Deliverable 8

Application of Dynamic Crash Prediction Methodologies to FDOT Safety and Transportation System Management & Operation (TSM&O) Programs

Project No.

FDOT BE548

Final Report

Prepared For

Florida Department of Transportation

November 2020

CUTR CENTER FOR URBAN TRANSPORTATION RESEARCH
USF UNIVERSITY OF SOUTH FLORIDA
CUTR.USF.EDU

Application of Dynamic Crash Prediction Methodologies to FDOT Safety and Transportation System Management & Operation (TSM&O) Programs

Final Report

Project Number
FDOT BE548

Prepared by

Zhenyu Wang, Ph.D.
Pei-Sung Lin, Ph.D.
Runan Yang
Abhijit Vasili

Mohammed Hadi, Ph.D.
Tao Wang, Ph.D.

Catalina Echeverri

Prepared For
Florida Department of Transportation
Nicole Forest, Project Manager

November 2020

Disclaimer

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

Metric Conversion Chart

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: volumes greater than 1000 L shall be shown in m^3				
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
TEMPERATURE (exact degrees)				
ºF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	ºC

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
BE548		

4. Title and Subtitle Application of Dynamic Crash Prediction Methodologies to FDOT Safety and Transportation System Management & Operation (TSM&O) Programs	5. Report Date November 2020 6. Performing Organization Code
7. Author(s) (USF) Zhenyu Wang, Pei-Sung Lin, Runan Yang, Abhijit Vasili (USF) (FIU) Mohammed Hadi, Tao Wang (FIU) Catalina Echeverri (Gannett Fleming)	8. Performing Organization Report No.
9. Performing Organization Name and Address Center for Urban Transportation Research (CUTR) University of South Florida (USF) 4202 E. Fowler Avenue, CUT100 Tampa FL 33620	10. Work Unit No. (TRAIS) 11. Contract or Grant No. BE548
12. Sponsoring Agency Name and Address Florida Department of Transportation (FDOT) 605 Suwannee St., MS 30 Tallahassee, FL 32399-0450	13. Type of Report and Period Covered Final Report, 05/2018–11/2020 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract
Dynamic crash prediction, a proactive safety management strategy, predicts crash risk based on prevailing traffic conditions and prevents crashes before occurrence. As an innovative technology, dynamic crash prediction provides a potential way for the Florida Department of Transportation (FDOT) to take advantage of information provided by intelligent transportation system (ITS) devices and other sources, combined with increasingly available big data/data analytics to effectively the safety and mobility of Florida roadway systems. This project documents and evaluates existing dynamic crash prediction methods and practices related to accuracy and timeliness and develops recommendations for implementing a proactive safety strategy in Florida. Information related to dynamic crash prediction, including previous studies, existing vendor and technologies, and current user and implementation experiences, was collected and assembled through literature review, online search, document review, and interviews. A comparison of existing dynamic crash prediction technologies was performed to identify the “best” technology (WayCare) for the pilot study. Historical traffic and crash data (2015-2019) were collected at two study sites, covering freeway segments (I-95) and an arterial corridor (E Sunrise Blvd) in FDOT District 4. Based on the historical data, WayCare built a prediction model that can predict crash risk for a three-hour time window based on traffic and crash conditions for nine hours prior to the time window. With the WayCare model, the research team conducted an offline test, using three-month data in 2020 (January, February, July), to evaluate the performance of dynamic crash prediction in the Florida roadway environment. The offline results showed that the WayCare model presented better performance for the I-95 site than the E Sunrise Blvd site due to the high number of crashes and relative simplicity of traffic patterns on freeways. The model can correctly predict 60% crashes during the PM period (3:00–6:00 PM) on I-95. Based on pilot study results, it is suggested to implement the dynamic prediction model preferentially on freeways but work with the WayCare team to improve performance of its model for periods other than the PM period. Three crash prediction actions (DMS safety messages, stationary police cars with flashing lights, and advance warning to Road Rangers) were proposed based on WayCare’s experience and the availability of TSM&O applications in FDOT District 4. A further study is needed to address the safety and mobility of crash prediction actions and real-time data connection.

17. Key Word	18. Distribution Statement
Dynamic crash prediction, proactive safety management, freeway, arterials, big data, machine learning, crash prevention	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		113	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Acknowledgments

The authors express their sincere appreciation to Project Manager (PM) Nicole Forest and three previous PMs (Melissa Ackert, Allison Glunt, and Daniel A. Smith) for their support and excellent guidance throughout the project. Special thanks go to Shayla Khalilahmadi of FDOT District 4, who helped collect traffic sensor data, and Vik Bhide (City of Tampa), Severyn Josette (City of For Lauderdale), Norman Jester (Pinellas County), and Deneta Jones (Pinellas County) for providing valuable information in interviews.

The authors also thank Raj Ponnaluri, Alan El-Urfali, Fred Heery, Lora Hollingsworth, Jonathan Overton, Joseph Santos, Rupert Giroux, Benjamin Jacobs, Yujing Xie, Alexandra Lopez, Hossam Abdel All AbouKorin, Mairelys Hill, and Arlene Willis for their worthy comments in coalition meetings and Darryll Dockstader, Jennifer Clark, Jason Tuck, and Amanda Ulmer of the FDOT Research Center for their project management.

The authors are also highly indebted to WayCare, especially Paul-Matthew Zamsky, Alex Smolyak, and Shai Suzan, whose efforts in model development and techncial support during the pilot study were invaluable. The authors also recognize the contributions of Dr. Cong Chen of the University of South Florida (USF) and Dr. Aluri Priyanka of Florida International University (FIU) in proposal preparation. The project would not have been possible without the kind support and help of many individuals and departments at USF, FIU, and Gannett Fleming, Inc. We extend our sincere thanks to all of them.

Executive Summary

The Florida Department of Transportation (FDOT) Safety and Transportation System Management and Operations (TSM&O) programs have been collecting, archiving, and analyzing a wide range of traffic, crash, event, and other data to improve congestion and safety on the State Highway System (SHS). Dynamic crash prediction, a proactive safety management strategy, predicts crash risk based on prevailing traffic conditions and applies crash prevention actions to prevent crashes before occurrence. As an innovative technology, dynamic crash prediction provides a potential way for FDOT to take advantage of information provided by intelligent transportation system (ITS) devices and other sources, combined with increasingly available big data/data analytics to effectively prevent crash occurrence and improve the safety and mobility of Florida roadways. Although many Florida traffic agencies have shown interest in dynamic crash prediction methods and have plans to implement them, there is no clear understanding on the applicability of dynamic crash prediction in performance, implementability, integrability, and impacts.

Realizing the challenges of using available big data to improve roadway safety and the potential benefits of proactive safety management strategies, this project aimed to (1) document the current state of practice of dynamic crash prediction methods and software based on a comprehensive review of the literature, practices, and tools; (2) compare existing dynamic crash prediction methods/software based on developed evaluation criteria and select methods/software for potential use in Florida; (3) coordinate with FDOT District 4 and local agencies to conduct a pilot study to demonstrate and evaluate selected dynamic crash prediction methods/software; (4) conduct a pilot study to evaluate identified dynamic crash prediction methods/tools at selected sites (covering both freeways and arterials) in FDOT District 4; and (5) develop recommendations for implementing dynamic crash prediction in Florida.

A comprehensive literature review was conducted to summarize state-of- the-art dynamic crash predictions, including a theoretical framework for dynamic crash prediction, data needs and sampling methods, modeling algorithms, and performance. Meanwhile, online searching, document review, and interviews were used to collect the information on the state-of-the practice of dynamic crash prediction. The study identifies existing vendors and technologies and understanding of implementation status and evaluates/compares identified technologies. Based on the results, one technology (WayCare) was selected to conduct the pilot study. Knowledge about dynamic crash prediction implementation was used to develop the pilot study plan.

Two study sites, covering freeway segments (I-95) and arterials (E Sunrise Blvd), were selected considering historical crash records, traffic demands, traffic sensor availability, and Transportation System Management & Operations (TSM&O) applications. Historical traffic data and crash data for five years (2015–2019) were collected and provided to waycare for model calibration. Using the calibrated model, the research team conducted an offline test using three-month data (January, February, July) in 2020. Two performance measures were used in the evaluation—Recall, the percentage of crash events that can be predicted by the model; higher is better, and Precision, the percentage of alarms (crash prediction) that are true; higher is better. False Alarm Rate (FAR) is the flip side of Precision (=1-Precision), which is defined as the percentage of alarms that are false; lower is better.

Based on the offline test, the following findings were obtained:

- The WayCare model presented better performance for the I-95 site than the E Sunrise Blvd site for Recall (25% vs. 11%) and FAR (83% vs. 93%). The high number of crashes and relative simplicity of traffic patterns on the freeway may explain why the WayCare model worked better on I-95.
- The WayCare model presented varying Recall performances by period for both I-95 and E Sunrise Blvd.
    - On the I-95 sites, the WayCare model presented "good" performance for the PM period (3:00 PM–6:00 PM), with 55–65% of crashes predicted for different months. These performances were close to WayCare's evaluation based on historical data for 2015–2019 (54% of crashes can be predicted for I-95, on average, without distinguishing periods), as shown in Appendix A.
    - The WayCare model had "poor" performance on I-95 for the Midday and Night periods. The model outputs could not predict any crashes in most scenarios for these periods except for the Midday period in July 2017 (6% of crashes can be predicted).
    - It is worth noting that the Recall performance for the PM periods in July 2020 (55% of crashes can be predicted) was lower than those for January and February 2020 (64% of crashes can be predicted) and July 2017 (61%). This comparison may imply that the COVID-19 pandemic event had an impact on model performance (Recall reduction of 6–9%) on I-95.
    - For E Sunrise Blvd, the WayCare model presented relatively "better" performance for Midday (12:00 PM–3:00 PM) and PM (3:00 PM–6:00 PM). Based on 2020 data, an average 20% of Midday crashes and 11% of PM crashes could be predicted. It was interesting to find that the model had better performance in July than in January or February 2020, which is the opposite of the finding for I-95.
- FAR were relatively high ($\geqslant$ 70%) across scenarios (Precision was relatively low, $\leq$ 30%). This implies that 70% (or higher) of alarms were not actually associated with a crash. The possible causes are:
    - Underreported crashes – some minor crashes tend to not be reported to police and thus are not included in the crash database but can be predicted by the WayCare model.

- Near-crash events – some near-crash events, such as serious conflicts, are high-risk events but do not necessarily result in crashes. Prediction of these near-crash events are useful to apply actions to prevent risky situations.

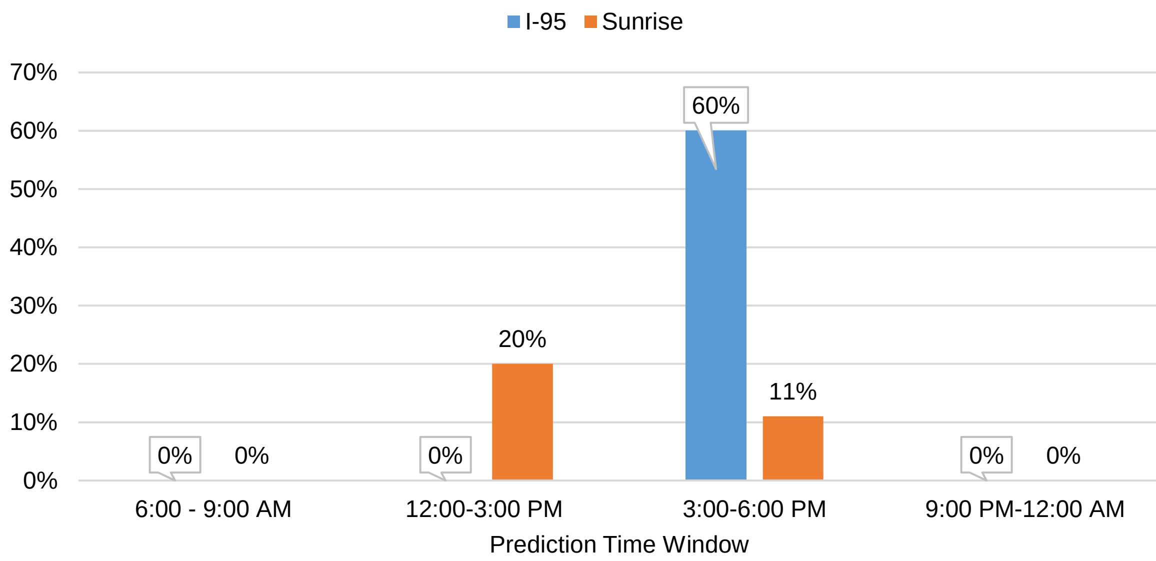


Figure ES-1. Percentage of crashes that could be predicted (Recall)

The recommendations for implementation developed based on the pilot study are as follows:

- Implement the dynamic prediction model preferentially on freeways but work with WayCare to improve model performance for periods other than PM considering the following:
    - Model produced good performance for the PM period (3:00–6:00 PM) on the tested freeway section (correctly predicted 60% of crash cases).
    - Local resources for model data input and crash prevention (i.e., traffic sensors, ITS/T&SMO actions, etc.) are plentiful on interstates.
    - Freeways experience high traffic volumes and excessive crash frequencies compared to other road facilities; implementation of dynamic crash prediction could bring significant safety and mobility benefits.
- Consider implementation of the dynamic prediction model on arterials but work with WayCare to improve model performance for periods other than Midday and PM, if traffic agencies have a high need for arterial safety management, considering the following:
    - Model showed “positive” performance for the two periods on arterials (correctly predicting 11–20% of crash cases).

