime crash frequency	• Horizontal illuminance characteristics have a significant impact on nighttime crash risk on roadway segments. • An increase in the mean of horizontal illuminance, indicating an improvement in average lighting level, tends to decrease nighttime crash risk; an increase in the standard deviation, representing a poor uniformity of lighting pattern on a roadway segment, is more likely to raise nighttime crash risk. • Because the two measures are strongly correlated in a low mean range (<0.44 fc), the two photometric measures need to be considered together to interpret the safety effects of lighting patterns. • The standard deviation shows better performance in measuring lighting uniformity on a roadway segment than the traditional ratios (max-to-min and mean-to-min).

1.5 Research Objectives

The "big" lighting data produced from ALMS bring opportunities to regularly monitor roadway lighting systems and deeply analyze lighting performance and challenges in data processing and analysis. First, existing lighting analysis tools focus on roadway lighting design rather than analysis of measured lighting data; there are no standard methods and computer tools to process measured big lighting data for decision-making support. Second, many studies have explored the connection between nighttime crash risk and lighting photometric measures; however, these photometric studies used simple lighting measures statistics (e.g., mean of illuminance/ luminance and ratio-based uniformity) that cannot use the full information of lighting data and capture the "true" lighting patterns that influence driver nighttime vision, consequently contributing to nighttime crash risks (Wang et al., 2019). Traditional models are not appropriate for analyzing big lighting data on disaggregate zones. Upgrading or maintaining lighting systems on disaggregate zones with high nighttime risks can better identify nighttime risks related to lighting patterns and more effectively allocate maintenance budgets. Thus, it is necessary to develop innovative methods to diagnose lighting system performance based on high-resolution lighting data and provide decision-making support in roadway lighting management and maintenance.

The major goal of the proposed project was to develop innovative methods and tools that automatically and intelligently diagnose roadway lighting performance based on big lighting data collected using the ALMS. The diagnosis results can provide decision-making support in roadway lighting system maintenance and management to improve nighttime safety and security. More specifically, the research objectives were as follows:

- Develop diagnosis algorithms that can effectively recognize lighting patterns and identify zones with poor lighting performance (e.g., do not meet standards).
- Develop machine learning models to automatically extract features from lighting patterns and precisely predict nighttime crash risk associated with given lighting patterns and other factors.
- Develop a prototype of computer tools for automatically processing and analyzing collected lighting data, including data mapping and pre-processing, intelligent diagnosis, nighttime risk assessment, and data visualization. The expected technology readiness level (TRL) is Level 7: Prototype demonstrated in operational environment.
- Implement the developed tools using Florida Department of Transportation (FDOT) lighting measurement projects as case studies.

Chapter 2: Tool Development

2.1 System Architecture

The system architecture of the Automated Roadway Lighting Diagnosis System is shown in Figure **5**. The system consists of four modules:

- *Data Mapping and Inventorying* – Imports raw lighting data from the ALMS data logger and map the data into ArcGIS layers. Besides the GIS coordinates, the roadway ID and milepost are produced for each lighting data item using the Linear Reference function of ArcGIS Pro. The mapped lighting data are stored in a geographic database.
- *Lighting Pattern Recognition* – Analyzes the lighting data for a roadway facility (i.e., segment or intersection) and identifies the sub-sections for which their lighting patterns do not meet FDOT standards. The lighting pattern diagnosis uses a machine learning algorithm developed in this study; details of the algorithm are provided in Chapter 3.
- *Risk Prediction* – Predicts the nighttime risk for a roadway facility with lighting data, traffic data, geometric data, and historical crash data. The Safety Performance Functions (SPF) and Empirical Bayesian (EB) model are used to predict the crash risk; their descriptions are provided in Chapter 3.
- *Data Visualization* – Presents the results of pattern diagnosis and crash risk prediction. Presentation methods include GIS maps, heat-maps, statistics, and figures.

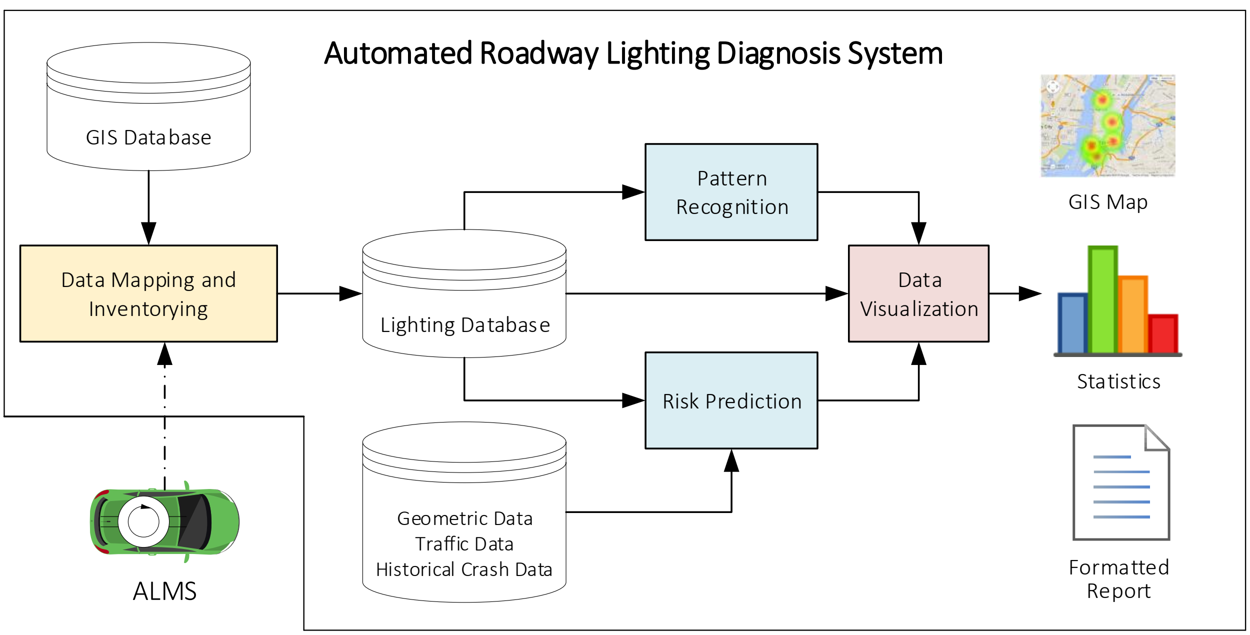

Figure 5. **System Architecture of Automated Roadway Lighting Diagnosis System**

2.2 System Development

The lighting diagnosis tools were designed to work on a web-GIS platform, as users can easily access the tools with a modern web browser (i.e., Google Chrome, Firefox, MS Edge) using a variety of devices (i.e., PC, Mac, tablets, mobile phones). The research team developed the tools based on the ESRI solution with the University of South Florida (USF) academic license. The tools and technologies for system development are summarized in Table 3.

Table 3. Summary of Environment & Technologies for Development

System Component	Tool/Technology
Deployment – Server Side	
Web Server	MS IIS 8.5 on Windows Server 2012 R2
GIS Server	ArcGIS Server 10.8
GIS Service	Customized geoprocessing service
Execute Mode	Asynchronous
Deployment – Client Side	
Device	Any device with an Internet connection
Software	Modern browsers
Development – Server Side	
Development/Testing Platform	ArcGIS Pro
Language	Python
Library	ArcPy, NumPy, Pandas
Development – Client Side	
Development/Testing Platform	MS Visual Studio Code
Language	JavaScript
Library	ArcGIS API for JavaScript 4.15, UIkit, Chart.js

2.3 System Functions

The system provides various functions to satisfy user needs in lighting data analysis. Users can access the functions through a web-GIS interface hosted on a CUTR server at http://its.cutr.usf.edu/lita. The major functions are summarized in Table 4.

Table 4. Summary of Major Functions

Function	Subject	Description
User Interface		
Display layer	• Signal • Segment • Crash	• Display subjects with lighting data on maps • Filter subjects by location, time, or type
Base-map		• Provide 24 base-maps • Flexible to switch base-maps
Selection	• Signal • Segment • Crash	• Specify subjects for analysis • Interactive selection on map
Clip	• Segment	• Specify a sub-section of a segment for analysis • Interactive selection on map
Split	• Segment	• Manually split a segment into multiple sub-sections for analysis • Interactive selection on map
Segmentation	• Segment	• Automatically split a segment into multiple sub-sections for analysis • By uniformed length • By the similarity of lighting patterns
Analysis		

Lighting diagnosis	• Signal
	• Segment
	• Crash
		• Calculate photometric statistics for selected subjects
		  – Mean, Max/Min, Avg/Min
		  – Customized buffer size
		• Display histogram charts for photometric values

Risk prediction	• Signal • Segment	• Predict nighttime crash risk (frequency) for selected subjects • Using EB model and SPFs • Input variables: lighting data, AADT, geometry, traffic control, and historical crash frequency • Display prediction results in table and figure
Heat-map	• Signal • Segment • Crash	• Display heat-maps of photometric data for selected subjects on maps • Intuitive presentation of lighting patterns

2.4 Data Inventory

The system maps ALMS data into GIS layers, which can be joined with other data (Annual Average Daily Traffic [AADT], geometry, crash, etc.) for analysis. All data are stored on the ArcGIS server in GIS formats. The map and geoprocessing services retrieve the data via the ArcPy library. Table 5 presents the data inventory used in the system.

Table 5. Data Inventory in Automated Roadway Lighting Diagnosis System

Data Item	Format	Description
Lighting Data Layer		
FC	Float	• Measured horizontal illuminance (foot-candle) at vehicle height
FC_6	Float	• Converted horizontal illuminance at 6-in. above ground
Coordinates	Float	• GIS coordinates (X, Y)
Roadway ID	Integer	• Unique number to identify a roadway segment in FDOT Roadway Characteristics Inventory (RCI) database
Milepost	Float	• Number to measure distance to a reference point
Date/Time	Integer	• Data collection year, date, time
Segment Layer		
Number of Lanes	Integer	• Number of through lanes
Area Type	Indicator	• Urban or not
Roadway Type	Indicator	• Divided or undivided
Access Points	Integer	• Number of intersections and driveways
AADT	Integer	• Average Annual Daily Traffic by year
T-Factor	Float	• Percentage of truck volume by year
Intersection Layer		
AADT1	Integer	• AADT on Street 1
AADT2	Integer	• AADT on Street 2
Intersection Type	Indicator	• Four-leg or three-leg
Crash Layer		
Year	Integer	• Crash year (2017–2019)
Severity	Indicator	• Fatal, incapacitating injury, non-incapacitating injury, possible injury, PDO
Crash Type	Indicator	• Rear-end, angle, pedestrian, …
Location	Integer/Float	• Roadway ID, milepost, coordinates

Chapter 3: Model Development

3.1 Lighting Diagnosis Model

3.1.1 Problem Statement

In street lighting management, engineers usually diagnose a street lighting system by reviewing photometric statistics on a roadway facility. If the photometric statistics cannot satisfy the DOT standards, shown in Table 5, the lighting system may need to be maintained or upgraded. The system provides the functions to calculate photometric statistics, such as mean of horizontal illuminance, standard deviation of horizontal illuminance, max-min ratio, avg-min ratio, and histogram, for a selected roadway facility. The functions work well for a small-size entity, such as intersections or short roadway segments; however, they cannot account for the diversity of lighting patterns on a long road segment.

Figure 6 is an example of illuminance distribution along a roadway segment, showing that the lighting level of the left section is higher than that of the right section. The photometric statistics of the whole segment cannot accurately describe the lighting pattern of segment. The mean of illuminance (0.85 fc), max-min ratio (14.19), and avg-min ratio (8.54) are all higher than the DOT lighting criteria (1.0 fc, 10, and 4, respectively). The statistics derive a conclusion that the lighting level of the whole segment does not meet DOT standards. However, the lighting statistics on the left section are 1.05 fc (mean), 3.48 (avg-min ratio), and 4.65 (max-min ratio), indicating that the left section does meet DOT standards. Comparatively, the right section (mean of 0.75 fc, max-min ratio of 14.19, and avg-min ratio of 7.52) does not meet the standards. In practice, engineers would expect to identify the right section for lighting maintenance or upgrade and exclude the left section.