- Relatively high volumes and crash frequencies on major arterials introduce the need for dynamic crash prediction and prevention; arterials have more complex traffic patterns.
- Traffic agencies should decide on implementation based on their arterial safety management goals and needs.

- Real-time implementation of the model at TMCs will require maintaining traffic and crash/incident data for the previous nine hours to predict crash rates for the next three-hour prediction window. The time interval of traffic sensor data is suggested to be 20 sec or 1 min. Longer time intervals can be applied; however, they may reduce prediction performance. The protocol for data transfer between TMC SunGuide software and databases and the WayCare web platform needs to be addressed.
- Three crash prediction actions (DMS safety messages, stationary police cars with flashing lights, advance warning to Road Rangers) were proposed based on WayCare's experience and the availability of TSM&O applications in FDOT District 4. A further study is needed to address the safety and mobility of the crash prediction actions.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

According to the 2060 Florida Transportation Plan, Florida's transportation system aims to evolve over the next 50 years to support the transformation of Florida's economy and communities and proposes a vision of a fatality-free and congestion-free transportation system in Florida. To support this goal, the Florida Department of Transportation (FDOT) has developed various strategy plans such as the Florida Strategic Highway Safety Plan (SHSP) and the Transportation Systems Management & Operations (TSM&O) strategy plan. These plans intend to integrate programs to optimize the performance of multimodal infrastructures through implementation of systems, services, and projects to preserve capacity and improve the security, safety, and reliability of Florida's transportation system. The TSM&O programs and supporting ITS strategies are collecting, archiving, and analyzing big-scale data regarding traffic, weather, crashes, construction and other events, signals, and videos to support traffic management strategies using a proactive approach.

Big challenges remain in achieving the goal of a fatality-free transportation system in Florida As shown in Figure 1, Florida experienced a rapid increase in traffic crashes for 2011–2018 (from 227,998 in 2001 to 403,626 in 2018) and a high yearly fatalities (> 3,000 in 2016–2018).

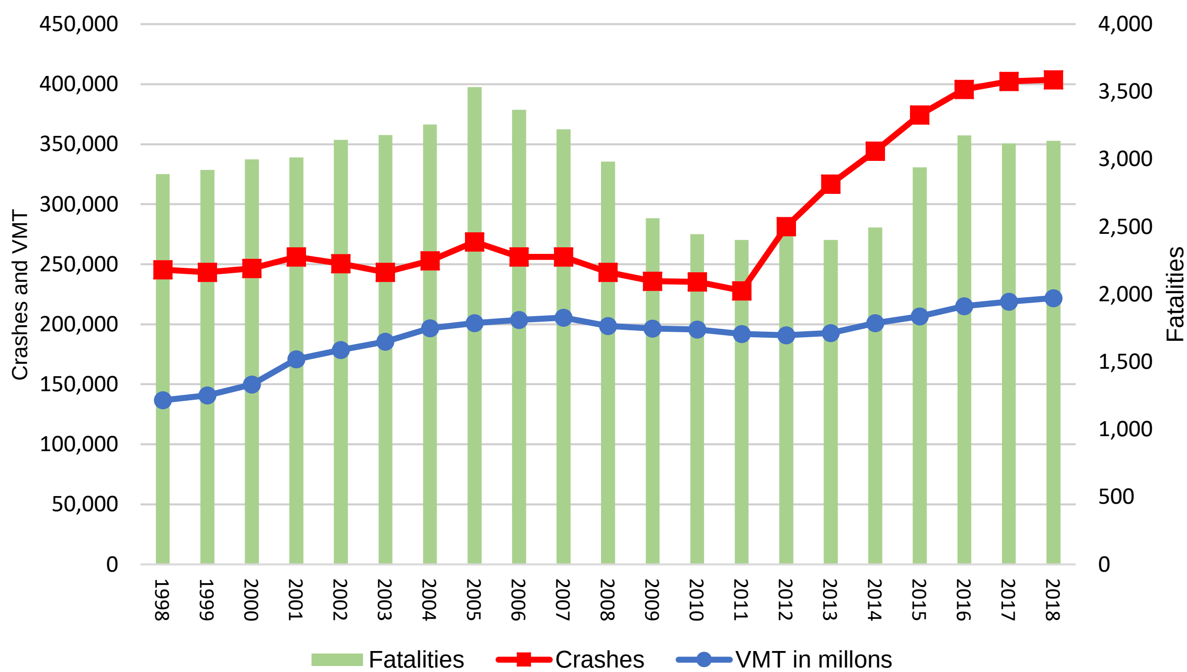


Figure 1. Trends of Crashes, VMT, and Fatalities in Florida, 1998–2018

(Source: FLHSMV Traffic Crash Facts 2018)

Principle roads in urban areas (including interstates, expressways, and major arterials) accounted for a major portion (41%, as shown in Figure 2) of traffic crashes, although these roads comprise only around 15% of center miles on the Florida roadway system (*1*). Thus, developing effective safety management strategies is an urgent task for FDOT and local agencies to reduce crashes and prevent fatalities on the Florida roadway system, especially for urban principal roads that carry high traffic volumes and suffer high crash risks.

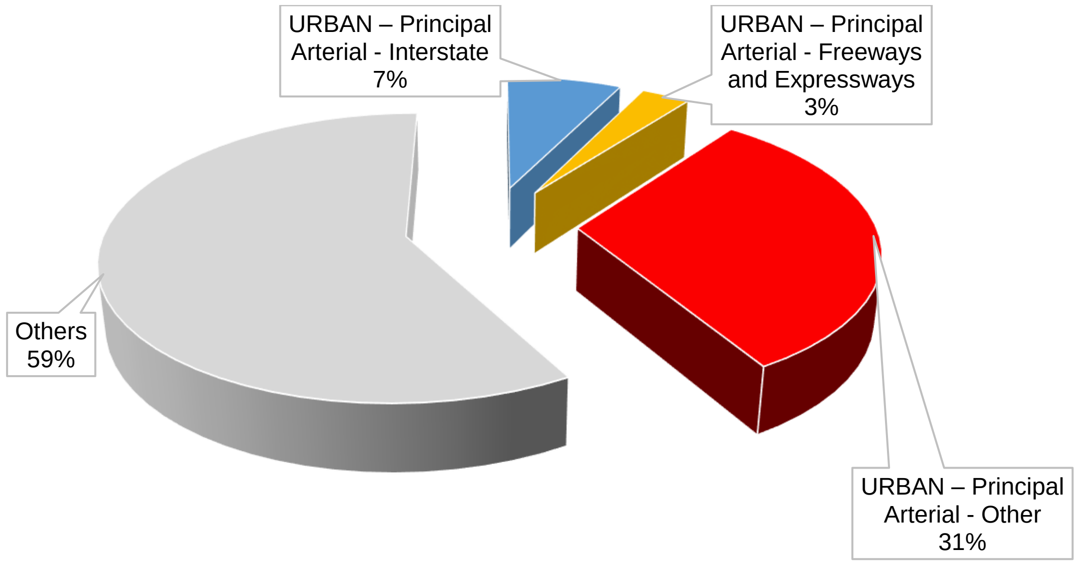


Figure 2. Distribution of Traffic Crashes by Roadway Type in Florida

(Source: FDOT Crash Analysis Reporting [CAR] System)

1.2 Proactive Safety Management Strategy

The mission of FDOT is to provide a safe transportation system that ensures the mobility of people and goods, enhances economic prosperity, and preserves the quality of environment and communities. FDOT's TSM&O programs have been collecting, archiving, and analyzing a wide range of traffic, crash, event, and other data to improve congestion and safety on the State Highway System (SHS). To use these big data to improve transportation safety, two strategies may be applied, per the Federal Highway Administration (FHWA) Road Safety Audit Guidelines:

- *Reactive Approach* – Based on analysis of existing crash data, focuses on identification of locations experiencing safety problems (screening), problem definition (diagnosis), and identification and implementation of countermeasures (cure).
- *Proactive Approach* – Aims to prevent safety problems before they manifest themselves in the form of a pattern of crash occurrence.

The *Highway Safety Manual* (HSM) includes procedures to support the reactive approach, and FDOT has associated data collection and archiving and has established processes in place for this purpose, as shown in Figure 3.

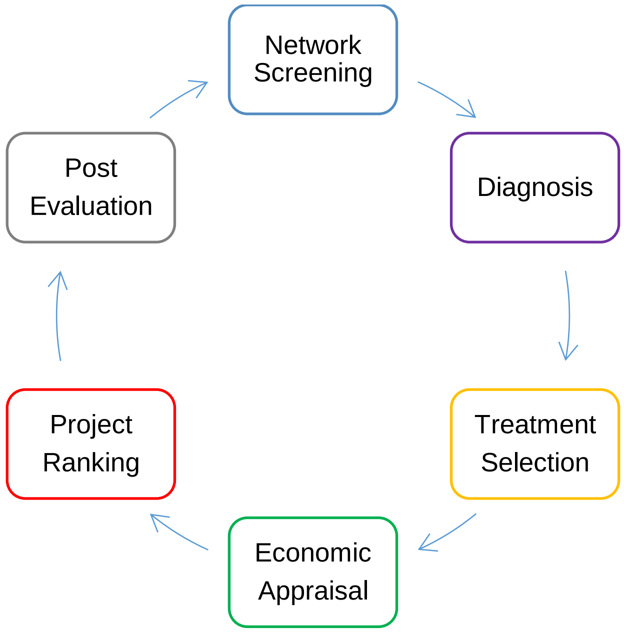


Figure 3. Reactive Safety Management Strategy

However, the reactive safety management strategy has some limitations:

- The Reactive safety strategy relies on historical crash data that are often inaccurate, incomplete, and outdated.
- The Reactive safety strategy is costly, as a long observation period (≥ three years) is needed to accumulate sufficient samples of historical crash data.
- The Reactive strategy does not fully use big traffic and other high-resolution real-time data; it is difficult to integrate with ITS/TMS&O applications to identify crash risk in real-time and prevent crashes before occurrence.

Compared to the Reactive approach, the Proactive safety management strategy, as shown in

Figure 4, provides an innovative way to reduce potential crash risk prior to crash occurrence. This approach has the following advantages:

- *Crash prevention* – Can prevent crash risks before crash occurrence and save life and property loss.
- *Relatively low cost* – Does not rely on historical crash data; its implementation is quicker and less costly.

- *Integration* – Is more effective in supporting the operations of FDOT ITS/TSM&O programs. By fully using big data, it provides decision-making for FDOT TSM&O actions to improve safety proactively.

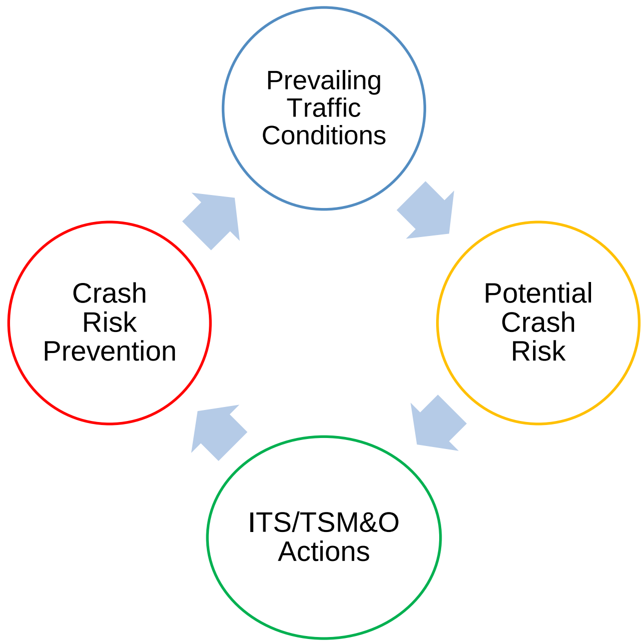


Figure 4. Proactive Safety Management Strategy

FDOT potentially could be more proactive through the use of dynamic crash prediction methodologies, an innovative safety management strategy, to take advantage of information provided by ITS devices and other sources, combined with increasingly available big data/data analytics to predict crash statistics, such as location, time, and severity in real-time prior to crash occurrence. Agencies using these methods or software can then proactively reduce the potential for crashes and enhance traffic flow by implementing strategies before crashes occur and can provide quick and effective responses if/when they do.

Although many Florida traffic agencies have shown interest in dynamic crash prediction methods and have plans to implement them, there is no clear understanding of the applicability of these existing methods in Florida in the following aspects:

- *Performance* – FDOT has limited knowledge of the accuracy and timeliness of existing methods. What method can provide the best, or at least acceptable, outcomes in quick and accurate real-time crash prediction?
- *Implementability* – FDOT has limited knowledge on how to easily implement the existing crash prediction methods and the required capabilities and resources, including data needs, software support, operation and management requirements, reliability and robustness, and output content and format.

- *Integrability* – FDOT has limited knowledge on integration of the existing prediction methods/software in FDOT TMCs and TSM&O/Safety programs, including input data interface, output format, hosting environment, compatibility to existing system and related standard operating guidelines (SOGs)/action plans, and needed resources from involved local agencies.
- *Impacts* – FDOT has limited experience on the impacts of dynamic crash prediction on safety and mobility on the Florida transportation system. What management strategies and data-sharing and dissemination should be applied after crash prediction? What is the effectiveness of the strategies in crash prevention and improvement of mobility?

This lack of knowledge prevents FDOT from implementing dynamic crash prediction to improve Florida highway safety and mobility. Thus, research is necessary to address the above aspects and evaluate the accuracy and timeliness of existing dynamic crash prediction methodologies, their applications at TMCs, and their impacts on safety and mobility.

1.3 Research Objectives

The primary goal of this research project is to evaluate existing dynamic crash prediction methods and practices related to accuracy and timeliness, use in TMCs, and impacts on safety and mobility for implementing a proactive safety strategy in Florida. To achieve this goal, the objectives of this research project are the following:

- Document the current state of practice of dynamic crash prediction methods and software based on a comprehensive review of the literature, practices, and tools.
- Provide an understanding of application uses, effectiveness, integration, and operations and management (O&M) requirements.
- Develop criteria for evaluation of existing dynamic crash prediction methods/software for potential use by FDOT TSM&O and Safety programs.
- Compare existing dynamic crash prediction methods/software based on developed evaluation criteria and select methods/software for potential use in Florida.
- Coordinate with FDOT District 4 and local agencies to conduct a pilot study to demonstrate and evaluate selected dynamic crash prediction methods/software.
- Estimate the safety and mobility benefits from the implementation of the developed dynamic crash prediction.
- Develop warrants, framework, and SOGs or an action plan that determine practical methods, needed resources, and operations/management procedures to provide guidelines on implementing dynamic crash prediction in Florida.

1.4 Report Organization

The report is organized as follows: Chapter 1 introduces the project background and research objectives, and Chapter 2 presents a comprehensive review of previous studies related to dynamic crash prediction, including theory framework, data needs, sampling methods, modeling technologies, and performance. Identification and comparison of existing dynamic crash prediction technologies are provided in Chapter 3, and Chapter 4 describes the pilot study conducted in FDOT District 4, including site selection, data collection, offline testing procedure, testing results, and suggested prevention actions. Finally, Chapter 5 presents conclusions and recommendations for implementing dynamic crash prediction in Florida.

2 Literature Review

2.1 Theoretical Fundamentals for Dynamic Crash Prediction

The principle of dynamic crash prediction assumes that crash occurrence is correlated to prevailing traffic conditions at a roadway facility. By investigating the traffic patterns prior to a crash, it could predict the risk (probability) of crash occurrence. Previous studies have proven the relationship between crash risk and macroscopic traffic flow characteristics such as volume, speed, and density.

2.1.1 Crash-Flow Relationship

It has been suggested in latest studies that traffic volume has a nonlinearly monotonic connection with crash count (*3*): with an increase in traffic volume, either at the aggregated (e.g., Annual Average Daily Traffic [AADT]) or disaggregated (e.g., hourly rate) levels, the likelihood of crashes tends to increase. High volumes signify frequent interactions among vehicles, resulting in increased vehicle conflicts and risk of crashes. It is worth mentioning that the increase rate of all crashes progressively diminishes when traffic volume increases; however, the increase rate of multi-vehicle (MV) crashes keeps nearly constant as traffic volume increases. Figure 5 presents the crash-flow relationship on urban road segments.

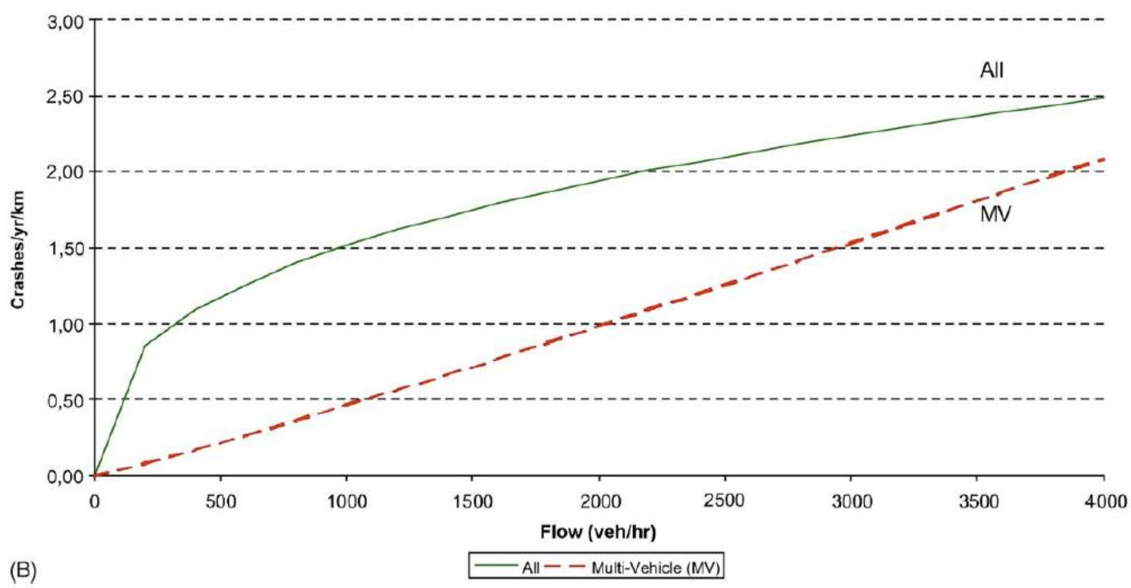


Figure 5. Crash-Flow Relationship for Urban Segments *(3)*

2.1.2 Crash-Speed Relationship

Several previous studies (*4–7*) found that higher mean speed is associated with an increased crash frequency. Examples of crash-flow relationships are given in Figure 6. However, the *Highway Safety Manual* (HSM) (*8*) argues that the relationship of crash-speed presents a

U-shaped curve: the crash rate reaches the lowest point at 60 mph and increases when speed is higher than 60 mph. Crashes related to low speed may be caused by low-speed-related maneuvers (e.g., turning movements), roadway conditions, and congestion.

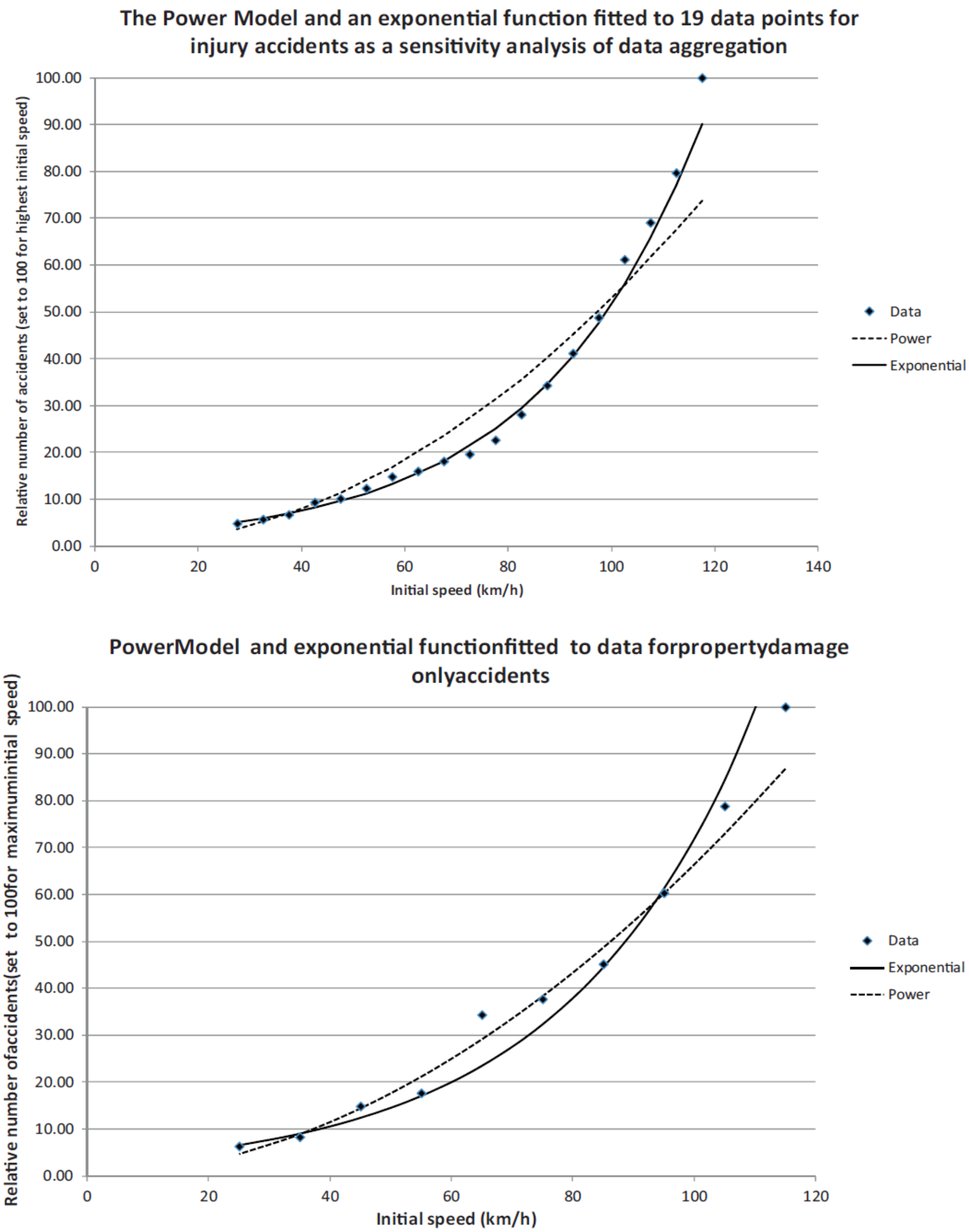


Figure 6. Crash-Speed Relationship *(5)*

Previous studies (*4, 8–11*) consistently found that large speed variation affects increased crash frequency. An example of the crash-speed variation relationship is given in Figure 7.

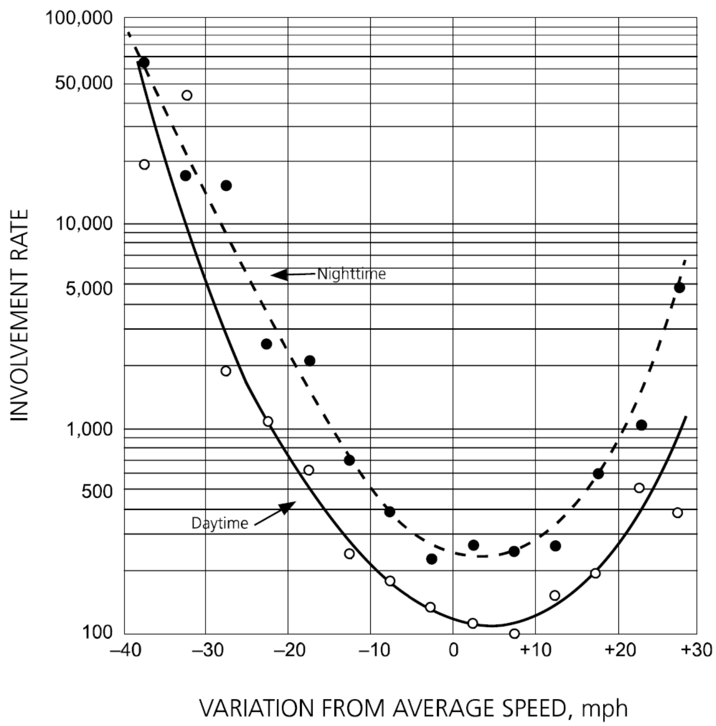


Figure 7. Crash-Speed Variation Relationship *(8)*

2.1.3 Crash-Density Relationship

A previous study (*3*) explored the relationship between crash frequency and traffic density and found that, as shown in Figure 8, an increased density results in the likelihood of single-vehicle (SV) increasing, peaking, and decreasing and increases the probability of MV crashes.