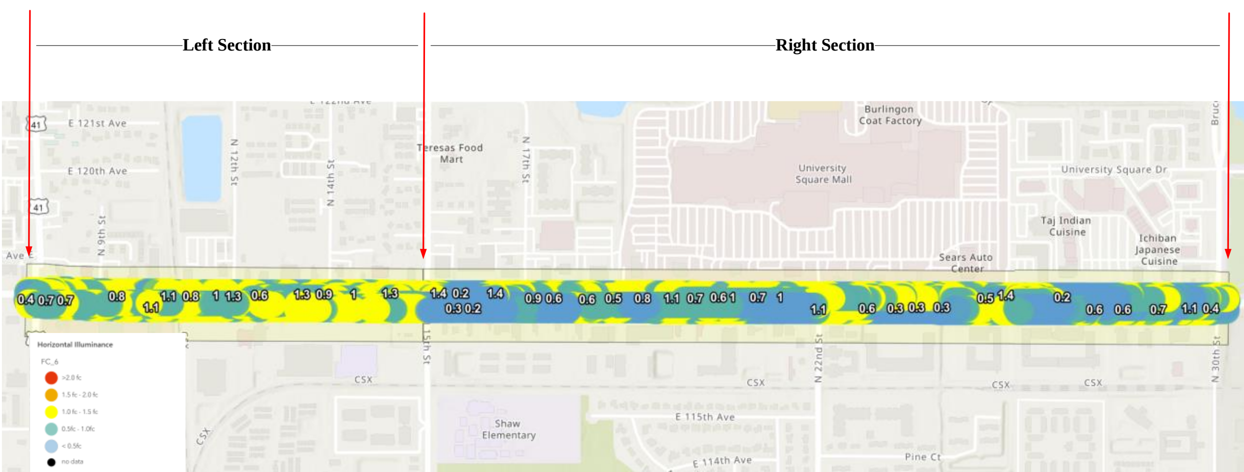

Figure 6. Example of Diverse Lighting Patterns along a Roadway Segment

The system provides functions of clip and segmentation. With these functions, users can manually split the whole segment into sub-sections (such as left and right). An intelligent diagnosis function that automatically identifies sub-sections that do not meet DOT standards is expected to improve lighting diagnosis efficiency and accuracy.

3.1.2 Automatic Segmentation Algorithm

A machine learning model was developed to automatically split the whole roadway segment into sub-sections. Each sub-section has similar photometric performance (either mean of illuminance or ratio-based uniformity measure), but the photometric measures of neighboring sections are different. The automatic segmentation model, as shown in Figure 7, adopts the hierarchical clustering algorithm that iteratively combines two neighboring sections based on their similarity of photometric measures. The details of important terms are interpreted as follows:

- *Parameters* – To provide flexibility for different situations, the algorithm has four parameters. Users can specify parameters based on their needs and get an optimal diagnosis result. The parameters are summarized in Table 6.

Table 6. Parameters for Automatic Segmentation Algorithm

Parameter	Description	Default Value
Photometric Measure	User specifies photometric measure to calculate similarity (distance) between two neighboring sections. Measure could be mean, standard deviation, max-min ratio, or avg-min ratio.	n/a
Initial Length	Whole segment is split into small sections based on initial section length; shorter initial length may recognize small sections with different lighting patterns from neighboring sections but may increase calculation time.	0.003 miles
Label Range	Label used to identify similarity in photometric measures of neighboring sections. Two neighboring sections falling into a category (label) are treated as similar sites and can be merged into one section. More labels levels can increase the resolution of lighting pattern diagnosis; however, more label levels may result in fragmentation.	[0, 0.5, 1, 1.5, 2, >2] for mean of horizontal illuminance
Minimum Length	Minimum length used as stopping criteria. Length of final sections required to be greater or equal to minimum length. Setting used to avoid issue of fragmentation.	0.135 miles

- *Similarity* – The absolute difference of photometric measures between two neighboring sections is an indicator of the similarity of the sections. The “closest” section pair is defined as two neighboring sections with the smallest difference of photometric measures.
- *Block* – If two neighboring sections have different label levels and both satisfy the minimum length, the two sections are indicated as an unsimilar pair (blocked). The algorithm ignores “blocked” section pairs in iterations.

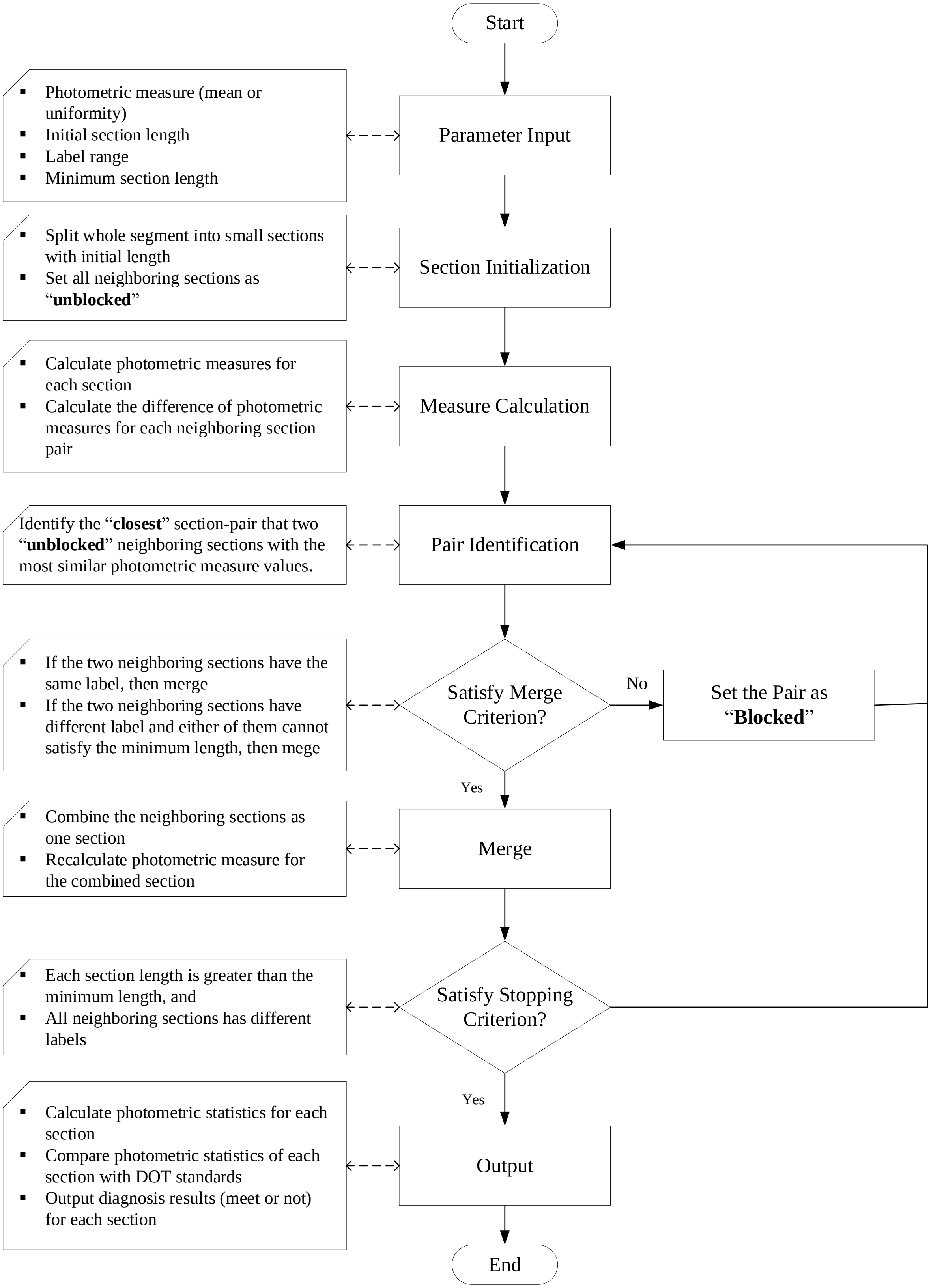

Figure 7. Flow Chart of Automatic Segmentation Algorithm

3.1.3 Example of Automatic Lighting Diagnosis Algorithm

An example is used to explain the automatic lighting diagnosis algorithm. Assuming that lighting data are collected at a roadway segment of 0.5 miles, the parameter inputs include the following:

- Mean of illuminance is selected as photometric measure for analysis.
- Initial section length is 0.1 miles.
- Label range is 1 – [0 fc, 0.5 fc), 2 – [0.5 fc, 1 fc), 3 – [1 fc, 1.5fc), 4 – [1.5 fc, ∞).
- Minimum length is 0.2 miles.

Based on the parameter inputs, the automatic lighting diagnosis procedure is as follows:

- *Iteration 1* – The whole segment is splatted into five sections based on the initial length (0.1 miles), and the mean of illuminance for each section is calculated. Based on the calculated mean, a label is assigned to each section. For example, Section 1 is labeled as 3 because its mean of illuminance is 1.17 falling in [1 fc, 1.5 fc). Meanwhile, the distance (the difference of illuminance mean) between two neighboring sections is calculated. Section 2 and Section 3 are identified the "closest" pair, as they have the smallest distance (0.1 fc). The pair satisfies the merging criteria: two neighboring sections have the same label (2). Thus, the two sections are merged into a single section.

Section	Length (miles)	Label	Mean of Illuminance (fc)	Similarity*
1	0.1	3	1.17	-
2	0.1	2	0.86	0.31
3	0.1	2	0.76	0.10
4	0.1	1	0.32	0.44
5	0.1	2	0.54	0.22

* Mean of Illuminance for Section$_i$ - Mean of Illuminance for Section$_{i-1}$

- *Iteration 2* – The mean of illuminance is recalculated for the new section (Section 2) that combines Section 2 and Section 3 in Iteration 1. Section 2 is labeled as 2 because its mean of illuminance is 0.81 fc. Meanwhile, the distance between any two neighboring sections is recalculated. Section 3 and Section 4 are identified the "closest" pair, as they have the smallest distance (0.22 fc). The "closest" pair satisfies the merging criteria: if two neighboring sections have the different label (1 vs 2), but either of them is shorter than the minimum length (0.2 miles). Thus, the two sections are merged into a single section.

Section	Length (miles)	Label	Mean of Illuminance (fc)	Similarity
1	0.1	3	1.17	-
2	0.2	2	0.81	0.36
3	0.1	1	0.32	0.49
4	0.1	2	0.54	0.22

- *Iteration 3* – The mean of illuminance is recalculated for the new section (Section 3) that combines Section 3 and Section 4 in Iteration 2. Section 3 is labeled as 1 because its mean of illuminance is 0.43 fc. Meanwhile, the distance between any two neighboring sections is recalculated. Section 1 and Section 2 are identified the "closest" pair since they have the smallest distance (0.36 fc). The "closest" pair satisfies the merging criteria: if two neighboring sections have the different label (3 vs 2), but either of them is shorter than the minimum length (0.2 miles). Thus, the two sections are merged into a single section.

Section	Length (miles)	Label	Mean of Illuminance (fc)	Similarity
1	0.1	3	1.17	-
2	0.2	2	0.81	0.36
3	0.2	1	0.43	0.38

- *Iteration 4* – The mean of illuminance is re-calculated for the new section (Section 1) that combines Section 1 and Section 2 in Iteration 3. Section 1 is labeled as 2 because its mean of illuminance is 0.93 fc. Meanwhile, the distance between two neighboring sections is recalculated. All neighboring sections have different labels, and each section reaches the minimum length. Thus, the iteration stops.

Section	Length (miles)	Label	Mean of Illuminance (fc)	Similarity
1	0.3	2	0.93	-
2	0.2	1	0.43	0.56

The mean of illuminance for Section 1 is 0.93 fc, which is close to the DOT standard (1.0 fc). However, the measure for Section 2 is 0.43 is much lower than the DOT standard (0.43 fc). Thus, Section 2 should be put in the priority list for upgrading and maintaining.