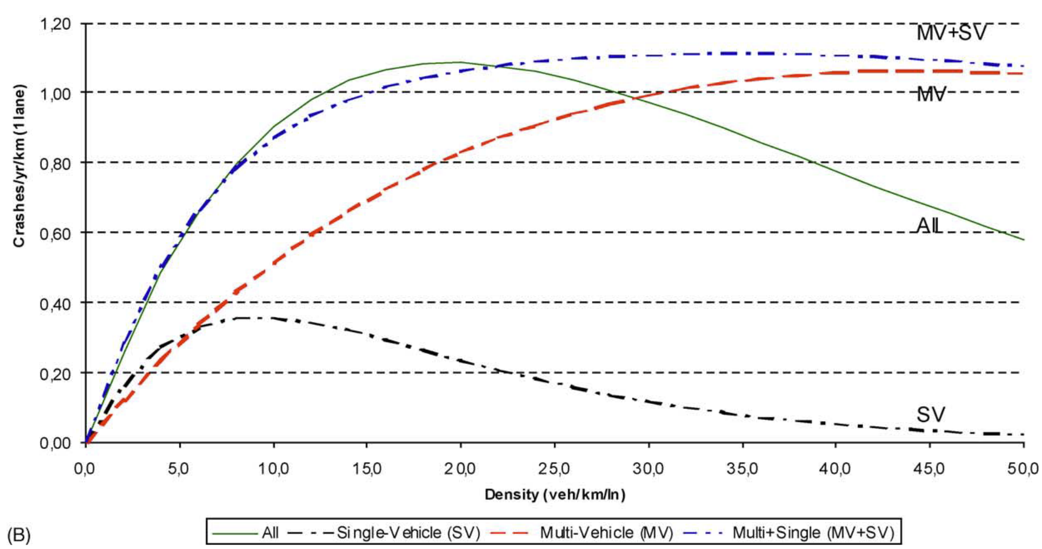


Figure 8. Crash-Density Relationship for Urban Segments *(3)*

2.1.4 Other Contributing Factors

Except for macroscopic traffic flow characteristics, crash occurrence is influenced by various factors such as human factors, vehicle characteristics, and roadway/environment features. Some of these factors are observed (e.g., geometric design, events, weather conditions), and some are difficult to collect (e.g., driving behaviors near crashes). The observed factors could be included in prediction models or using different models to address the impacts of these factors.

2.2 Fundamental Dynamic Crash Prediction Development

The occurrence of a traffic crash is a complex process and is caused by numerous factors, including behavior, vehicle, traffic, roadway geometry, and environment. In all likelihood, human error is the most significant factor contributing to traffic crashes and is estimated to account for around 93% of all crashes (*12*). In practice, behavioral and vehicle factors are often omitted because collection of information on the two factors in real time is difficult. The current practices of dynamic crash prediction attempt to predict crash risk based on real-time traffic conditions for different geometry and environmental conditions with the following assumption (*13*):

> A significant relationship exists between crash (occurrence) risk and traffic conditions prior to a crash. Traffic conditions during a certain time interval immediately before a crash, as a direct contributor, can be measured and linked to crash likelihood, given roadway and environmental conditions.

Base on this assumption, numerous dynamic crash prediction methods with various technologies have been developed since 2002. The fundamental components in the development of dynamic crash prediction are shown in Figure 9.

2.3 Facility Type

In total, 36 previous studies indicate the roadway facility types for which their prediction models were developed, as summarized in Table 1. Most previous studies (92%) focused on freeways, including freeway basic segments, merge/diverge segments, and ramps. Only 8% of previous studies (3 papers) investigated dynamic crash prediction on urban arterials. This phenomenon is caused by the following factors:

- Uninterrupted traffic flow on freeways regulated by vehicle-vehicle and vehicle-roadway interactions has simpler characteristics than the surface roads regulated by traffic signals and conflicts of side traffic. The relative simplicity of traffic operations makes dynamic crash prediction easier.
- Most important, traffic surveillance systems (e.g., loop detectors) are widely implemented on freeway facilities (for example, interstates), and traffic data resolutions (spatial and temporal) on freeways are higher than those on arterials. Data availability and integrity resulted in most previous studies focusing on freeways.

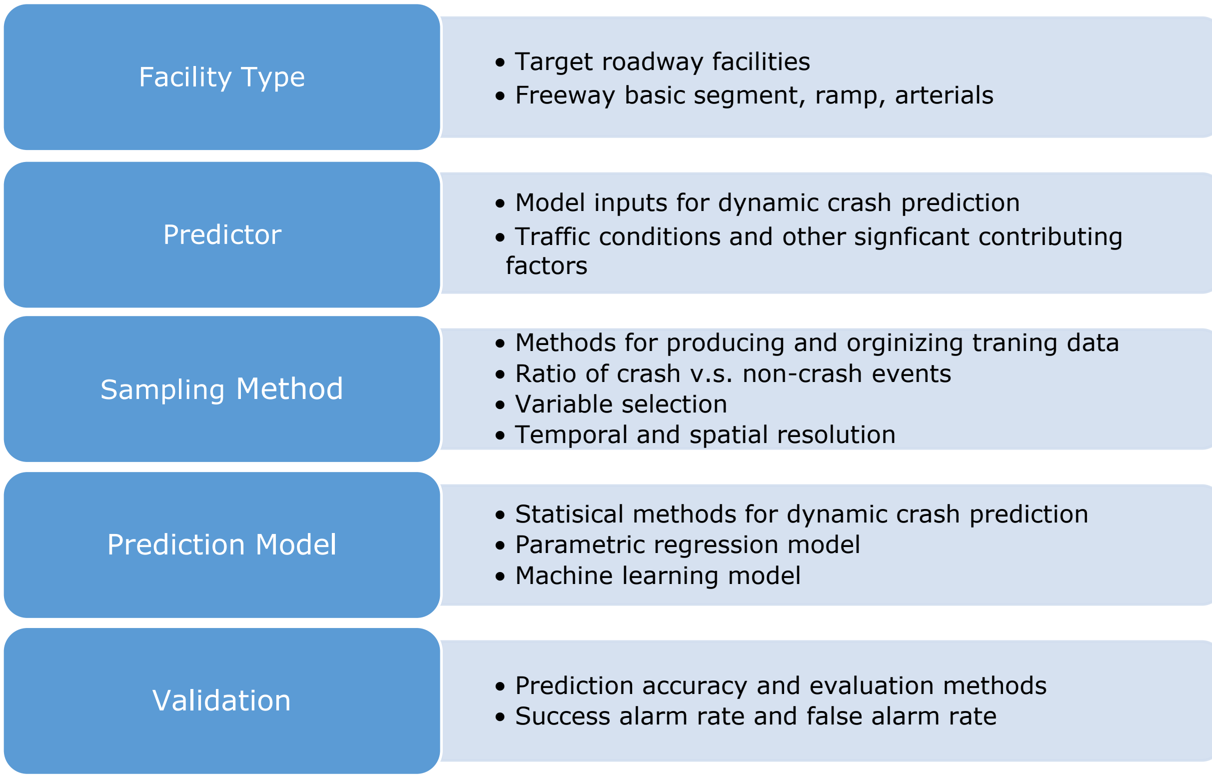


Figure 9. Fundamental Components of Dynamic Crash Prediction Development

Table 1. Summary of Roadway Facility Types in Dynamic Crash Prediction

Facility Type	Number of Previous Studies	Percentage
Freeway basic segment	16	44%
Freeway segment (basic + merge & diverge)	3	8%
Freeway interchange (mainline and ramp)	14	39%
Arterial	3	8%
Total	36	

2.4 Predictors

Predictors, as the data fields (variables) for model inputs, are usually significant contributors to crash occurrence. Based on the literature review, significant factors contributing to crash risk include traffic conditions prior to crash occurrence, geometry, time, and environment conditions. In most studies (94% of 36 papers), only traffic condition variables were treated as predictors in prediction models; different models were developed to address the variation of other factors. Only two studies (*14*, *15*) used a location variable (ramp or not) as predictors in addition to traffic conditions. Traffic condition predictors are summarized in Table 2.

A meta-analysis (*13*) showed that speed variation (including standard deviation and coefficient of variance) highly affects the likelihood of crash occurrence. Average/median speed, average density, and traffic volume have moderate impacts on crash occurrence, and the impact of traffic volume is considerably small. The variation of density and volume was beyond the scope of the meta-analysis.

Table 2. Summary of Traffic Condition Predictors

Variable		Number of Studies	Significance Level*
Speed	Mean	20	Moderate
	Median	1	Moderate
	Standard deviation	14	High
	Coefficient of variance	18	High
Density	Mean	26	Moderate
	Variance	9	-
Volume	Mean	17	Low
	Standard deviation	11	-
	Variance	8	-

*Significance level obtained from meta-analysis in a review paper (13).

Table 3 summarizes other factors identified in previous studies. Most were used to split models (developing different models to account for different factor values) and/or be matched () to eliminate their confounding influence. These factors include roadway geometry, environment, and time of day.

Table 3. Summary of Other Factors

Variable		Number of Studies	Usage
Roadway	Ramp	3	Predictor
	Curve	6	Model split or match
	Pavement condition (dry or wet)	4	Model split or match
Environment	Peak hour	6	Model split or match
	Lighting	4	Model split or match
	Weather	5	Model split or match

2.5 Sampling Method

Dynamic crash prediction is a data-driven method. Sampling quality, which means how to select and assembly data for training and prediction, is critical to dynamic crash prediction development and implementation. In crash data sampling, the following should be considered:

- *Data Balance* – Traffic crashes are rare and random events. Previous studies collected data on historical crashes for several years, but the number of crash events is still limited (up to hundreds). On the other hand, non-crash events have massive data. Without control of crash-to-non-crash ratios in sample data (data balance), the prediction model may produce biased outputs (predominant zero-crashes).

- *Confounder Control* – Traffic crash occurrence is caused by various factors. To investigate the relative crash risk due to a change of traffic conditions (predictors), it is necessary to fix other factors (confounders), such as weather and roadway conditions.
- *Temporal Slice* – Current traffic surveillance systems (e.g., loop detectors) can collect traffic data in very short time intervals (≤ 1s). Raw data were often aggregated into longer periods (e.g., 5 or 10 mins prior to crash) to suppress noise (*13*). In the development and implementation of dynamic crash prediction models, it is necessary to determine appropriate time slices of traffic condition data to capture the most significant impacts of traffic conditions on crash occurrence.
- *Spatial Range* – Traffic conditions associated with a traffic crash usually are collected from detectors near the crash location. The prediction method needs to determine the spatial range of traffic data that significantly influence crash risk, such as detector location (upstream and/or downstream) and number of detectors. The spatial range is determined by configurations of traffic surveillance systems.
- *Sampling Rate* – Traffic sensors collect traffic data at a given time internal. The shorter the time internal is, the better the model addresses data variance. However, not all data sources support high-resolution data.

Figure 10 shows the distribution of sampling methods in previous studies for data balance and confounder controls. Due to its simplicity, cost-effectiveness, and theoretical soundness (*13*), the matched case-control method was predominantly used in previous studies (75%). In one paper, bootstrap sampling technology was used to increase sample size. Eight studies did not control the crash-to-non-crash event ratio and adopted an unbalanced sample for model development, and Figure 11 shows the distribution of non-crash-to-crash ratios in sample data for model training in previous studies.

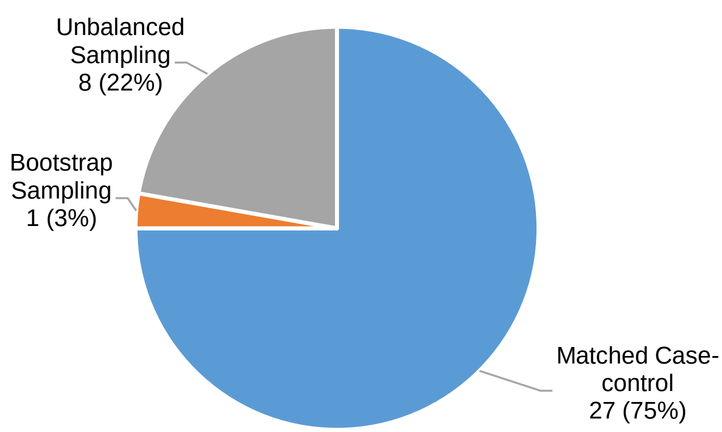


Figure 10. Summary of Sampling Methods

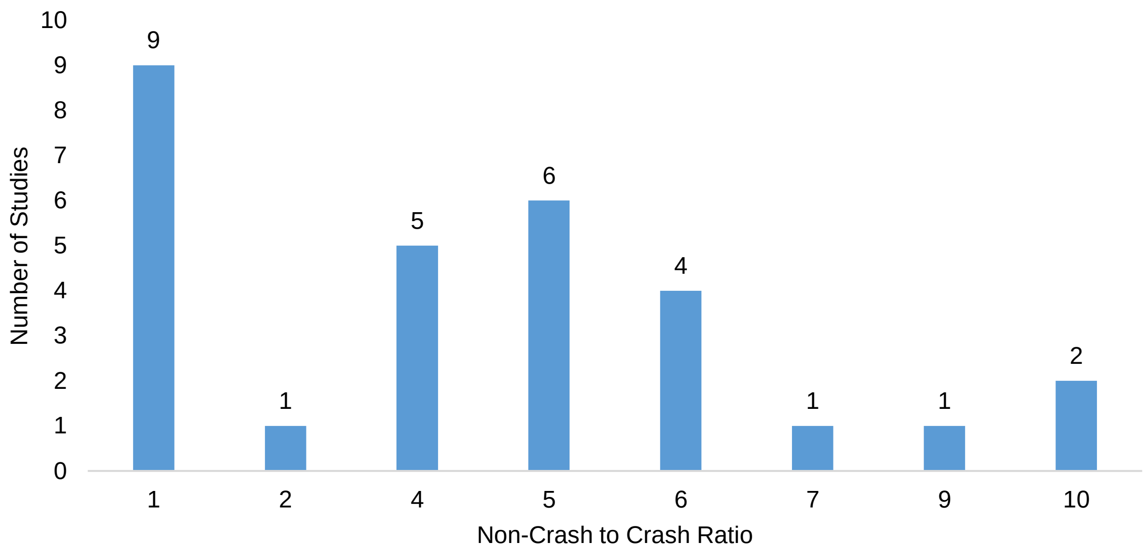


Figure 11. Summary of Non-Crash-to-Crash Ratio in Sample Data

Figure 12 shows the time slices used in previous studies. Almost 66% of previous studies found that traffic conditions within the time slice of 5–10 mins prior to a crash had the most significant impact on crash occurrence than other slices, including 0–5 mins (17%), 10–15 mins (11%), 0–10 mins (3%), and 15–20 mins (3%).

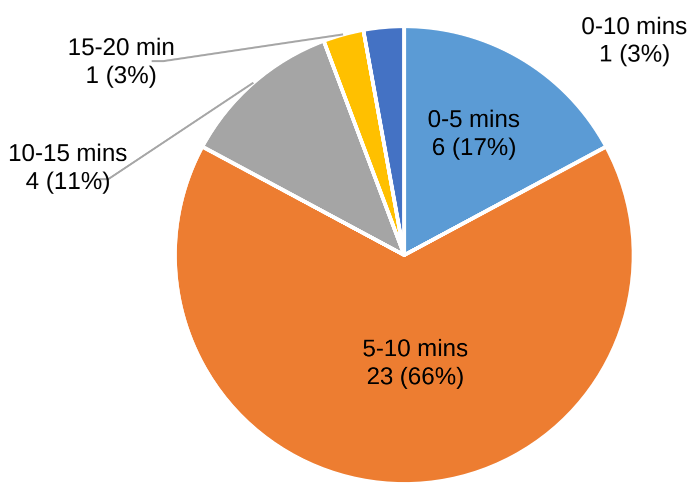


Figure 12. Summary of Time Slice in Dynamic Crash Prediction

The distribution of spatial intervals of traffic conditions in previous studies are presented in Figure 13. Most previous studies (97%) collected traffic condition information upstream of a crash, and 68% collected traffic conditions information downstream; only one study considered download stream only.

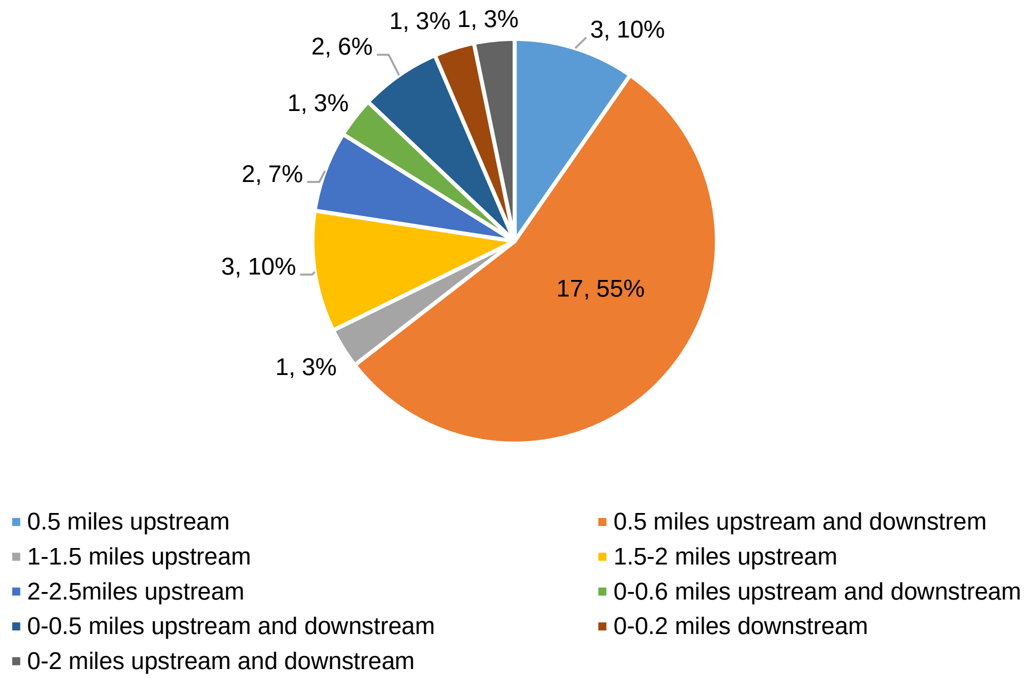


Figure 13. Summary of Spatial Range

Figure 14 presents the percentage of sampling rates from raw data sources in previous studies. In these studies, 10-, 20-, and 30-sec and real-time traffic data were considered; however, half of studies received data per 30 sec:

- Based on 30-sec traffic data, researchers would aggregate data and take measures to improve the mobility or safety. Some de-noising strategies are employed in these studies since near real-time raw data.
- The 10-, 20-, and 30-sec raw data have random noise since it is nearly real-time data during short period. Generally, radar would archive speed, volume and occupancy information at given short period.
- Raw data are difficult to work with in a modeling framework in the optimization system (*16*). Thus, raw data would first aggregate into a given interval, such as 5 or 10 mins.

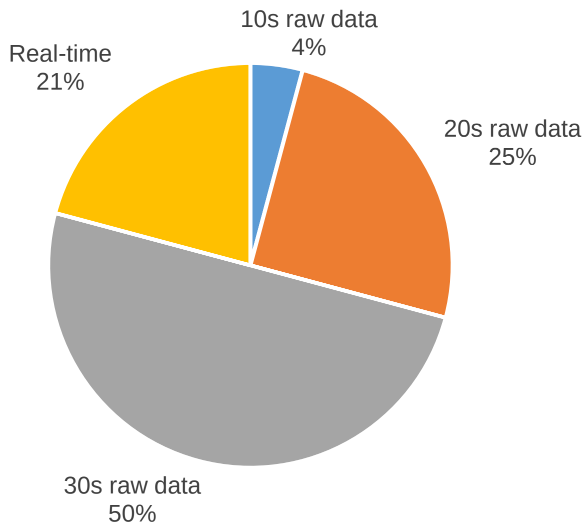


Figure 14. Summary of Sampling Rate

2.6 Prediction Models

A prediction model is used to address the relationship between input data (traffic conditions) and output (crash occurrence risk). Supervised machine learning methods have been widely used. Table 4 summarizes prediction models developed in previous studies. As the output of dynamic crash prediction is usually expressed as a binary variable (e.g., crash occurrence or not, alarm or not, etc.), 33% of previous studies used discrete choice models (such as binary logistic model). Data-driven classification models, such as Support Vector Machine, Neural Network, Bayesian Network, CART, etc., were also developed.

Table 4. Summary of Prediction Models

Category	Method	Years	Number of Studies	Percent
Parametric Regression	Discrete Choice Model	2004–2015	14	33%
	Other Regression Model	2003–2012	6	14%
Data-Driven Method	Support Vector Machine	2014–2017	3	7%
	Neural Network	1999–2014	9	21%
	Bayesian Network	2004–2015	8	19%
	CART and Others	2010–2011	2	5%

Table 5 summarizes the performance (accuracy) of prediction models. Two measures were investigated in previous studies:

- Successful Alarm Rate = number of predicted crashes that are true crashes ÷ number of predicted crashes
- False Alarm Rate = number of predicted crashes that are not “true” crashes ÷ number of predicted crashes

Different prediction models have diverse accuracy. A Bayesian Network has the highest successful alarm rate (92%) but its range is wide (55–92%). A Support Vector Machine and a Neural Network have similar performance. Only a few previous studies provided a False Alarm Rate. It is worth noting that the performance data were derived from selected testing data in research projects rather than real implementation; the value may not present the real performance of the models in practice. A pilot study is needed to evaluate the different methods using more diverse data for obtaining "real" performance.

Table 5. Summary of Prediction Model Performance

Category	Method	Successful Alarm Rate	False Alarm Rate
Parametric Regression	Discrete Choice Model	58–82%	20%
	Other Regression Model	65–78.3%	
Machine-Learning Method	Support Vector Machine	67–88%	20.9%
	Neural Network	70–86%	
	Bayesian Network	55–92%	10-23.7%
	CART and Others	70–74%	

2.7 Summary

This chapter summarized findings from a comprehensive literature review; a more detailed summary is shown in Appendix A. The major findings are as follows:

- Traffic crash occurrence is associated with prevailing traffic flow characteristics (e.g., speed, density, volume). The risk of traffic crash in a short term can be predicted based on real-time traffic flow data.
- Most previous studies focused on freeway segments due to the relatively simple crash-traffic relationship and data availability. Limited studies were found to apply on arterials.
- Prediction inputs mainly include speed variation, average speed, density (occupancy), and volume. These data were collected primarily from fixed vehicle detectors (loops, Microwave Vehicle Detection Systems, or Bluetooth devices). The sampling rate (time interval for collecting raw traffic data) is 10–30 secs.
- Most studies adopted supervised machine learning models to predict the crash occurrence risk. The successful alarm rate reaches 58–92%.

It worth noting that the previous studies were developed and tested on limited datasets (Florida, California, Germany), and the evaluation results may not represent their real performance in a more "generalized" traffic condition. In addition, these studies focused on modeling and algorithm research rather than products. The applicability of the dynamic crash prediction models in real traffic conditions was not proven from the previous studies.

3 Evaluation of Existing Dynamic Crash Prediction Technologies

The chapter summarizes the evaluation of existing dynamic crash prediction technologies. Unlike the literature review, which focused on academic research, the evaluation aimed to identify vendors that provide dynamic crash prediction products and compare different systems. The evaluation results were used to understand the application status of dynamic crash prediction and select products/vendors for the pilot study.

3.1 Evaluation Procedure

The evaluation procedure is shown in Figure 15. The evaluation included three major steps—Search, Interview, and Evaluation.

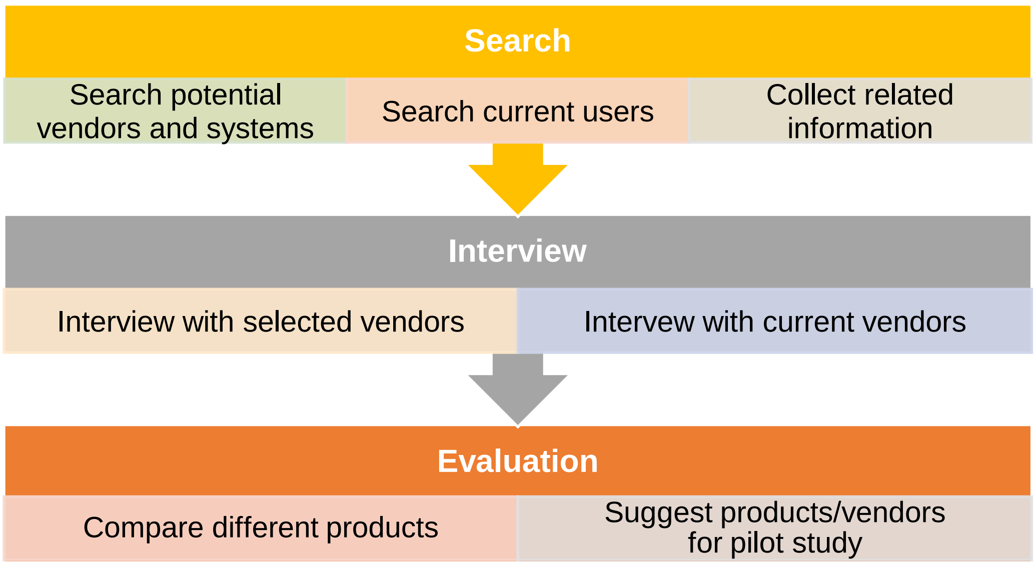


Figure 15. Procedure of Evaluation of Existing Crash Dynamic Prediction Technologies

3.1.1 Search

The research team searched vendors that potentially provide dynamic crash prediction function and/or traffic data support using the Google search engine, advertisements, news, and government reports. All information related to potential vendors (e.g., official websites, news, flyers, and reports) was collected and reviewed to identify vendors and technologies that met study needs. Meanwhile, to obtain practical experience of dynamic crash prediction implementation, users that have implemented dynamic crash prediction or that planned to implement/test a system were searched and identified. Information related to current users was also collected and reviewed.