3.2 Crash Risk Prediction Model for Roadway Corridor

3.2.1 Safety Performance Function

A Safety Performance Function (SPF) developed in a previous study (Yang et al., 2019a) was used to evaluate the nighttime crash risk. The SPF adopts the Random Parameter Negative Binominal technology to predict the nighttime crash frequency, given lighting photometric measures (mean and standard deviation), traffic conditions (AADT and truck percentage), roadway type (divided or not), area type (urban or rural), access density, and segment length. The SPF equation is given below:

$$\begin{aligned}\mu = \ \exp\,(&-4.969 - 0.42 \cdot M + 0.769 \cdot S + 0.526 \cdot LN(AADT) \\ &+ 0.236 \cdot HV + 1.161 \cdot L \ + 0.036 \cdot \ ACCESS \\ &+ 0.456 \cdot UNDIVIDED + 0.283 \cdot URBAN)\end{aligned} \tag{1}$$

where μ – predicted nighttime crash frequency (per four years)

M – mean of horizontal illuminance (fc)

S – standard deviation of horizontal illuminance

$LN(AADT)$ – natural logarithm of AADT

HV – heavy vehicle indicator (1 – if heavy vehicle percentage > 3%, 0 – otherwise)

L – segment length in miles

$ACCESS$ – access density (number of access points per mile)

$UNDIVIDED$ – undivided road indicator (1 – if roadway is undivided, 0 – otherwise)

$URBAN$ – urban road indicator (1 – segment is located inside urban limits but not inside city limits, 0 – otherwise)

An Empirical Bayesian (EB) model is applied to provide more reliable inference for nighttime crash risk at a given corridor. The EB model, which is a standard method define in the *Highway Safety Manual* (HSM), combines the crash risk information from the SPF and from local historical crash data. The EB model is described below:

$$N_e = w \times N_p + (1 - w) \times N_o$$
$$w = \frac{1}{1 + N_p' \times Y \div \phi} \tag{2}$$

where N_e – expected nighttime crash frequency (per year)

N_p – predicted nighttime crash frequency (per year) by SPF, $= \frac{\mu}{4}$

N_o – average historical crash frequency (per year)

w – weighting factor

N_p' – predicted crash frequency (per year) for unit length (one mile), $= \frac{N_p}{L}$

Y – number of years for historical crash data collection

ϕ – overdispersion parameter estimated in SPF development, = 3.604

The expected nighttime crash frequency (N_e) represents the expected nighttime risk of a road segment under prevailing lighting, traffic, and roadway conditions. In addition, the probability (risk) of different nighttime crash number is calculated by:

$$P(N) = \frac{\Gamma(\phi + N)}{\Gamma(\phi) \cdot N} \left(\frac{\phi}{\phi + N_e}\right)^{\phi} \left(\frac{N_e}{\phi + N_e}\right)^{N} \tag{3}$$

where $P(N)$ – probability of N nighttime crashes occurring

$\Gamma()$ – Gamma function

3.2.2 Implementation of Crash Risk Prediction Model

The crash risk prediction model (SPF and EB model) was coded as an ArcGIS Geoprocessing Tool (risk prediction module) in the system. The module retrieves crashes, lighting, traffic, and geometry data from the system databases and display the prediction results on web. The flow chart of crash risk prediction model is shown in Figure 8.

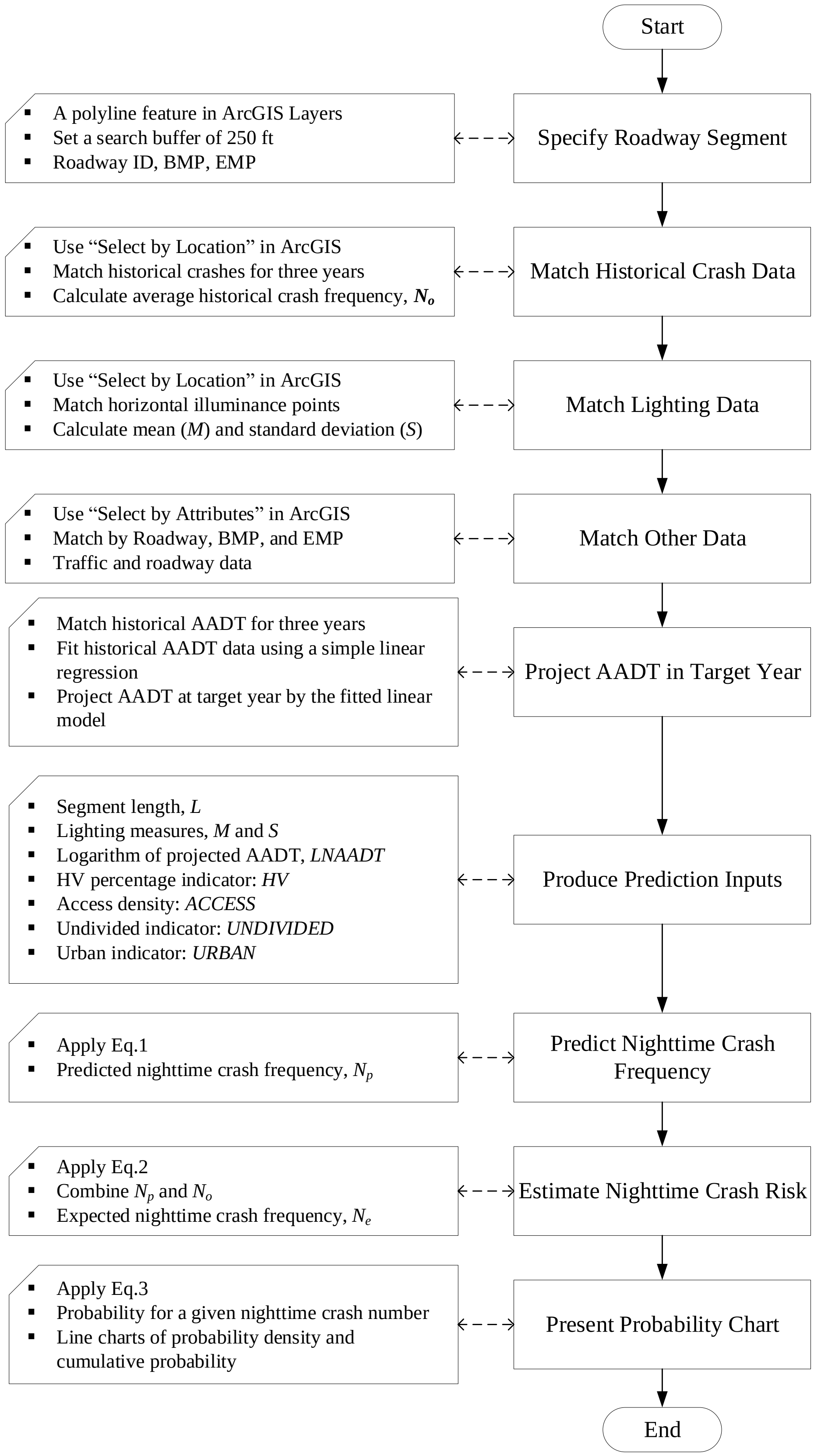

Figure 8. Flow Chart for Crash Risk Prediction

3.3 Crash Modification Factors for Street Lighting Photometric Measures

3.3.1 Problem Statement

Crash modification factors (CMFs) are a measure used to evaluate the safety performance of a treatment in crash reduction. Usually, it is necessary to isolate the effect of a factor of interests from other contributing factors. Lighting safety studies were based on cross-sectional observation data, as it was difficult to collect lighting photometric data in street lighting improvement projects for before-after studies. Street lighting photometrics, including street lighting uniformity, are associated with many confounding factors; for example, transportation agencies usually require high lighting levels on major roads that serve high traffic volume. Thus, AADT is a typical confounder connecting both street lighting level and uniformity and nighttime crash frequency. Observational cross-sectional studies without any control on confounding factors may lead to counterintuitive conclusions such as high lighting level associated with high nighttime crash frequency (Janoff et al., 1978; Wei et al., 2016) or no association between street lighting level and nighttime crashes (Keck, 2001). Night-to-day crash ratios were widely used in previous studies (Box, 1971; Keck, 2001; Scott, 1980; Wang et al., 2017; Zhao et al., 2015) to exclude the unnecessary influence of factors that impact both nighttime and daytime crashes (e.g., roadway maintenance quality, functional class, etc.) at the same sites. However, night-to-day ratios cannot address confounders that connect nighttime crashes only. For instance, the standard deviation of illuminance (a measurement of uniformity) tends to increase with a high mean of illuminance in lighting data samples. Both variables significantly influence driver detection ability at night and, consequently, contribute to nighttime crashes but not daytime crashes. Night-to-day ratios cannot control the influence from the illuminance mean when examining the safety effect of the standard deviation of illuminance, or vice versa.

The primary objective of this study was to investigate the safety effects of street lighting uniformity on nighttime crash occurrence on roadway segments. A matched case-control method was used to address the critical issues in cross-sectional street lighting data. CMFs of uniformity were developed for roadway segments.

3.3.2 Matched Case-Control Study

A matched case-control study is commonly used in epidemiology to quantify the risk of disease given certain characteristics related to an individual (Schlesselman, 1982). It has been used recently in the highway safety field to investigate the risk of vehicle crashes given certain characteristics related to a roadway entity (Abdel-Aty et al., 2004; Davis et al., 2006; Gross, 2013; Gross and Donnell, 2011; Gross and Jovanis, 2007). Unlike a cross-sectional study that proceeds from cause to effect—although a matched case-control study is based on cross-sectional data—this method adopts an opposite procedure (from effect to cause), attempting to identify pre-condition factors contributing to the outcomes (crashes). The basic steps of a matched case-control study are illustrated as follows:

- *Step 1 – Defining*: Roadway entities are split into two groups: 1) case – an entity (roadway segment) that experienced at least one crash in a given period, and 2) control – an entity (roadway segment) that did not experience a crash in the same period.

- *Step 2 – Matching*: Multiple control entities are randomly matched to each case based on the similarity of confounding factors correlated to both the risk factor of interest (e.g., lighting photometrics) and the outcome (e.g., nighttime crash). This matching scheme is intended to eliminate the biased estimations on the association between the risk factor of interest and the outcome through mitigating the disturbance from confounders (Schlesselman, 1982). Since it often is impractical to match the exact value for confounding variables, a category-matching scheme is usually implemented to stratify each variable (stratification) and randomly pair cases and controls that fall into the same cell created by the multiple cross-classification. The case-control ratio is constant by the minimum ratio of controls to cases among all cross-classification categories.
- *Step 3 – Modeling*: The conditional logistic regression model is estimated based on the matched case-control pairs to address the relative risk of unmatched risk factors (lighting photometrics and other geometric factors) rather than the probability of a crash in terms of expected frequency (Gross et al., 2010).

The matched case-control method has some distinctive features that lead to being more valid to infer causality than the cross-sectional study. First, this method adopts a comparison group (controls) to support or refute an inference of a cause for any risk factor (Schlesselman, 1982). Second, a case-control study is powerful for studying rare events since a pre-specified number of cases (roadway entities experiencing nighttime crashes in a given period) enrolled in the study can ensure an adequate sample size for analysis (Woodward, 2013). This feature is valuable for addressing the issues of excess zero observations in the previous cross-sectional street lighting safety studies. Third, a matched design in a matched case-control study can directly control for confounding variables (including temporal instability) because each matched stratum has similar values for each confounding variable (Schlesselman, 1982). Fourth, matched sampling in a matched case-control study leads to a balanced number of cases and controls, which can reduce the variance in the parameters of interest (horizontal curve design features) and improve statistical efficiency in model estimation (Sahai and Khurshid, 1995).

A careful design is needed to avoid overmatching and residual confounding that may cause biased inference and inefficiency of study. Overmatching may be caused by matching variables associated only with or having equal status with either the risk factor of interest or the outcome; consequently, the relationship between exposure and outcome will be obscured (Marsh et al., 2002). In addition, the strong correlation between matching variables may also result in overmatching (Schlesselman, 1982). If additional confounders are not controlled or stratification of matching variables is too loose, residual confounding that the effects of confounding factors could not be eliminated from the risk-outcome effect of interest may result in biased inference (Psaty et al., 1999). However, if too many confounders are controlled by matching or stratification of matching variables is too tight, the case-control ratio in each matching cell would decrease dramatically and reduce the power of analysis (Woodward, 2013), even with insufficient controls for matching. In practice, a trade-off analysis of the number of and the stratification width for matching variables should be made based on statistical analysis and professional knowledge to avoid residual confounding and/or insufficient controls.

Let i $(i = 1, 2, \cdots, I)$ be an index to represent the matched case-control stratum. In each stratum, J controls are randomly matched to one case based on the similar values of confounding variables. Let j $(j = 0, 1, 2, \cdots, J)$ be an index to represent the observation record within each stratum. Let k $(k = 0, 1, 2, \cdots, K)$ be an index to represent the unmatched explanatory variablex_k. The probability of a binary outcome associated with the unmatched explanatory variables for j^{th} observation of the i^{th} stratum can be given as

$$\Pr(y_{ij} = 1) = 1 \Bigg/ \left\{1 + exp\left[-\left(\alpha_i + \sum_{k=1}^{K} \beta_k x_{ijk}\right)\right]\right\} \tag{4}$$

where $\Pr(y_{ij} = 1)$ is the probability that the j^{th} observation in the i^{th} stratum is a case; x_{ij} is a row vector for k unmatched explanatory variables $x_{ij} = (x_{ij1}, x_{ij2}, x_{ij3}, \cdots, x_{ijk})$; x_{ijk} is the specific value of k^{th} unmatched explanatory variable for j^{th} observation in the i^{th} stratum; α_i is the stratum-specific interpretation term reflecting the different combination effects of confounding variables for different strata; and β_k is estimated parameters for unmatched explanatory variables.

The conditional likelihood for each stratum i is based on the matched case-control design that the case is the one with the row vector x_{i0} and the controls are those with the other row vectors x_{ij} $(j = 1, 2, \cdots, J)$ for k unmatched explanatory variables. Because each observation within the stratum shares the same characteristics of the confounding variables, the effects of the confounding variables on conditional probability cannot be estimated. The conditional likelihood $L(Y_i|\beta_k)$ of the stratum i can be calculated as

$$L(Y_i|\beta_k) = \left[1 + \sum_{j=1}^{J} exp\left(\sum_{k=1}^{K} \beta_k (x_{ijk} - x_{i0k})\right)\right]^{-1} \tag{5}$$

where x_{i0k} is the value of x_k for a case in the i^{th} stratum, and x_{ijk} is the value of x_k for the j^{th} matched control in the i^{th} stratum. Because the strata are assumed to be independent from each other, the conditional log-likelihood function $LL(Y|\beta_k)$ over the population of *I* strata can be written as (Schlesselman, 1982)

$$LL(Y|\beta_k) = -\sum_{i=1}^{I} ln\left[1 + \sum_{j=1}^{J} exp\left(\sum_{k=1}^{K} \beta_k (x_{ijk} - x_{i0k})\right)\right] \tag{6}$$

Because the interpretation term α_i in the above equation cannot be estimated, the absolute probability of crash occurrence cannot be calculated in a matched case-control study. Alternatively, the odds ratio is calculated to evaluate the relative effects of unmatched explanatory variables on crash occurrence. For a dummy variable, the odds ratio represents the odds that a motorcycle crash will occur given a roadway characteristic k $(x_k = 1)$ compared to the odds of crash occurring in the absence of that roadway characteristic ($x_k = 0$), holding other variables constant. The odds ratio for a dummy variable can be written as

$$OR(x_k) = \frac{\Pr(y_{i0} = 1|x_k = 1, Z)/[1 - \Pr(y_{i0} = 1|x_k = 1, Z)]}{\Pr(y_{i0} = 1|x_k = 0, Z)/[1 - \Pr(y_{i0} = 1|x_k = 0, Z)]} = exp(\beta_k) \tag{7}$$

where Z represents the vector of explanatory variables other than x_k, and β_k is the estimated parameter for dummy variable x_k. Based on this definition, the odds ratio can be used as the direct estimation of the CMF.

3.3.3 Case and Control

In this study, 440 roadway corridors in urban and/or suburban areas with street lighting data were identified based on the following criteria: (1) roadway sections between two successive signalized intersections, (2) 600 ft or longer, (3) equipped with High Pressure Sodium (HPS) light bulbs, and (4) no upgrade on street lighting in past several years. To exclude the influence from adjacent signalized intersections, a 250-ft buffer was subtracted from the two ends of the roadway corridors.

The 440 measured corridors were split into 2,440 segments with a uniform length of 1,200 ft. Nighttime crash data for 2011–2014 was matched to each segment. Within the 2,444 segments, a case was defined as a segment in which at least one nighttime crash occurred, and a control was defined as a segment in which no nighttime crashes occurred. It is noted that segment length should be carefully determined; a reasonable length cannot be too short or too long. In the former case, the ratio of the number of cases to the number of controls would be too small, and zero-inflated observations would be produced. Therefore, the analysis power would greatly decrease. In the latter case, the variation of each risk factor within a segment would be large, whereas each risk factor should be the same or similar within the segment; thus, the inference would be biased. The segment length of 1,200 ft was used in this study to satisfy the above requirement.

A segment was allocated as a case or control for each year; thus, it could be a case for one year and a control for another year. The detailed number of cases and controls in different years are provided in Table 7. The measured street lighting points that fall into the same segment were calculated for the illuminance mean, the illuminance standard deviation, and the illuminance maximum-minimum ratio (max-min ratio) of the segment. The illuminance standard deviation and the max-min ratio were used as two measurements of the illuminance uniformity of the segment.

Table 7. Number of Cases and Controls by Year

Year	Number of Cases	Number of Controls	Total
2011	279	332	611
2012	272	339	611
2013	301	310	611
2014	319	292	611
Total	1,171	1,273	2,444

3.3.4 Confounder Matching

In this study, confounders were defined as variables associated with both the illuminance mean and nighttime crashes. The confounders must be controlled so that the effects of illuminance mean on nighttime crashes can be isolated and properly addressed. Several variables can be the potential confounding variable of the illuminance mean, such as the illuminance standard deviation, speed limit, number of lanes, AADT, and max-min ratio. To identify the real confounding variables of the mean of illuminance, a Pearson's correlation test was conducted. The correlation matrix is shown in Figure 2.

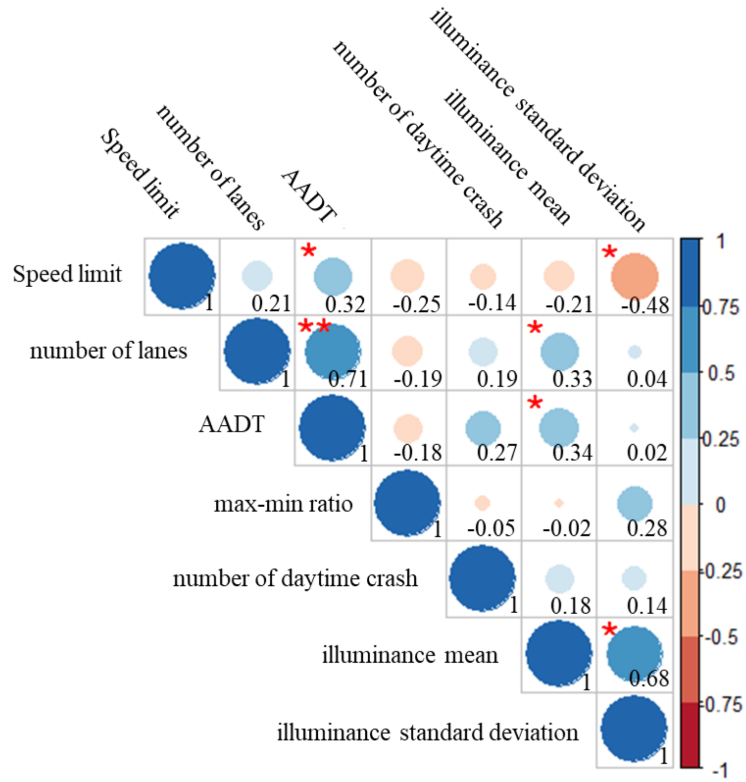

Figure 9. Explanatory Variable Correlation Matrix

It is noted that the illuminance mean is moderately and positively correlated with the illuminance standard deviation, as indicated by a Pearson's coefficient of 0.68. The reason is that the illuminance standard deviation tends to be smaller when the illuminance mean is close to zero given that the horizon illuminance is non-negative data. As the illuminance mean increases, the distribution of the horizon illuminance gets wider and, thus, the illuminance standard deviation becomes greater. In addition, the significant effects of the illuminance standard deviation on nighttime crash risk were confirmed by conducting a negative binomial model. The results showed that the illuminance standard deviation was positively associated with the number of nighttime crashes, which is consistent with previous studies (Yang et al., 2019b; Zhao et al., 2015). Therefore, the illuminance standard deviation was defined as a confounder of the illuminance mean in this study.

Further, the illuminance mean was weakly and positively correlated with the number of lanes and AADT, as indicated by a Pearson's coefficient of 0.33 and 0.34, respectively. Traffic agencies usually install high-level lighting systems on high-grade roadways that consist of more traffic lanes and bear higher traffic demand. It is noted that the number of lanes and AADT were strongly and positively correlated with each other with a Pearson's coefficient of 0.71, which implies that only one of these two variables could be matched to avoid the overmatching issue. In this study, AADT was chosen as another confounder in addition to the number of lanes so the confounding effects could be properly eliminated and, thus, the residual confounding issue does not happen. The significant effects of AADT on nighttime crash risk was also confirmed by conducting a negative binomial model. The results showed that AADT was positively associated with the number of nighttime crashes, which is consistent with previous studies (Yang et al., 2019b).

It is impractical to match the illuminance standard deviation and AADT by their exact values. Thus, the two confounders were categorized into five levels using the following three steps:

1. Transform the original data to a normal distribution by calculating square root values.
2. Estimate the standard deviation after the above transformation.
3. Categorize the transformed data based on +/- a standard deviation from the mean, as illustrated in Figure 10.

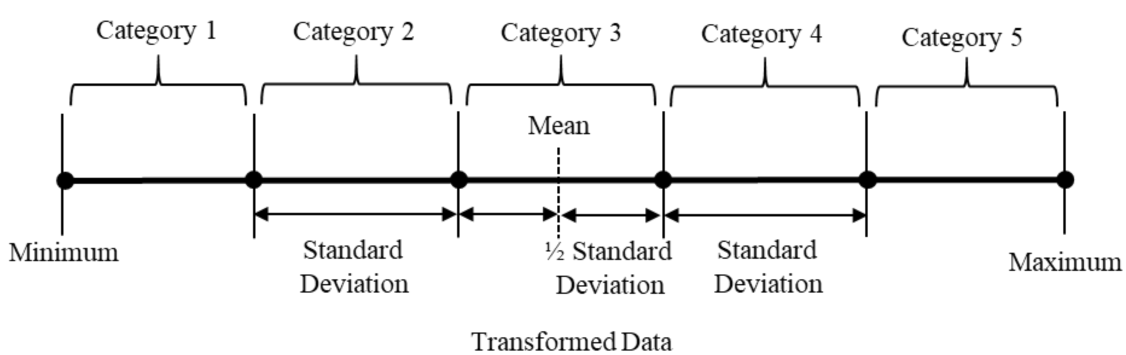

Figure 10. Data Categorization

With these matching categories, controls were randomly matched to cases by a case-control ratio of 1:1, which makes the power of design achieve approximately 90% (Woodward, 2013). Matched categories and sample sizes for AADT and the illuminance standard deviation are shown in Table 8.

After random matching, 1,046 cases and 1,046 controls were identified for modeling. The descriptive statistics of unmatched risk factors for cases and controls are presented in Table 9. The matched case-control method indicates the likelihood of a risk factor by assessing where this factor is disproportionally distributed between the cases and control. As shown, only 11.6% of cases (roadway segments that experienced at least one nighttime crash) had no access points, but about 26.6% of controls (roadway segments that experienced no nighttime crashes) had no access points. A conditional logistic model was estimated to investigate and quantify the risk associated with one of these factors while holding others constant. The estimation results are presented and discussed in the following section.

Table 8. Matched Categories and Sample Sizes for AADT and Illuminance Standard Deviation

Illuminance Standard Deviation	AADT										Total
	Category 1 (<7,569)		*Category 2 (7,569–19,321)*		*Category 3 (1,9321–36,481)*		*Category 4 (36,481–59,049)*		*Category 5 (>59,049)*		
	Case	**Control**	**Case**	**Control**	**Case**	**Control**	**Case**	**Control**	**Case**	**Control**	
Category 1 (<0.067)	10	30	39	73	30	54	44	40	4	0	324
Category 2 (0.067–0.240)	13	3	55	80	24	31	13	10	2	1	232
Category 3 (0.240–0.518)	24	24	41	35	166	218	246	218	31	17	1020
Category 4 (0.518–0.902)	44	45	104	97	117	165	146	102	3	1	824
Category 5 (>0.902)	10	26	5	3	0	0	0	0	0	0	44
Total	101	128	244	288	337	468	445	374	40	19	2,444

Table 9. Descriptive Statistics of Key Variables in Matched Case-Control Study

Variable Description	Case (n = 1,046)		Control (n = 1,046)	
	Mean	***SD***	***Mean***	***SD***
Illuminance mean less than 0.5 fc indicator (1 if illuminance mean less than 0.5 fc in this segment; 0 otherwise)	0.371	0.483	0.358	0.480
Illuminance mean 0.5–1.0 fc indicator (1 if illuminance mean 0.5–1.0 fc in this segment; 0 otherwise)	0.499	0.500	0.505	0.500
Illuminance mean greater than 1.0 indicator (1 if illuminance mean greater than 1.0 fc in this segment; 0 otherwise)	0.130	0.337	0.138	0.345
No access point indicator (1 if segment has no access point; 0 otherwise)	0.116	0.320	0.266	0.442
One or two access points indicator (1 if segment has one or two access point[s]; 0 otherwise)	0.487	0.500	0.486	0.500
Three or four access points indicator (1 if segment has three or four access points; 0 otherwise)	0.288	0.453	0.187	0.390
Five or more access points indicator (1 if segment has five or more access points; 0 otherwise)	0.110	0.313	0.061	0.240
Max-min ratio no less than 10 indicator (1 if ratio of maximum illuminance to minimum illuminance is greater than 10 in this segment; 0 otherwise)	0.887	0.317	0.855	0.353
Daytime crash indicator (1 if segment experienced at least one daytime crash; 0 otherwise)	0.923	0.267	0.592	0.492
Wider shoulder width indicator (1 if segment shoulder width greater than 16 ft; 0 otherwise)	0.420	0.494	0.520	0.500
High-density commercial location indicator (1 if segment located in high-density commercial area; 0 otherwise)	0.239	0.427	0.245	0.430
Year 2012 indicator (1 if segment is observed in 2012; 0 otherwise)	0.231	0.422	0.270	0.444

3.3.5 Model Estimation

The software package STATA 15 was used to estimate the conditional logistic model using the matched case-control data. The max-min ratio was significant at a 90% confidence level; all other explanatory variables are significant at a 95% confidence level, including illuminance mean, number of access points, daytime crashes, shoulder width, land use type, and observed year. The detailed estimation results and confidence interval (CI) of odds ratio are provided in Table 10.