3.1.2 Interview

To obtain more detailed information on vendors and current users, the research team interviewed selected vendors and current users through teleconferences, in-person meetings, and email questionnaires. The interviews aimed to (1) confirm vendor technologies/systems satisfying the study objective, (2) collect detailed information on vendor technologies/systems that are unavailable in documents, and (3) understand the practice of dynamic crash prediction from current users, including successful experience and lessons they learned from the practice.

3.1.3 Evaluation

By assembling the information collected from searches and interviews, the research team identified vendors that provided dynamic crash prediction technologies. The research evaluated each identified vendor and its technologies based predefined criterions. Evaluation results were used to select vendors for the pilot study.

3.2 Evaluation Criteria

Evaluation of the selected dynamic crash prediction systems was based on the following criteria:

- Functionality
- Performance and impacts
- Data and local resource needs
- Usability
- Maturity

3.2.1 Functionality

This category indicates the available functions provided by the selected systems. The expected functions include the following:

- *Dynamic crash prediction* – A system can dynamically predict crash risk based on real-time traffic and environmental data. This function emphasizes a critical prediction before crash occurrence in real-time rather than a long-term prediction used in a traditional safety study. This function is the minimum (enforced) requirement for dynamic crash prediction.
- *Crash risk alarm* – A system can send out an alarm when a predicted crash risk is higher than a configurable threshold. The system alarm could be an alert message to operators or a signal to trigger actions. This function is required in the system.
- *Crash prevention actions* – A system can performance actions to prevent crash occurrence after the prediction. This function is an optional module that enhances the functionality of dynamic crash prediction.
- *Incident detection* – A system can detect incident occurrence based on prevailing traffic conditions or video detection as quickly as possible for emergency responders. This function aims to reduce incident detection time and prevent secondary crashes and recurring congestion. This function module is optional.

- *Long-term crash prediction* – A system can analyze historical crash data and predict the crash risk of roadway sites for the long term (e.g., monthly or yearly). This function does not emphasize a prediction in real-time (dynamic) and is an optional module.
- *Presentation* – A system can present prediction results and statistics in various formats (text, figures, heat maps, etc.) on GIS maps or in printable reports. Other information presentations, such as CCTV monitoring, are optional add-ons.
- *Roadway facility* – The dynamic crash function should be implemented on various roadway facilities, such as interstate highways, expressways, arterials, and signalized intersections.

3.2.2 Performance

This category indicates the performance of dynamic crash prediction of the selected systems. The major prediction performance measures include the following:

- *Prediction accuracy* – Prediction accuracy is defined as the percentage of crash events that can be successfully predicted. This criterion is a key performance measure, and a high prediction accuracy is expected.
- *False Alarm Rate (FAR)* – A FAR is defined as the percentage of predicted crash events that are not true. A low FAR is expected.
- *Prediction threshold* – The prediction systems alerts a crash occurrence if the predicted crash risk is higher than the threshold. Threshold is a critical factor influencing prediction performance (prediction accuracy and false alarm rate). Increasing the threshold can reduce false alarms but may result in failure of alerting true crash events. Decreasing the threshold may have an opposite effect. The threshold should be configured to allow users to determine the best tradeoff between the two performance measures.
- *Timeliness* – A prediction system can predict crash risk in advance of crash occurrence. A long warning time allows traffic agencies to have enough opportunities to apply actions for preventing crash occurrence.

3.2.3 Benefits

This category indicates the benefits of the dynamic crash prediction systems related to safety and operations:

- *Safety impact* – The safety impact of the dynamic crash prediction systems can be measured by the number of crashes prevented and surrogate safety indicators such as reduction in average speed and speed variance:
- *Operations impact* – The operational impact of the dynamic crash prediction systems can be measured by flow rate and average speed.

- *Incident management impact* – The impact of the systems on incident management can be measured by the reduction in incident reaction time due to dynamic crash prediction if a crash prevention action fails.

3.2.4 Data Needs

This category indicates the expected data for the implementation of dynamic crash prediction:

- *Historical data* – Historical data are needed to calibrate the prediction model and include crash data, traffic data, weather data, construction activities, traffic signal, and incident events.
- *Real time data* – Real-time data are used as model inputs. The calibrated model predicts crash risk based on the real-time inputs. Real-time data are the same as historical data.
- *Primary data sources* – The required data can be retrieved from local data sources (State database and TMC sensors) or third-party data sources. Vendors having independent third-party data sources can operate their systems on roadway facilities where local data sources are unavailable.

3.2.5 Usability

This category indicates the usability of the three systems; measures of usability include the following:

- *Platform* – The system can be implemented on a cloud platform that does not need additional hosting resources:
- *Data Application Programming Interface (API)* – The system should provide API to connect local data sources for real-time data feeding:
- *User interface* – A user interface allows users to monitor system outputs and set system configurations.
- *Integration with TSM&O systems* – The system can be integrated into existing or planning TSM&O systems.
- *Implementation without local data* – The system can be implemented on roadway segments with local data sources.

3.2.6 Technical Maturity

This category indicates the technical maturity of the three systems. Maturity measures include the following:

- *Pilot study of dynamic crash prediction* – If the dynamic crash prediction function has been tested in a pilot study.

- *Implementation in Florida* – If the system (non-function of dynamic crash prediction) has been implemented or tested in Florida.
- *Implementation in other states or countries* – If the system (non-function of dynamic crash prediction) has been implemented or tested in other states or countries.

3.3 Evaluation Results

The research identified 3 technologies from 11 potential vendors (see Appendix B). The three vendors stated that they have the dynamic crash prediction functions but only one vendor (WayCare) has implantable systems. Evaluation of the three technologies is shown in Table 6 through Table 11.

Table 6. Comparison of Selected Systems for Functionality

Function	Requirement	WayCare	Vendor 2	Vendor 3
Dynamic crash prediction (DCP)	Required	Included and tested	Stated	Stated
Crash risk alarm	Required	Included and tested	Stated	Stated
Crash prevention action	Optional	Tested	Not included	Not included
Police high-visibility		Yes	No	No
Dynamic message		Yes	No	No
Incident response		Yes	No	No
Incident detection	Optional	Included	No	No
Long-term crash analysis and prediction	Optional	Included	Included	Included
Web-based GIS map	Required	Yes	Yes	Yes
Formatted report	Required	Yes	Yes	Yes
CCTV	Optional	Yes	No	No
Roadway facility types for DCP				
Interstate	Required	Yes, tested	Yes, but not tested	Yes, but not tested
Arterial	Required	Yes, not tested	Yes, but not tested	Yes, but not tested
Intersection	Required	Yes, but not tested	Yes, but not tested	Yes, but not tested

Table 7. Comparison of Selected Systems for Performance

Function	WayCare	Vendor 2	Vendor 3
Prediction accuracy	56%*	Unknown	Unknown
False Alarm Rate	Unknown	Unknown	Unknown
Prediction threshold	Unknown	Unknown	Unknown
Timeliness	2 hours, but may vary over sites	Unknown	Unknown

* Source: WayCare pilot study in Las Vegas

Table 8. Comparison of Selected Systems for Benefits

Function	WayCare	Vendor 2	Vendor 3
Primary crash reduction	17%*	Unavailable	Unavailable
Secondary crash reduction	23%*	Unavailable	Unavailable
Speed reduction	91% of drivers reduce speed to 65 mph or lower*	Unavailable	Unavailable
Operational impact	Unavailable	Unavailable	Unavailable
Incident reaction time reduction	12%*	Unavailable	Unavailable

* Source: WayCare pilot study in Las Vegas

Table 9. Comparison of Selected Systems for Data Needs

Function	WayCare	Vendor 2	Vendor 3
Historical Data			
Crashes	Required for 3–5 yrs; the more years, the better performance	Required for county and state to get enough samples	Required
Traffic	Optional*	Required	Required
Weather	Optional*	Required	Unclear
Construction events	Optional*	Unclear	Unclear
Traffic signaling	Optional*	Unclear	Unclear
Incident	Optional*	Unclear	Unclear
Real-time Data			
Crash	Required (for model fine-tune)	No	No
Traffic	Optional*	Required	Required
Weather	Optional*	Unclear	Unclear
Construction events	Optional*	Unclear	Unclear
Traffic signaling	Optional*	Unclear	Unclear
Incident	Optional*	Unclear	Unclear
Others			
Primary data sources	TMC + third party	TMC	TMC

*WayCare has third-party data sources for historical and real-time data.

Table 10. Comparison of Selected Systems for Usability

Function	WayCare	Vendor 2	Vendor 3
Platform	Cloud-based	Cloud-based	Cloud-based
Data API	Yes	Yes	No, TMC should provide
User interface	Web	Web	Web
Integration with TSM&O systems/ devices	Yes, tested	Unclear	Unclear
Implementation without local data sources	Yes	No	No

Table 11. Comparison of Selected Systems for Maturity

Function	WayCare	Vendor 2	Vendor 3
Previous pilot studies of DCP	Yes, Nevada	No	No
Implementation (non-DCP) in Florida	Yes, Tampa, Pinellas, District 4	No	No
Implementation (non-DCP) in other states or countries	Yes	Chicago, IL	CA, Canada

3.4 Existing Users

The research team identified and interviewed four local agencies that implemented or are interested in dynamic crash prediction. A summary of these users is given in Table 12.

Table 12. Summary of Existing Users

	Las Vegas, NV	Tampa, FL	Pinellas County, FL	Chicago, IL
System	WayCare	WayCare	WayCare	Open Data Nation
Implementation of DCP	Pilot study	No	No	No
Current application	Unknown	Incident identification	Incident identification	Long-term crash prediction (monthly)
Have plan to implement DCP	Unknown	Yes	Yes	Unknown
Facility type	Freeway	Freeway and major arterials	Freeway and major arterials	Roadway network
Dynamic crash prevention actions with DCP	DMS, stationary police car, incident management	No	No	Freeway and major arterials

3.5 Summary

Based on the evaluation results, major conclusions are as follows:

- A limited number of vendors provide dynamic crash prediction functions that are an innovative technology. Only one vendor (WayCare) has relatively mature systems for dynamic crash prediction functions, although two other vendors stated that they have similar technologies. A summary of these three technologies is shown in Table 13.

Table 13. Summary of Comparison

Function	WayCare	Vendor 2	Vendor 3
Maturity of dynamic crash production	Best	In development	In development
Crash prevention actions after prediction	Tested	No	No
Documented performance and benefits	Yes	No	No
Additional functions	Yes, incident detection	Yes, long-term crash prediction	Yes, long-term crash prediction
Data requirement	Relatively low	High	High/medium
Third-party data sources	Yes	No	No
Implementability without local data sources	Yes	No	No
Easy to deploy	Yes	Yes	Yes

- No current users were found to implement dynamic crash prediction systems, although many agencies showed interest. Only the Nevada Department of Transportation (NDOT) tested the functions in a pilot study. Two local agencies in Florida (City of Tampa, Pinellas County) have implemented the WayCare system; however, they do not apply the dynamic crash prediction.
- Only one pilot study was found that tested WayCare’s dynamic crash prediction functions in Las Vegas. The pilot study produced some preliminary results (see Table 7 and

- Table 8) and proved the concepts of dynamic crash prediction. Detailed information on the Las Vegas pilot study is given in Appendix C. Information on dynamic crash prediction is still limited because:
  - The pilot study was conducted on freeway only, so performance of dynamic crash prediction on arterials is unknown.
  - Performance results (crash reduction) were based on a three-month pilot study and are not very accurate and reliable.
  - Evaluation results were reported by WayCare; no independent third-party evaluation was found.

4 Pilot Study

This chapter describes the pilot study conducted in FDOT District 4 that aimed to demonstrate dynamic crash prediction in the Florida roadway environment and evaluate the performance of dynamic crash prediction technologies with "real" traffic conditions. Pilot study results were used to develop recommendations for implementing dynamic crash prediction in Florida.

4.1 Pilot Study Procedure

The pilot study procedure, as shown in Figure 16, consisted of three stages—Planning and Preparation, Training, and Testing. First, the research team, in collaboration with the Project Manager, determined vendors/technologies for the pilot study and invited the three vendors for evaluation. Only WayCare committed to completing the pilot study within the project budget and timeline. Thus, WayCare was selected to conduct the pilot study.