Table 10. Matched Case-Control Conditional Logistic Regression Model

Variable	Coefficient	*z*	*p*-value	OR	95% CI of OR
Illuminance mean less than 0.5 fc indicator	**Baseline**				
Illuminance mean 0.5–1.0 fc indicator	-0.387	-2.04	0.041	**0.679**	[0.468, 0.984]
Illuminance mean greater than 1.0 indicator	-0.543	-2.30	0.021	**0.581**	[0.367, 0.922]
Max-min ratio no less than 10 indicator	0.330	1.81	0.070	**1.391**	[1.031, 1.876] *
No access point indicator	**Baseline**				
One or two access points indicator	0..042	3.15	0.002	1.043	[1.016, 1.070]
Three or four access points indicator	1.009	5.33	0.000	2.744	[1.893, 3.979]
Five or more access points indicator	1,382	5.17	0.000	3.983	[2.360, 6.724]
Daytime crash indicator	2.053	12.70	0.000	7.794	[5.68, 10.7]
Wider shoulder width indicator	-0.270	-2.03	0.042	0.763	[0.587, 0.991]
High-density commercial location indicator	-0.293	-2.05	0.040	0.746	[0.565, 0.987]
Year 2012 indicator	-0.285	-2.35	0.019	0.752	[0.593, 0.954]
Model Statistics					
Number of observations					2,092
Log-likelihood					-531.446
Pseudo R^2					0.267

* 90% Confidence Interval

The illuminance mean was aggregated into three levels: < 0.5 fc, 0.5 fc–1.0 fc and > 1.0 fc. As shown in Table 9, when the illuminance mean is less than 0.5 fc, the percentage of cases is more than the percentage of controls by 1.3% (37.1% for cases and 35.8% for controls), but when the illuminance mean is 0.5–1.0 fc, the percentage of cases is less than the percentage of controls by 0.6% (49.9% for cases and 50.5% for controls). Similarly, when the illuminance mean is more than 1.0 fc, the percentage of cases is less than the percentage of controls by 0.8% (13.0% for cases and 13.8% for controls). This trend implies that nighttime crashes are less likely to be observed on segment with relatively great illuminance mean values, i.e., no less than 0.5 fc.

The model estimation results presented in Table 10 confirmed this trend. The coefficients of all the illuminance mean levels are significantly negative at a confidence level of 95% and become smaller with an increase in the illuminance mean. This indicates that the increase in the illuminance mean is more likely to decrease the relative risk of nighttime crashes on roadway segments. This finding is consistent with previous studies (Jackett and Frith, 2013; Sullivan and Flannagan, 2007) and the common sense that driver vision and sight distance considerably improves as the illuminance mean increases, which contributes to a lower nighttime crash risk. The odds ratio gives a significant (95% confidence level) and valid (95% confidence interval excluding one) estimation on the CMF for the illuminance mean. If the illuminance mean on a roadway segment increases from < 0.5 fc to 0.5 fc–1.0 fc, the relative risk of nighttime crashes is 0.679 times as many as before. If the illuminance mean on a roadway segment increases from < 0.5 fc to >1.0 fc, the relative risk of nighttime crashes is 0.581 times as many as before. Compared with the previous study that developed joint CMFs for horizontal illuminance, including both the illuminance mean and standard deviation (Yang et al., 2019b), the CMFs for the illuminance mean was independently derived in this study by eliminating the effects of the illuminance standard deviation during the matched case-control design. Therefore, the derived CMFs can be used to better understand and assess the isolated effects of the illuminance mean on nighttime crashes.

In addition to the illuminance standard deviation identified as a confounder and matched in this study, the ratio of the maximum illuminance to the minimum illuminance also is an illuminance uniformity measurement. The coefficient of a max-min ratio no less than 10 indicator is significantly positive at a confidence level of 90%. If the max-min ratio of a segment is no less than 10, the relative risk of nighttime crashes of this segment is 1.391 times as many as the segments with the max-min ratio less than 10. This aligns with the standard suggested by the Illuminating Engineering Society (IES) that the illuminance max-min ratio on the roadway must not exceed 10 (IESNA, 1993).

3.4 Nighttime Crash Severity Diagnosis Model

3.4.1 Problem Statement

Most previous studies (see Table 2) focused on exploring the impacts of lighting patterns on nighttime crash occurrence; few investigated the impacts of lighting patterns on injury severity, which is another important risk measure of nighttime crashes. Wei et al. (2016) concluded that an increase in intersection illuminance was an effective countermeasure to reduce the probability of fatality and severe injury in a nighttime crash at a signalized intersection, especially for pedestrian- or bicycle-involved, head-on, and angle crashes. For these crash types, intersection illuminance kept at 0.9 fc or higher tends to reduce the probability of fatality, severe injury, and non-severe injury by 10.7%, 9.0%, and 6.3%, respectively. If alcohol or drugs were involved, these reductions were even larger. Xin et al. (2018) applied a random parameter ordered probit model to describe the connection between the injury severity of a nighttime crash on a roadway segment and the lighting level of a 1,000-ft zone in the upstream of a crash. They found that increasing average horizontal illuminance from 0.4 fc or less to 0.4–0.8 fc can significantly reduce the probability of injury severity in a nighttime crash by 4.05% (fatal or incapacitating injury) and 6.62% (non-incapacitating injury or possible injury).

Illumination distribution along a roadway corridor presents an intricate pattern due to lighting depreciation, obstacles (e.g., tree branches), and external lighting resources. Figure 11 shows an example of street lighting patterns influencing nighttime driving safety. Along the car travel route, the horizontal lighting level suddenly changes from 0.1 fc to 0.9 fc; the driver needs several seconds to adapt to the new lighting condition, and, during this period, the driver's vision deteriorates. If an object presents in front of the car during this time, the driver may not be able to avoid a collision, potentially resulting in a severe injury. The average illuminance used in studies by Wei et al. (2016) and Xin et al. (2018) represented the overall lighting level of the segment and could not capture this local feature. Ratio-based uniformity measures (max/min and average/min) maintain an extreme value (minimum or maximum) and may misinterpret a dangerous pattern. Traditional photometric statistics are applicable in a grid analysis for a narrow zone (single lighting pole design or isolated intersection), but they may not be appropriate for a safety diagnosis for lighting patterns along a roadway corridor.

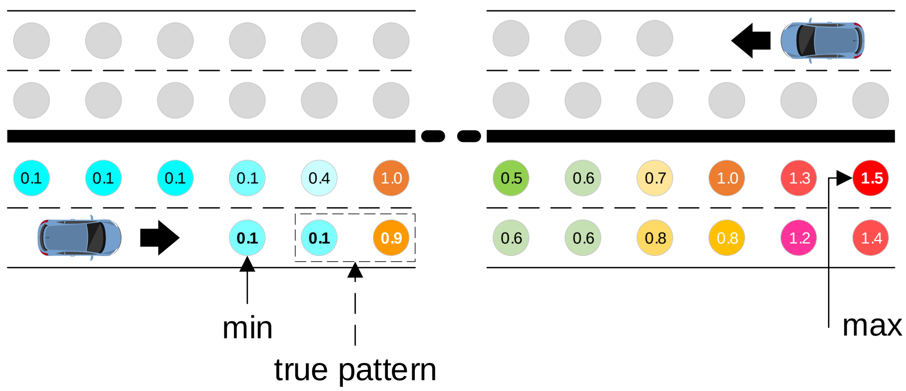

Figure 11. Example of Illuminance Patterns Influencing Driver Vision

This study aimed to develop a novel method to diagnose street lighting patterns that result in a severe injury in a nighttime crash. A machine-learning algorithm was used to process the high-resolution lighting data and increase prediction accuracy.

3.4.2 Data Preparation

Data for 2,862 nighttime crashes that occurred between 2012 and 2014 were collected from roadway corridors with measured lighting data. For each crash, a rectangle buffer was created towards the upstream direction of the crash in ArcGIS. The measured lighting data that fell into the buffer were spatially matched to the associated crash. The buffer represents the range of lighting patterns that influences crash injury. To capture the safety impacts of lighting patterns more accurately, each buffer was divided into several sub-zones. The mean of matched lighting data (horizontal illuminance) was calculated for each sub-zone to represent the lighting level of the local zone. An example of the lighting buffer and sub-zones is shown in Figure 12.

In addition to lighting and crash data, geometric, traffic control, and environmental data were also collected and matched to nighttime crashes. The collected data are described in Table 11.

(a) Lighting pattern associated with a non-severe injury crash

(b) Lighting pattern associated with a severe injury crash

Figure 12. Lighting Buffers and Sub-zones Associated with Nighttime Crashes

Table 11. Description of Collected Data

Category	Values
Crash severity	1 – severe injury (fatal or incapacitating injury), 0 – others
Lighting	Average horizontal illuminance (fc) of sub-zones
Weather condition	Clear, cloudy, rainy, fog, smog, or smoke
Speed limit (mph)	30, 35, 40, 45, 50, 55, 60
Drug-, alcohol-involved	Drug-involved, alcohol-involved, drug-and-alcohol-involved
Junction event	On roadway, off roadway
Site location	Not at intersection/railroad crossing/bridge, at intersection, influenced by intersection, driveway access, railroad, bridge, on-ramp, off-ramp
Related junction	Non-junction, intersection, intersection-related, driveway/alley access-related, railway grade crossing, entrance/exit ramp, crossover-related, shared-use path or trail, through roadway, narrative
Traffic device	No controls, school zone sign/device, traffic control signal, Stop sign, Yield sign, flashing signal, railroad crossing device, person (including flagman, officer, guard, etc.), warning sign and narrative
Work zone	Yes, no
Roadway surface condition	Dry, wet, mud, dirt or gravel roadway, water, narrative
Vehicle driveway	Two-way not divided; two-way not divided with continuous left-turn lane; two-way divided, unprotected median; two-way divided, positive median barrier
Number of lanes (bidirectional)	1-10

3.4.3 Methodology

The Support Vector Machine (SVM) classifier was applied to predict injury severity based on prevailing lighting patterns, roadway conditions, and environmental characteristics. Figure 13 illustrates the procedure for the prediction model development; the basic steps include data processing, SVM model training, and SVM model testing. A grid search was conducted to find the "best" model that reaches the highest prediction performance.

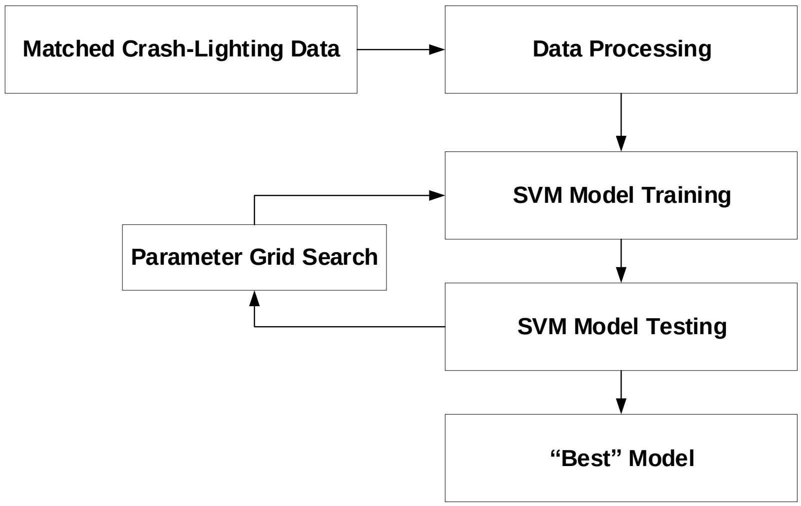

Figure 13. Flow Chart of Technical Approach for Model Development

Data Preprocessing

- *Feature Selection* – Model input features should be the factors that contribute to crash injury severity. Crash-, person-, and vehicle-specific factors were excluded from the feature selection. Although these variables have significant impacts on crash injury severity, it is difficult to collect data for the model to diagnose lighting patterns, as some features cannot be observed before crash occurrence (for example, impact point on vehicles in a crash). To minimize the cost for implementing the model, only factors that can be retrieved from roadway inventory databases are included in the model, such as geometric data, traffic control devices, and environmental factors.
- *Encoding and Normalization* – Feature variables can be divided into two categories: categorical and continuous. Unlike some machine learning algorithms (e.g., decision tree), SVM cannot directly learn from categorical data. A data transformation from categorical data to a numerical format is conducted. Since no ordinal relationship exists in categorical variables, one hot encoding is applied to categorical variables to provide a new representation for machine learning tasks. In one hot encoding, each unique value in a category is converted to a new binary variable, and a categorical variable with *k* different unique values is converted to *k* binary variables. For continuous variables, all variables are scaled in [0, 1] to prevent variables in different magnitudes from affecting model performance. Assume x is a continuous variable, min (x) and max (x) are the minimum and maximum values; the scaled value x' is given below:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (8)$$

Support Vector Machine

SVM (Chang and Lin, 2011) is a highly-preferred machine learning algorithm that produces significant accuracy with less computation. It defines a hyperplane to distinguish the data in a *d*-dimensional space, where *d* is the number of variables. The hyperplane is found by maximizing the margin for all classes to provide more classification confidence. For the binary classification, given *n* observations:

$$(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), y_i \in \{0, 1\} \quad (9)$$

where $y_i = 1$ indicates a serve injury crash; $y_i = 0$ indicates others; x_i represents the input features. The decision hyperplane is defined as:

$$f(x) = w^T x + b \quad (10)$$

where *w* is the weight vector and *b* is the intercept. To solve the following optimization problem SVM training becomes:

$$Min\ \frac{1}{2} w^T w + C \sum_{i=1}^{n} \varepsilon, C > 0$$

(11)

$$\text{Subject to } y_i(w^T \emptyset(x_i) + b) \geq 1 - \varepsilon$$

where $\phi(x_i)$ maps x_i into a higher dimension space; ε is the error constraint and *C* is the regularization parameter. By introducing a Lagrange multiplier, the optimization problem becomes:

$$Min\ \frac{1}{2} \alpha^T \alpha Q - e^T \alpha$$

(12)

$$\text{Subject to } y^T \alpha = 0, 0 < \alpha_i < C$$

where *e* is a vector of all ones; *Q* is an *n* x *n* matrix, $Qij \equiv y_i y_j K(x_i, x_j)$, $K(x_i, x_j)$ is the Radial-Basis Function (RBF) kernel, shown in Eq. 6.

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad (13)$$

where γ is kernel parameter. After solving the optimization problem, the hyperplane is defined as:

$$f(x) = w^T x + b = \sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b \tag{14}$$

This hyplane is the decision boundary of the model, which helps for label prediction.

Grid Search

A grid search is conducted to find the "best" SVM model that can most accurately predict the injury severity of nighttime crashes based on input features. The search scope includes four parameters, including buffer length (*l*), number of sub-zones (*m*), SVM kernel parameter (*g*), and SVM regularization parameter (*C*). The first parameter (*l*) defines the influence area of lighting patterns on a crash. The number of sub-zones describes lighting patterns more precisely within a buffer. The grid search traverses the given ranges for the two variables and finds the best values. The search range is given below:

- *Buffer length* – 0.1, 0.125, 0.15, 0.175, 0.2, 0.225 miles
- *Number of sub-zones* – 5, 10, 15

The buffer and sub-zones were produced for each crash by combining buffer lengths and zone numbers. The lighting data and other features were matched to each generated buffer and associated sub-zones. Finally, 18 crash-lighting datasets were generated.

SVM parameters g and C are model hyperparameters that may depend on the data but cannot be estimated from the data. The grid search is used to find the optimal value of hyperparameter to address the most accurate model. For each crash-lighting dataset, 80% of observations are randomly selected to train the SVM model. The LIBSVM JAVA Library (Chang and Lin, 2011) is used to perform model training. The grid search traverses different SVM parameters g and c in the range of $[2^{-10}, 2^{10}]$ to explore more space and find the hyperplane that provides better discriminative classification.

To identify the well-trained model with different parameters, each model with the specified parameter is validated using tenfold cross-validation to test the effectiveness of the model. Since severe injury crashes are rare events, the crash-lighting datasets are unbalanced; non-severe injury crashes (86% of the sample) occur more than severe ones (14%). The unbalance may result in inaccurate prediction. To address this issue, the F1 score (Powers, 2011) combines both the precision p and the recall r and better reflects the effectiveness of a model. Precision p is true positive samples over the sum of true positive and false-positive samples. r is true positive samples over the sum of true positive and false negative samples. The F1 score is a harmonic average of the precision and recall, as shown below:

$$\begin{gathered} Precision\ (p) = \frac{True\ positive}{True\ positive + False\ positive} \\ Recall\ (r) = \frac{True\ positive}{True\ positive + False\ Negative} \end{gathered} \tag{15}$$

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

After training, the SVM model is applied to the remaining 20% of observations to test the prediction accuracy of the model with the highest validation results. To avoid randomness, each crash-lighting dataset is trained 10 times, and the average testing accuracy on the trained model is calculated as the model testing accuracy.

3.4.4 Results

Grid search results are presented in Table 12. Each row represents the best training result for the 18 combinations of buffer length and sub-zone number. The results indicate that the trained SVM models had similar performance. The SVM model trained with a buffer of 0.125 miles and 10 sub-zones had the best performance.

Table 12. Grid Search Results

Dataset ID	Buffer Length (mi)	Number of Zones	Accuracy (%)	Sample Size
1	0.1	5	78.8	2,416
2	0.1	10	78.3	2,400
3	0.1	15	77.8	2,364
4	0.125	5	78	2,381
5	***0.125***	***10***	***80.1***	***2,375***
6	0.125	15	78.6	2,363
7	0.15	5	78.9	2,329
8	0.15	10	78.8	2,326
9	0.15	15	79.1	2,315
10	0.175	5	78.2	2,297
11	0.175	10	79	2,292
12	0.175	15	78.1	2,281
13	0.2	5	77.7	2,242
14	0.2	10	78.1	2,241
15	0.2	15	78.6	2,235
16	0.225	5	78.9	2,206
17	0.225	10	77.9	2,194
18	0.225	15	78.1966	2,191

An example of a lighting pattern diagnosis is provided in Table 13. The best model predicted the injury severity of the two crashes. For the possible injury crash, the predicted probability of label "1" is 5.6%; the predicted probability for the fatal crash is 73.3%. This example shows that the SVM model effectively classified the crash severity level based on lighting patterns and other features. By checking the photometric statistics for the two crashes, it was found that the possible injury crash experienced a higher lighting level (1.298 fc) and much better uniformity (11.4 for avg/min and 19.5 for max/min) compared to the fatal crash. Figure 12(a) shows that the possible injury crash with the lighting level along the buffer was kept at a high level. Five zones (50%) were higher than 1.0 fc, and four fells in the range of 0.75–1.0fc. Only one zone was lower than 0.7 fc. In contrast, Figure 12(b) represents the lighting pattern of the fatal crash and expresses a diverse lighting distribution; three sub-zones had a lighting level of 0.5 fc or less, and, most importantly, a very low-lit zone (≤ 0.5fc) directly connected to a high-lit zone (≥ 1.0fc). The contrast illumination pattern causes high risk for drivers. The SVM model captured the traits of the two lighting patterns.

Table 13. Comparison of Lighting Pattern Diagnosis for Two Crashes

Category	Item	Crash I	Crash II
Crash Features	Figure	Figure 12(a)	Figure 12(b)
	True severity	Possible Injury	Fatal
SVM Model	Buffer length	0.125 mi	
	Number of sub-zones	10	
	Predicted probability of label "1"	5.6%	73.3%
	Final label	0 – Non-severe Injury	1 – Severe Injury
Statistics for whole buffer	Average horizontal illuminance (fc)	1.298	0.705
	Average/Min	11.4	43.5
	Max/Min	19.5	134.5

The developed SVM model effectively captured the traits of lighting patterns that are complexly distributed over space. Even without knowing factors during and after a crash, the developed SVM model effectively predicted the crash injury severity based on lighting patterns. SVM model inputs include lighting patterns and other data that can be quickly retrieved from roadway inventory databases. The model is easy to implement for scanning the lighting patterns of a roadway corridor and identifying zones with high injury risk if a crash occurs.

The grid search defined the best lighting pattern diagnosis settings: 0.125 miles of buffer length and 10 sub-zones within one buffer. However, this study did not consider the lateral range of lighting patterns. In addition, only the mean of horizontal illuminance for each sub-zone was used to describe the lighting patterns. The simple measure may lose lighting information; more informative lighting measures, such as a histogram, and configurable buffer width will be considered in a future study.

Chapter 4: Case Studies

The developed tools were applied on selected corridors to demonstrate and validate the tool's operation and performance. Corridors were selected from CUTR's lighting data inventory, including additional lighting data collected for FDOT District 7 and Johnson, Mirmiran & Thompson (JMT) in 2020. The diagnosis results were provided to FDOT and JMT for supporting their decisions related to street lighting management.

4.1 Lighting Diagnosis and Proposed Lighting Improvement Evaluation on W Busch Boulevard

W Busch Boulevard in Tampa is a principle arterial in an urban area. Characteristics of Busch Boulevard are shown in Table 14.

Table 14. Busch Boulevard Characteristics

Item	Description
Functional Classification	Principal Arterial
Area type	Urban
Boundary	Dale Mabry Hwy – Florida Ave
Roadway ID in database	10310000
BMP – EMP	0.087 – 2.847
Length	2.76 mi
Number of through lanes	4 (bidirectional)
Speed limit	45 mph
AADT	42,500
Lighting technology	HPS
DOT lighting standards	Mean ≥ 1.5 fc, max/min ≥ 10, avg/min ≥ 4

The CUTR team collected lighting data (horizontal illuminance) on this segment using the ALMS. The illuminance points were read every 10 ft per lane along the corridor, and data for 12,286 illuminance points were collected. The photometric statistics in Table 15 clearly indicate that the lighting patterns do not satisfy FDOT standards in either mean and uniformity.

Table 15. Photometric Statistics for Whole Segment on W Busch Boulevard

Item	Description
Average Illuminance	0.72 fc (< 1.5 fc)
Avg/Min	198.77 (>>4)
Max/Min	391.78 (>>10)

A heatmap was produced based on the measured illuminance data, as shown in Figure 14, and indicates that the lighting pattern is unbalanced along the segment. The lighting level on the right side is significantly higher than that on left side; thus, statistics for the whole segment may not be reasonable to describe the lighting condition along the corridor.

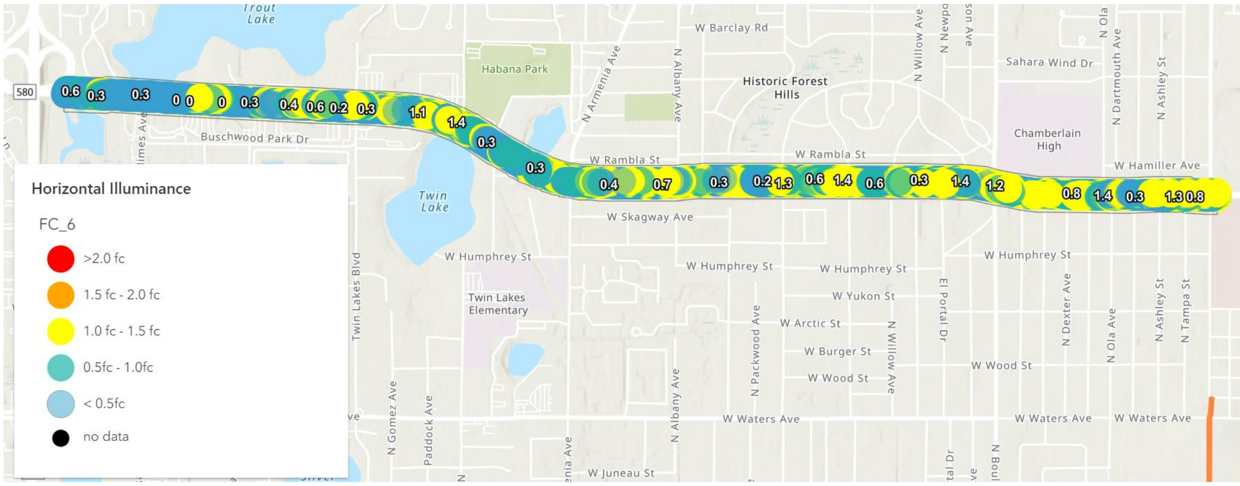

Figure 14. Heatmap for W Busch Boulevard

The Automatic Lighting Diagnosis Module was applied to split the segment into sub-sections based on similarity of average lighting level. The default values of the parameters (Table 6) were used to conduct the diagnosis; the diagnosis results are shown in Figure 15.