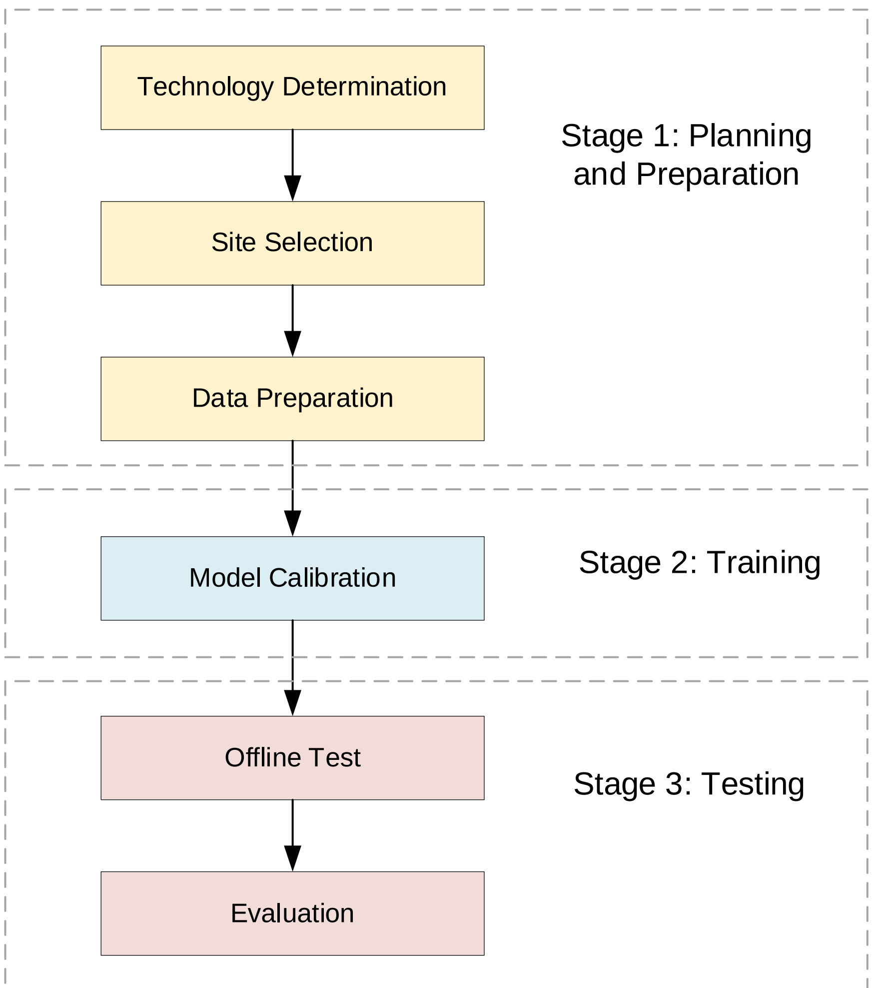


Figure 16. Procedure of Pilot Study in FDOT District 4

4.2 Study Sites

The research team selected study sites for the pilot study based on the following criteria:

- *Historical records* – Testbeds should have significant crash records and high traffic volumes such that enough sample data can be collected for model training.
- *Diversity* – Testbeds should cover various roadway types (e.g., interstates and major arterials) and geographic zones in the transportation network managed by FDOT District 4.
- *Local data resources* – Testbeds should be equipped with traffic monitoring systems (e.g., point detectors, Bluetooth, etc.) and potentially other data collection resources, if available (e.g., weather station).
- *Traffic management capabilities* – ITS should be available for use for applying actions to reduce crash risk after prediction for the testbeds and should be connected to the SunGuide system at the TMCs.

Three segments in District 4 (I-95, Sunrise Blvd, PGA Blvd) were identified initially based on the selection criteria; however, the traffic sensors on PGA Blvd could not provide qualified traffic data. Thus, two segments, covering freeways and arterials, were selected for the pilot study.

4.2.1 Site 1 – I-95

4.2.1.1 Overview

The first testbed, as shown in Figure 17, is an interstate freeway segment along I-95 in Broward County. The boundary limits are Hallandale Blvd (Exit 18) to Davie Blvd (Exit 26). The site includes the first, second, and fifth highest crash segments based on crash data for 2015–2018. The characteristics of Site 1 are summarized in Table 14.

Table 14. Basic Characteristics of Site 1, I-95

Boundary	Hallandale Blvd (S) to Davie Blvd (N)
Facility type	Interstate
Length	8.516 mi
Lane configuration (one-direction)	4 (general use) + 2 (express)
Number of interchanges	9 (including two ends)
Speed Limit	65 mph
AADT	275,000–319,000 vpd

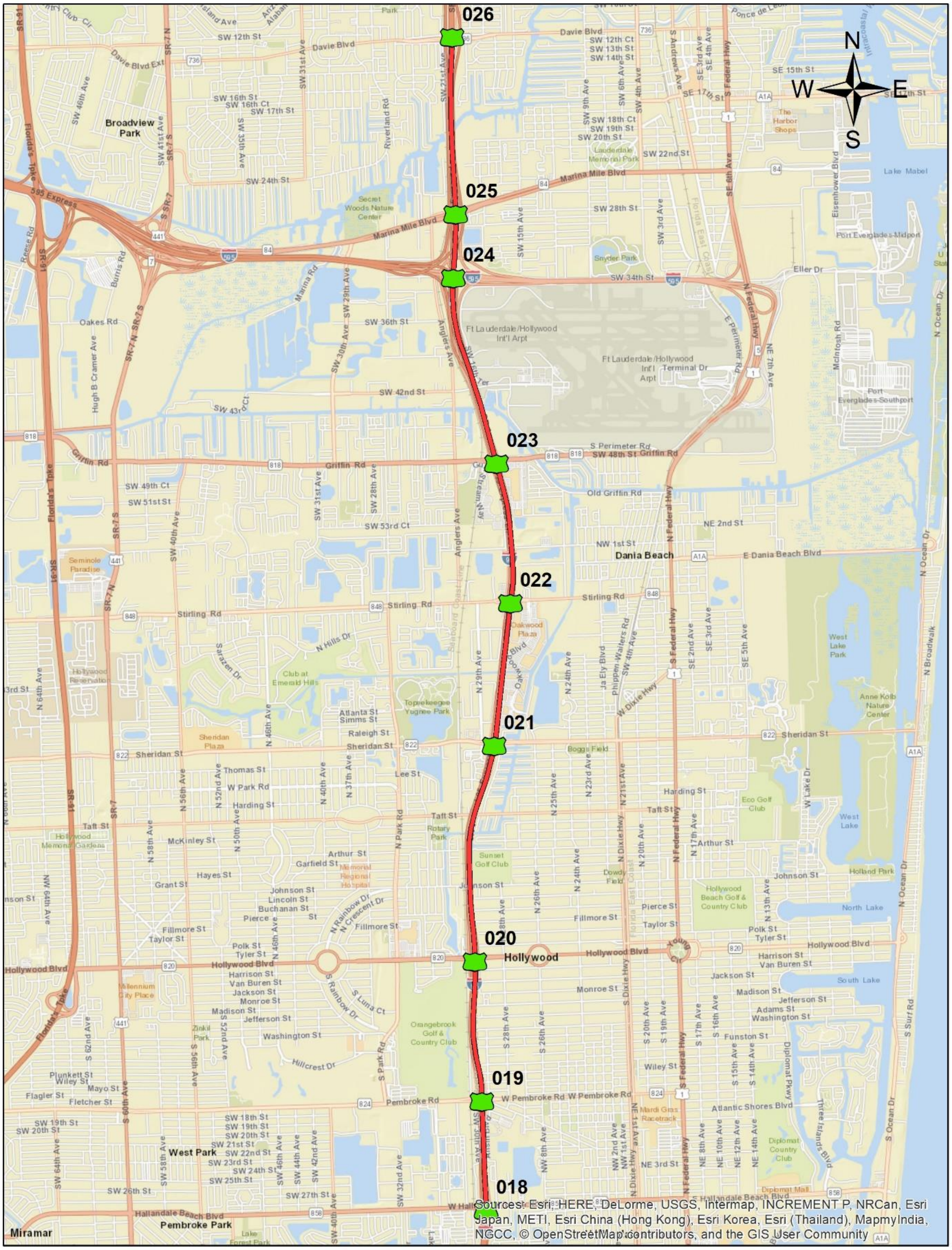


Figure 17. Site 1 – I-95 Segment (Hallandale Blvd to Davie Blvd)

4.2.1.2 Historical Crash Data

Historical crash data show that the I-95 segment experienced very high crash frequencies for 2015–2018. Average yearly crash frequency was 2,617 per year, as shown in Figure 18, which is more than 3,000 crashes per year after 2016. The monthly trend (Figure 19) shows that the top crash months were October, November, December, and January, each having 250+ crashes per month. Based on data collected from Signal Four Analytics, the I-95 study site includes five segments ranked 1st, 2nd, 5th, 6th, and 7th among the top 25 highest crash segments on I-95 in Broward County for 2015–2018, as shown in Figure 20. Spatial analysis of crashes over 0.1-mi segments, as shown in Figure 21, indicates two sub-segments experiencing 300 or more crashes per year, three experiencing 100–300 crashes per year, and six experiencing 50–100 crashes per year.