The diagnosis module recognized the diversity of lighting patterns and split the whole segment into four sections. Zone 1 represented the “worst” portion on W Busch Boulevard; the average lighting level in Zone 1 was 0.25 fc, significantly lower than the FDOT standard (1.5 fc for major arterials). The section also had “poor” uniformity: max-min ratio (391.8) and avg-min ratio (70.6) were much higher than FDOT standards (10 and 4, respectively). Zone 2 was brighter than Zone 1; its average lighting level is 0.73 fc, still lower than the standard. The uniformity measures were 30.4 (max/min) and 15.6 (avg/min). The lighting pattern in Zone 4 was slightly better than Zone 2, with a mean of 0.84 fc, max/min of 12.7, and avg/min of 7.5. However, the two zones still did not satisfy FDOT standards. Zone 3 had the “best” lighting performance; the average lighting level was 1.04 fc, lower than FDOT standards. However, the uniformity satisfied FDOT standards in either max/min (5.2 < 10) or avg/min (3.8 < 4). The Automatic Lighting Diagnosis System produced histograms of horizontal illuminance for each zone, which show that 65% of the area in Zone 1 was lower than 0.25 fc; 40% of the area in Zone 3 was 1.0–1.5 fc, and Zone 2 and Zone 4 had a relatively uniform distribution of lighting level over categories.

The crash risk prediction module was applied to analyze the nighttime crash risk for each section. Zone 3, which had the best lighting performance, was expected to have the lowest nighttime crash frequency (4.3 crashes per year per mile) in the target year (2020). Zone 1 and Zone 2 had similar safety performance (expected nighttime crash frequency in 2020)—11 crashes per year per mile and 10 crashes per year per mile, respectively. Zone 4 had the “worst” safety performance; it was expected to have 14.8 crashes per year per mile in 2020.

In summary, the lighting patterns on W Busch Boulevard (Dale Mabry Highway to N Florida Avenue) did not meet FDOT standards; most of the segment had a high nighttime crash risk (≥ 10 crashes per year per mile).

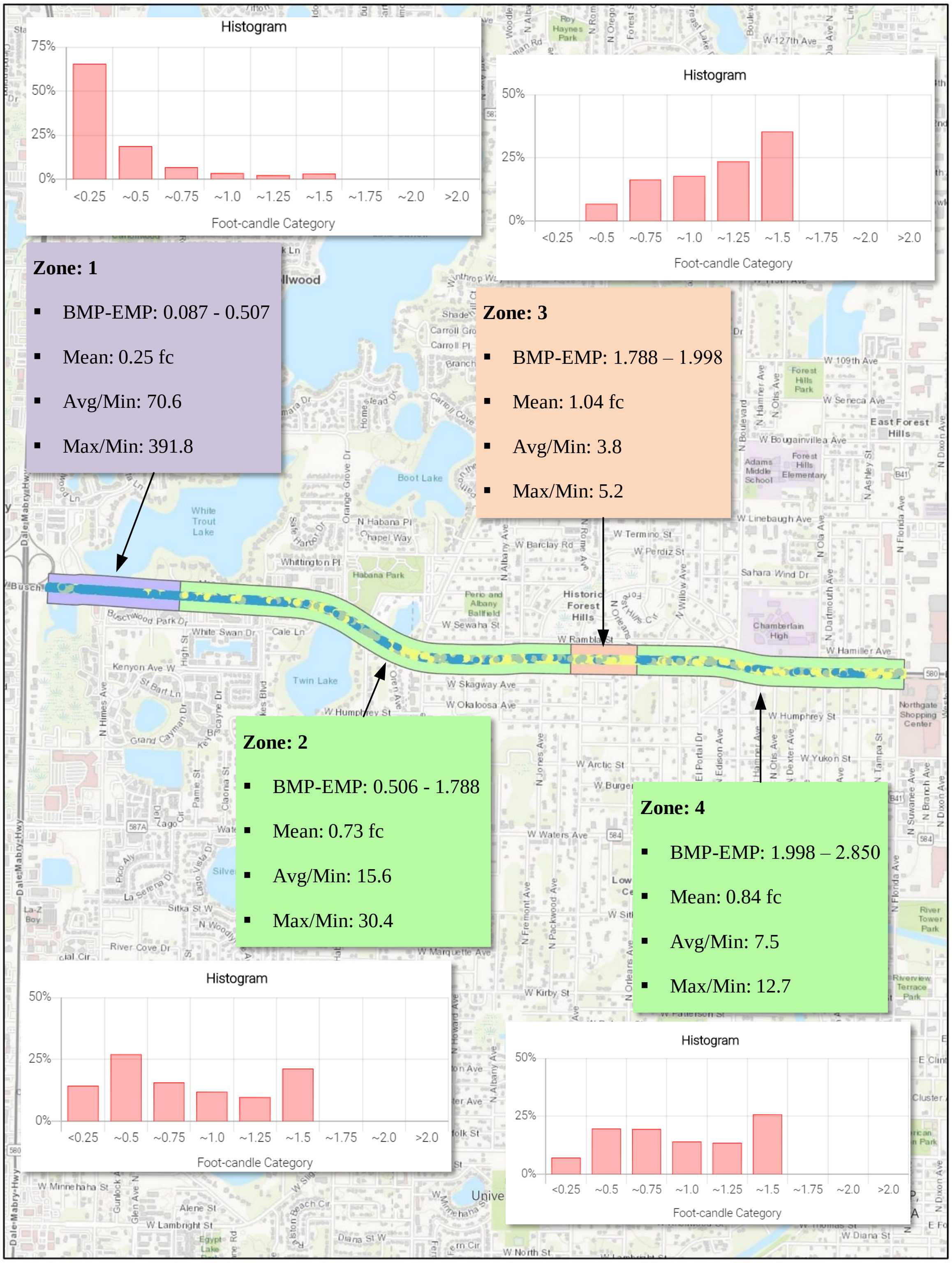

Figure 15. Automatic Lighting Diagnosis on W Busch Boulevard

Figure 16. Nighttime Crash Risk Prediction for W Busch Boulevard

Upgrading the street lighting system is a potential countermeasure to improve nighttime safety on W Busch Boulevard. The CMFs developed in this study were used to estimate nighttime crash reduction due to the lighting system upgrade. Assuming a lighting upgrading project is proposed to increase the lighting patterns to FDOT standards, the estimated nighttime crash reduction is given in Table 16.

Table 16. Estimation of Nighttime Crash Reduction with Proposed Lighting Improvement

		Zone 1	Zone 2	Zone 3	Zone 4
Mean of Illuminance	Existing	0.25 fc	0.73 fc	1.04 fc	0.84 fc
	Proposed	1.5 fc	1.5 fc	1.5 fc	1.5 fc
	CMF_M	0.581	0.856	1	0.856
Max/Min	Existing	391.8	30.4	5.2	12.7
	Proposed	10	10	10	10
	CMF_U	0.718	0.718	1	0.718
Expected Nighttime Crash Frequency (per yr)		4.6	12.8	0.9	12.6
$CMF_M \times CMF_U$		0.417	0.615	1	0.615
Crash Reduction Factor		0.583	0.385	0	0.385
Estimated Crash Reduction (per yr)		2.7	4.9	0.0	4.9
Total Crash Reduction		12.5 crashes per year			

4.2 Validation of LED Street Lighting Upgrade

E 7th Avenue from Nuccio Parkway to 24th Street in Tampa is a historic community with many bars, clubs, and recreational stores. The street experienced significant pedestrian crashes and injuries, especially at night. The City of Tampa upgraded the street lighting technology from HPS to LED in 2015. The CUTR team help the City measure the lighting patterns on 7th Avenue in 2014 (before upgrading), 2015 (after upgrading), and 2019 (the second measure after upgrading) through a contract with FDOT District 7. This study diagnosed and compared lighting patterns collected in three stages to verify LED lighting patterns. The basic information of E. 7th Avenue is shown in Table 17.

Table 17. E 7th Avenue Characteristics

Item	Description
Functional Classification	Major Collector with frequent pedestrian traffic
Area type	Urban
Length	One mile
Number of through lanes	2 (bidirectional)
Speed limit	30 mph
AADT	5,600
Lighting technology	HPS (before 2015), LED (since 2015)
DOT lighting standards	Mean ≥ 1.0 fc, max/min ≥ 10, avg/min ≥ 4

The system produced heatmaps for the three lighting measures, as shown in Figure 17. The comparison shows the following findings:

- The mean of horizontal illuminance satisfied FDOT standards in the three stages. The mean of horizontal illuminance decreased after LED upgrading (1.71 fc to 1.3 fc).

- The uniformity improvement was very significant (avg/min from 106.9 to 9.7 and 8.8, max/min from 143 to 10.5 and 9.4). Based on the max-min ratio, the uniformity satisfied FDOT standards after LED upgrading.
- According to the CMFs developed in this study (Table 10), improved uniformity can reduce nighttime crash frequency by 28%, and reduction of illuminance mean has no significant influence on safety performance.
- Depreciation of LED lighting performance was slight, as the lighting patterns in 2015 and 2019 are similar.

2014, HPS, E 7th Ave：

Max	2.289
Min	0.016
Avg	1.710
Avg/Min	106.875
Max/Min	143.063

2015, LED, E 7th Ave：

Max	1.410
Min	0.134
Avg	1.306
Avg/Min	9.746
Max/Min	10.522

2019, LED, E 7th Ave：

Max	1.410
Min	0.150
Avg	1.316
Avg/Min	8.773
Max/Min	9.400

Figure 17. Comparison of Lighting Patterns on E 7th Avenue, Tampa

Chapter 5: Summary and Conclusions

5.1 Summary and Conclusions

Roadway lighting is a conventional roadway infrastructure to ensure nighttime safety and security for multimodal road users (motorists, pedestrians, cyclists, transit passengers). To cost-effectively maintain a roadway lighting system, key tasks in infrastructure management include periodically measuring roadway lighting levels, diagnosing lighting performance based on collected data, and providing decision-making support for maintenance and improvement.

The ALMS developed by CUTR provides a low-cost and time-effective solution for collecting high-resolution lighting data for a big-scale roadway network. CUTR contracted with FDOT and JMT to measure lighting data; since 2012, 400 center-mile measurements have been completed using the ALMS. This lighting measurement is expanding to other districts.

This project developed a computer tool to effectively analyze big-scale lighting data. The tool provides three core functions – lighting pattern diagnosis, crash risk prediction, and data visualization. A summary of the functions is given in Table 18. The tool was developed on an ArcGIS web platform, and the analysis functions were coded as ArcGIS geoprocessing tools that can be accessed by desktop and web applications. A web-GIS application provides interactive interface to receive user command and present analysis results.

Table 18. Summary of Core Functions

Function	Algorithm/Technology	Description
Lighting Pattern Diagnosis	Hierarchical Clustering Model	Distinguish lighting patterns that do not satisfy FDOT standards
Crash Risk Prediction	• Safety Performance Function • Empirical Bayesian Model • CMF developed by Case-Control Study	• Predict nighttime crash frequency by lighting, traffic, and geometry conditions • Estimate nighttime crash reduction due to lighting pattern improvement
Data Visualization	Web-GIS	Present analysis results on GIS map

The developed tool was implemented on the segments studied by CUTR for FDOT and JMT in 2019. Two case studies are presented in this report to demonstrate the performance of tool in lighting management and evaluation projects. The tool successfully diagnosed the lighting pattern on W. Busch Boulevard in Tampa, predicted nighttime crash risk with existing lighting conditions, and estimated the benefit (crash reduction) of a proposed lighting improvement project. The tool was also used to compare the lighting patterns of E. 7th Avenue in Tampa over three years to validate the LED upgrading project, which showed that the tool reached Technology Readiness Level 7: Prototype Demonstrated in Operational Environment.

5.2 Implementation

The CUTR team will implement the tools and methods developed in this project in new lighting measurement and analysis tasks funded by FDOT and JMT. To implement the tool, CUTR will complete the following steps:

- *Debugging* – As a prototype, it is normal for errors and bugs to exist. The research team will continue to test the system to identify and fix bugs in the current version. A stable version is expected to be implemented in a production environment.
- *Standard Operation Procedure* – The research team will develop a Standard Operating Procedure (SOP) to guide the implementation step by step. The SOP will include importing raw data from the ALMS, converting them to a standard data format, conducting analysis based on needs, and producing formatted reports. Operators will be trained according to the SOP to provide a high-quality services in data collection, processing, and analysis.
- *Implementation* – The CUTR team will apply the SOP and tools in the lighting data collection tasks. Analysis results will be provided to FDOT/JMT for their decision-making in street lighting management. Experiences from the implementation will be used to update the system.

5.3 Future Study

Although the prototype successfully demonstrated its functions in lighting pattern analysis, there are some limitations requiring further study:

- The current lighting diagnosis algorithm considers only one measure (mean or uniformity) in calculating similarities. A study is needed to develop a new lighting diagnosis algorithm that recognize lighting patterns by mean and uniformity simultaneously to provide a more reasonable pattern recognition method.
- Crash risk prediction in the current version is based on nighttime crash frequency. Although the injury severity model was developed in this study, it was not integrated in the computer tool due to data availability in practice. A future study will develop a new crash risk index to combine the information of frequency and severity. The new crash risk prediction can provide a more reasonable measure for scaling the safety performance of street lighting patterns.
- The current computer tool can be used for single segment analysis; however, a decision-making support system is needed to analyze the lighting patterns for a large-scale area (i.e., city, county, or district). A future study would develop analysis functions at the planning level, including identifying segments with high nighttime crash risk, diagnosing lighting patterns (if lighting is a major cause), estimating benefits/cost for proposed lighting improvement projects, and ranking the proposed project based on benefit-cost ratios. FDOT managers can select top improvement projects from a ranking list to optimize street lighting management and maintenance.

References

AASHTO, 2011. *Roadway Lighting Design Guide*, 4th ed. American Association of State Highway and Transportation Officials, Washington, DC.

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transportation Research Record*, 1897 , 88–95.

Anarkooli, A.J., Hadji Hosseinlou, M., 2016. Analysis of the injury severity of crashes by considering different lighting conditions on two-lane rural roads. *Journal of Safety Research*, 56, 57–65. doi:10.1016/j.jsr.2015.12.003.

Bhagavathula, R., Gibbons, R.B., Edwards, C.J., 2015. Relationship between roadway illuminance level and nighttime rural intersection safety. *Transportation Research Record*, 2485, 8–15. doi:10.3141/2485-02.

Bhagavathula, R., Gibbons, R.B., Nussbaum, M.A., 2018. Effects of intersection lighting design on nighttime visual performance of drivers. *LEUKOS*, 14(1), 25–43.

Box, P.C., 1976. Effect of lighting reduction on an urban major route. *Traffic Engineering*, 46(10), 26–27.

Box, P.C., 1971. Relationship between illumination and freeway accidents. *Illumination Engineering*, 66(5), 365–393.

Bruneau, J.-F., Morin, D., 2005. Standard and nonstandard roadway lighting compared with darkness at rural intersections. *Transportation Research Record*, 1918, 116–122. doi:10.3141/1918-15.

Bullough, J.D., Donnell, E.T., Rea, M.S., 2013. To illuminate or not to illuminate: Roadway lighting as it affects traffic safety at intersections. *Accident Analysis & Prevention*. 53, 65–77. doi:http://dx.doi.org/10.1016/j.aap.2012.12.029

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3) , 27.

Davis, G.A., Davuluri, S., Pei, J., 2006. Speed as a risk factor in serious run-off-road crashes: Bayesian Case-Control Analysis with Case Speed Uncertainty. *Journal of Transportation Statistics*, 9(1) , 17.

Donnell, E.T., Porter, R.J., Shankar, V.N., 2010. A framework for estimating the safety effects of roadway lighting at intersections. *Safety Science*, 48(10) , 1436–1444. doi:10.1016/j.ssci.2010.06.008.

Edwards, C.J., 2015. Lighting levels for isolated intersections leading to safety improvements. Minnesota Department of Transportation.

Elvik, R., 1995. Meta-analysis of evaluations of public lighting as accident countermeasure. *Transportation Research Record*, 1486(1), 112–113.

Federal Highway Administration (FHWA), 2012. *FHWA Lighting Handbook*. Washington, DC.

Gross, F., 2013. Case-control analysis in highway safety: Accounting for sites with multiple crashes. *Accident Analysis & Prevention*, 61, 87–96.

Gross, F., Donnell, E.T., 2011. Case–control and cross-sectional methods for estimating crash modification factors: Comparisons from roadway lighting and lane and shoulder width safety effect studies. *Journal of Safety Research*, 42(2), 117–129.

Gross, F., Jovanis, P.P., 2007. Estimation of the safety effectiveness of lane and shoulder width: Case-control approach. *Journal of Transportation Engineering*, 133(6), 362–369.

Gross, F., Persaud, B., Lyon, C., Groos, F., Persaud, B., Lyon, C., Gross, F., Persaud, B., Lyon, C., 2010. A guide to developing quality crash modification factors. FHWA, Washington DC.

IESNA, I.E.S., 1993. *Lighting Handbook.*

Isebrands, H.N., Hallmark, S.L., Hans, Z., McDonald, T., Preston, H., Storm, R., 2006. Safety impacts of street lighting at isolated rural intersections, Part II. Minnesota Department of Transportation.

Jackett, M., Frith, W., 2013. Quantifying the impact of road lighting on road safety - A New Zealand Study. IATSS Res. 36 2 , 139–145. doi:10.1016/j.iatssr.2012.09.001.

Janoff, M.S., Koth, B., McCunney, W., Berkovitz, M.J., Freedman, M., 1978. The relationship between visibility and traffic accidents. *Journal of the Illumination Engineering Society*, 7(2) , 95–104.

Johnson, M., Fabregas, A., Wang, Z., Katkoori, S., Lin, P.-S.S.P.-S., 2014. Embedded system design of an advanced illumination measurement system for highways. *8th Annual IEEE International Systems Conference, SysCon 2014 - Proceedings*, 579–586. doi:10.1109/SysCon.2014.6819314.

Keck, M.E., 2001. A new visibility criteria for roadway lighting. *Journal of the Illumination Engineering Society*, 30(1), 84–89.

Kim, D.-G., Washington, S., 2006. The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Accident Analysis & Prevention*, 38(6), 1094–1100. doi:http://dx.doi.org/10.1016/j.aap.2006.04.017.

Mace, D.J., 1997. Safety benefits of roadway lighting using small target visibility (STV) design. FHWA, Washington, DC.

Marsh, J.L., Hutton, J.L., Binks, K., 2002. Removal of radiation dose response effects: an example of over-matching. *BMJ Br. Med. J* 325(7359), 327.

Monsere, C.M., Fischer, E.L., 2008. Safety effects of reducing freeway illumination for energy conservation. *Accident Analysis & Prevention*, 40(5), 1773–1780. doi:10.1016/j.aap.2008.06.018.

National Safety Council, 2018. Drive at night. https://www.nsc.org/road-safety/safety-topics/night-driving, accessed 8/9/18.

NHTSA, 2015. Fatality Analysis Reporting System. http://wwwfars.nhtsa.dot.gov/QueryTool/QuerySection/Report.aspx, accessed 1/1/15.

Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Computer Science*.

Preston, H., Schoenecker, T., 1999. Potential safety effects of dynamic signing at rural horizontal curves. Minnesota Local Roads Research Board.

Psaty, B.M., Koepsell, T.D., Lin, D., Weiss, N.S., Siscovick, D.S., Rosendaal, F.R., Pahor, M., Furberg, C.D., 1999. Assessment and control for confounding by indication in observational studies. *Journal of the American Geriatric Society*, 47(6), 749–754.

Sahai, H., Khurshid, A., 1995. *Statistics in Epidemiology: Methods, Techniques and Applications*. CRC Press.

Schlesselman, J.J., 1982. *Case-Control Studies: Design, Conduct, Analysis*. Oxford University Press.

Scott, P.P., 1980. The relationship between road lighting quality and accident frequency. Transport and Road Research Laboratory.

Sullivan, J.M., Flannagan, M.J., 2007. Determining the potential safety benefit of improved lighting in three pedestrian crash scenarios. *Accident Analysis & Prevention*, 39(3), 638–647.

Sullivan, J.M., Flannagan, M.J., 2002. The role of ambient light level in fatal crashes: Inferences from daylight saving time transitions. *Accident Analysis & Prevention*, 34(4), 487–498. doi:10.1016/S0001-4575(01)00046-X

Wang, Z., Li, M., Yang, R., Lin, P.-S., Xin, C., Hsu, P., Wong, E., 2019. A matched case-control safety study of street lighting uniformity along urban roadway segments. Transportation Research Board (TRB) 98th Annual Meeting, Washington DC.

Wang, Z., Lin, P.-S., Chen, Y., Hsu, P.P., Ozkul, S., Bato, M., 2017. Safety Effects of Street Illuminance on Roadway Segments in Florida. 96th Transportation Research Board Annual Meeting, Washington, DC.

Wanvik, P.O., 2009. Effects of road lighting: An analysis based on Dutch accident statistics 1987-2006. *Accident Analysis & Prevention*, 41(1), 123–128. doi:10.1016/j.aap.2008.10.003

Wei, F., Wang, Z., Lin, P.-S.P.-S., Hsu, P.P.P.P., Ozkul, S., Jackman, J., Bato, M., 2016. Safety Effects of Street Illuminance at Urban Signalized Intersections in Florida. Transportation Research Board 95th Annual Meeting, 2555 16–6376 . doi:10.3141/2555-13.

Woodward, M., 2013. *Epidemiology: Study Design and Data Analysis*, 3rd ed. CRC Press, London.

Xin, C., Wang, Z., Lin, P.-S., Koh, B., Hsu, P.P., Wong, E., Chen, T., 2018. Exploring the impacts of street illuminance on nighttime crash severity in roadway segments using a random parameter ordered probit model. Center for Urban Transportation Research, University of South Florida.

Xu, Y., Ye, Z., Wang, Y., Wang, C., Sun, C., 2018. Evaluating the influence of road lighting on traffic safety at accesses using an artificial neural network. *Traffic Injury Prevention*, in press.

Yang, R., Wang, Z., Lin, P.-S., Li, X., Chen, Y., Hsu, P.P., Henry, A., 2019a. Safety effects of street lighting on roadway segments: Development of a crash modification function. *Traffic Injury Prevention*. 0(0), 1–7. doi:10.1080/15389588.2019.1573317.

Yang, R., Wang, Z., Lin, P.-S., Li, X., Chen, Y., Hsu, P.P., Henry, A., 2019b. Safety effects of street lighting on roadway segments: Development of a crash modification function. *Traffic Injury Prevention*, 20(3), 296–302.

Yu, L., Li, Z., Bill R, A., Noyce A, D., 2015. Development of freeway and interchange safety performance functions with respect to roadway lighting: A pilot study. *Transportation Research Record*, 2485,16–25. doi:10.3141/2485-03.

Zhang, K., Yao, L., Li, G., 2012. Relationship between road lighting and traffic accidents. ICLEM 2012: Logistics for Sustained Economic Development—Technology and Management for Efficiency, ASCE, 640–646.

Zhao, J., Zhou, H., Hsu, P., 2015. Correlating the safety performance of urban arterials with lighting. *Transportation Research Record*, 2482, 126–132. doi:10.3141/2482-16.

CENTER FOR TRANSPORTATION EQUITY, DECISIONS & DOLLARS

The **Center for Transportation, Equity, Decisions and Dollars (CTEDD)** is a USDOT University Transportation Center, leading **transportation policy research** that aids in **decision making** and **improves economic development** through **more efficient, and cost-effective use of existing transportation systems,** and offers **better access to jobs and opportunities.** We are leading a larger consortium of universities focused on **providing outreach** and **research to policy makers**, through **innovative methods** and **educating future leaders of the transportation field.**

UNIVERSITY OF TEXAS ARLINGTON

CAL POLY SAN LUIS OBISPO

Georgia Tech

USF UNIVERSITY OF SOUTH FLORIDA

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

C-TEDD@UTA.EDU | CTEDD.UTA.EDU | @UTACTEDD | @C__TEDD | @ CENTER FOR TRANSPORTATION, EQUITY, DECISIONS AND DOLLARS | 817.272.5138