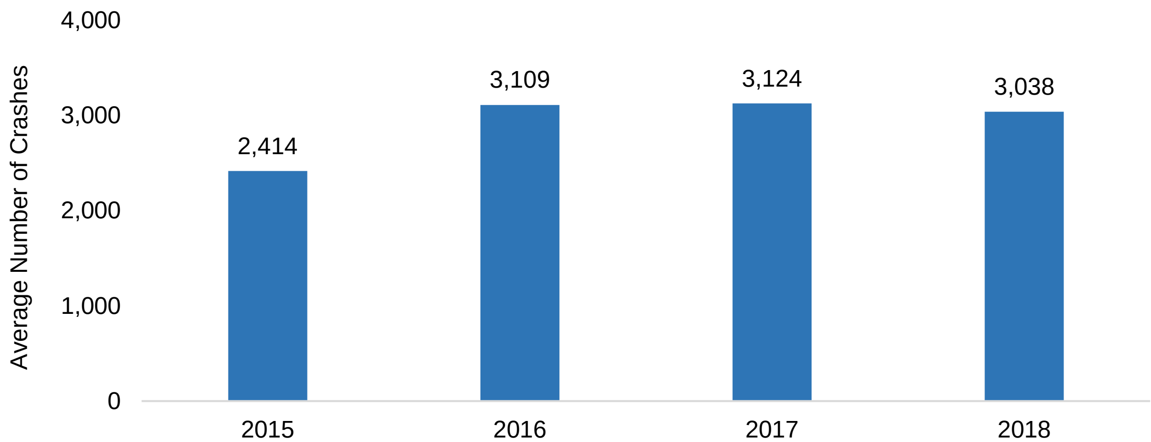


Figure 18. Average Yearly Crashes, I-95 Study Site

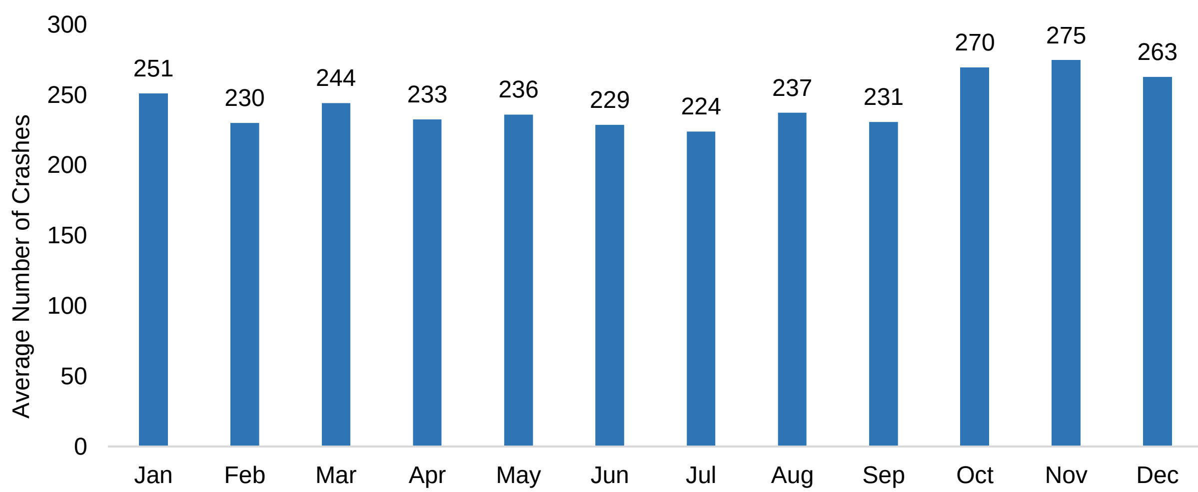


Figure 19. Average Monthly Crashes, I-95 Study Site

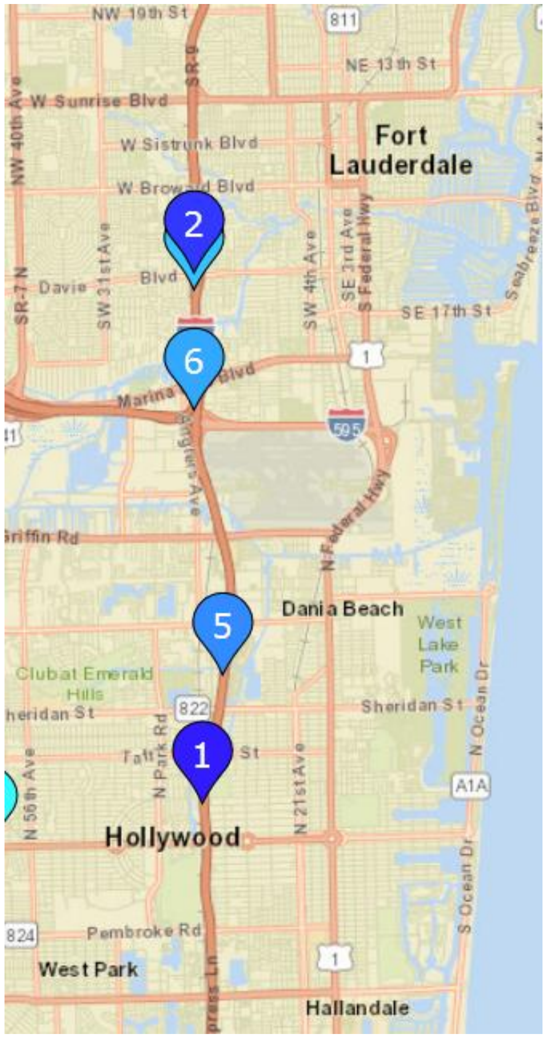


Figure 20. Top Crash Segments, I-95 Study Site

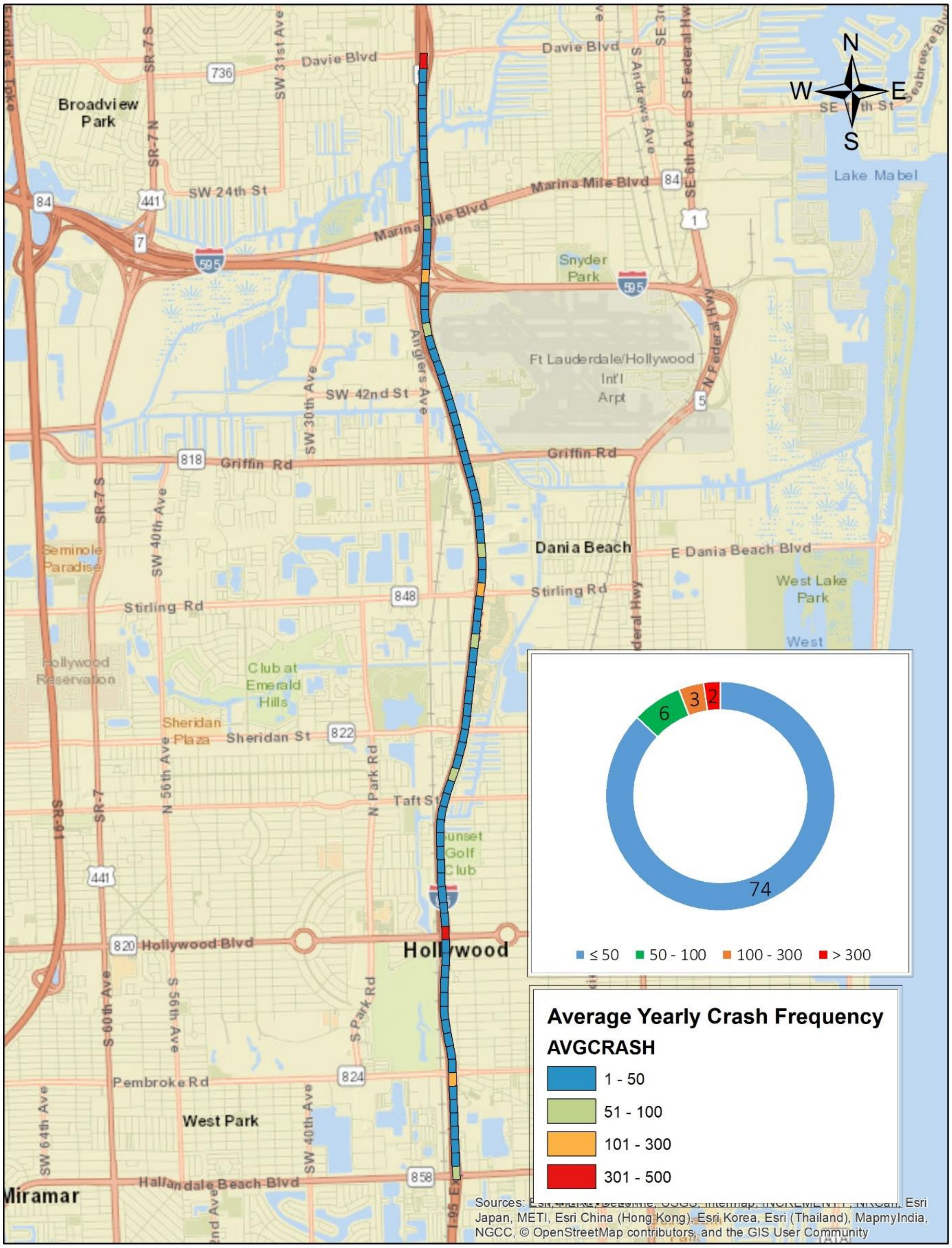


Figure 21. Spatial Distribution of High Crash Segments, I-95 Study Site

4.2.2 *Local Data Sensors*

The I-95 study site is equipped with 56 microwave vehicle detection systems (MVDS) in the NB direction and 53 MVDS in the SB direction. The average distance between two MVDS sensors is approximately 0.4 mi. The layout of the traffic sensor locations is shown in Figure 22.

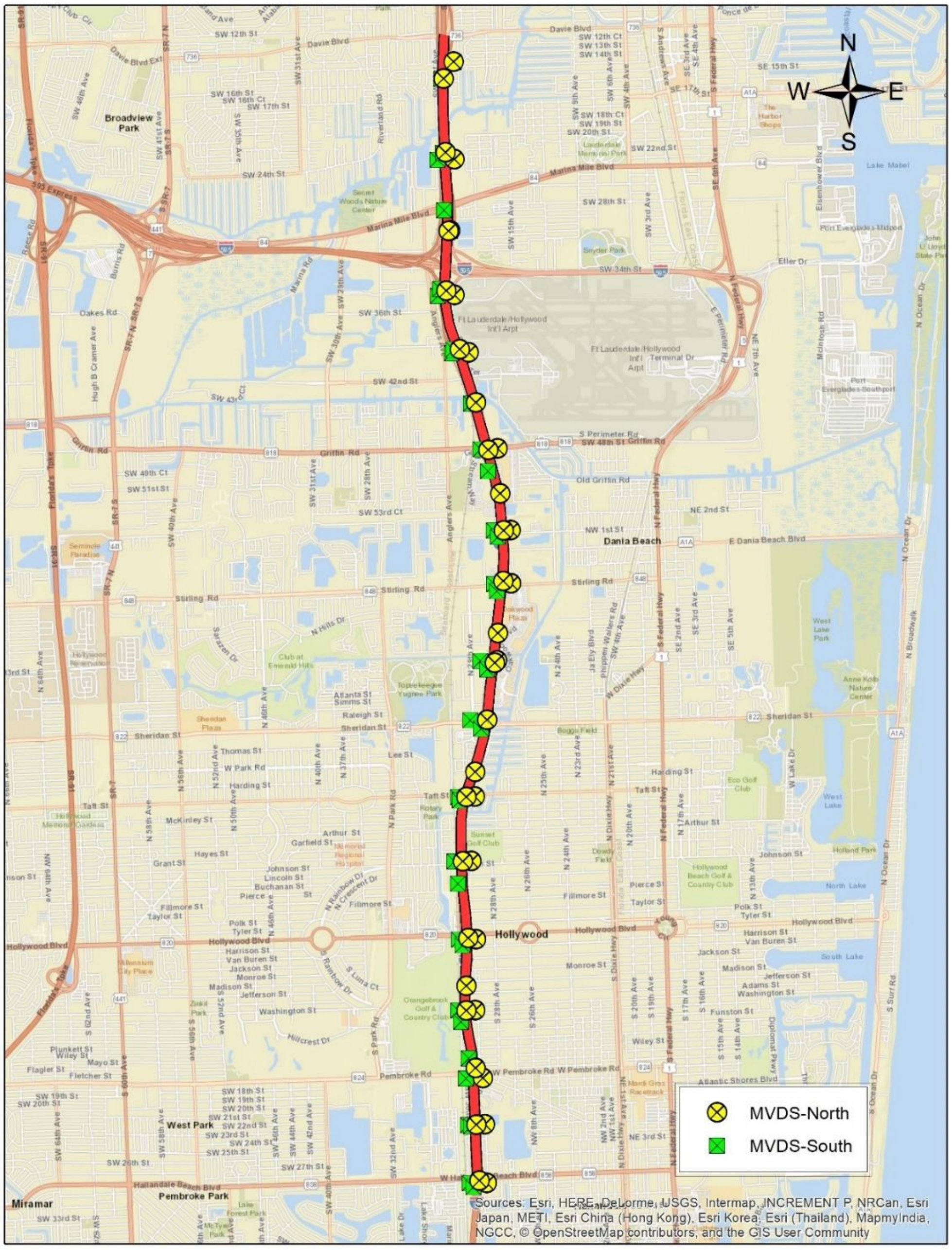


Figure 22. Layout of Traffic Sensors, I-95 Study Site

4.2.2.1 TSM&O Programs and Devices

The I-95 site includes 10 Dynamic Message Sign (DMS) devices at the following locations, as shown in Figure 23:

- I-95 N of I-595
- I 95 N of Griffin St
- I-95 NB S of Griffin Rd
- I-95 S of Green St
- I-95 SB N of Sheridan St
- I 95 NB S of Hollywood Blvd
- I 95 SB S of Hollywood Blvd
- I-95 SB at Pembroke Rd

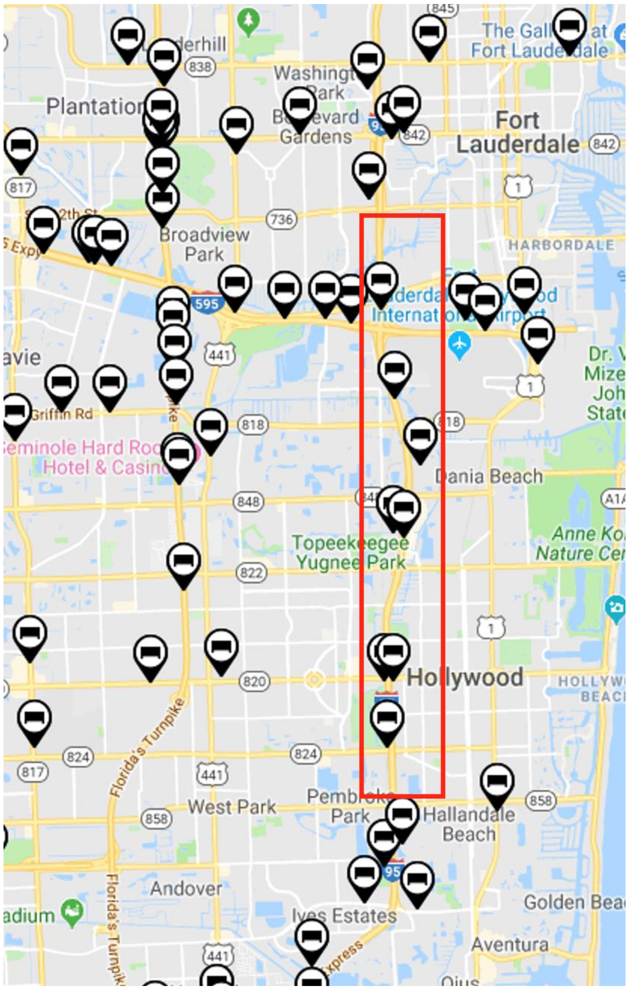


Figure 23. Locations of DMS Devices, I-95 Study Site

4.2.3 Site 2 – E Sunrise Blvd Site

4.2.3.1 Overview

The second site is a principal arterial segment along E Sunrise Blvd between I-95 and US-1 within the boundaries of Fort Lauderdale. The site includes 16 signalized intersections, 2 pedestrian signals, and 1 railroad crossing. Characteristics of Site 2 are presented in Table 15.

Table 15. Summary of Characteristics, E Sunrise Blvd Study Site

Category	Characteristics	Value
Geometry	Boundary	I-95 – US-1
	Facility type	Principal arterial
	Length	3.023 mi
	Lane configuration	3 per direction
	Number of signalized intersections	16
	Number of pedestrian signals	2
	Number of railroad crossings	1
	Median Configuration	Raised median + directional opening
	Speed limit	40 mph (I-95 – N Federal Hwy), 35 mph (N Federal Hwy – US-1)
Traffic	AADT	45,000–58,000 vpd

4.2.3.2 Historical Data

Average yearly crash frequency on the E Sunrise Blvd segment (see Figure 24) was 180 crashes for 2014–2018), as shown in Figure 25. The monthly trend (Figure 26) shows that each month experienced 50 or more crashes on this segment. Based on crash data from Signal Four Analytics, the E Sunrise study site includes three intersections that are ranked the 57th, 72nd, and 89th among the top 100 highest crash intersections in Broward County for 2015–2018, as shown in Figure 27. Spatial analysis of crashes over 0.1-mi segments indicates 11 sub-segments experiencing 30 or more crashes per year, as shown in Figure 28.

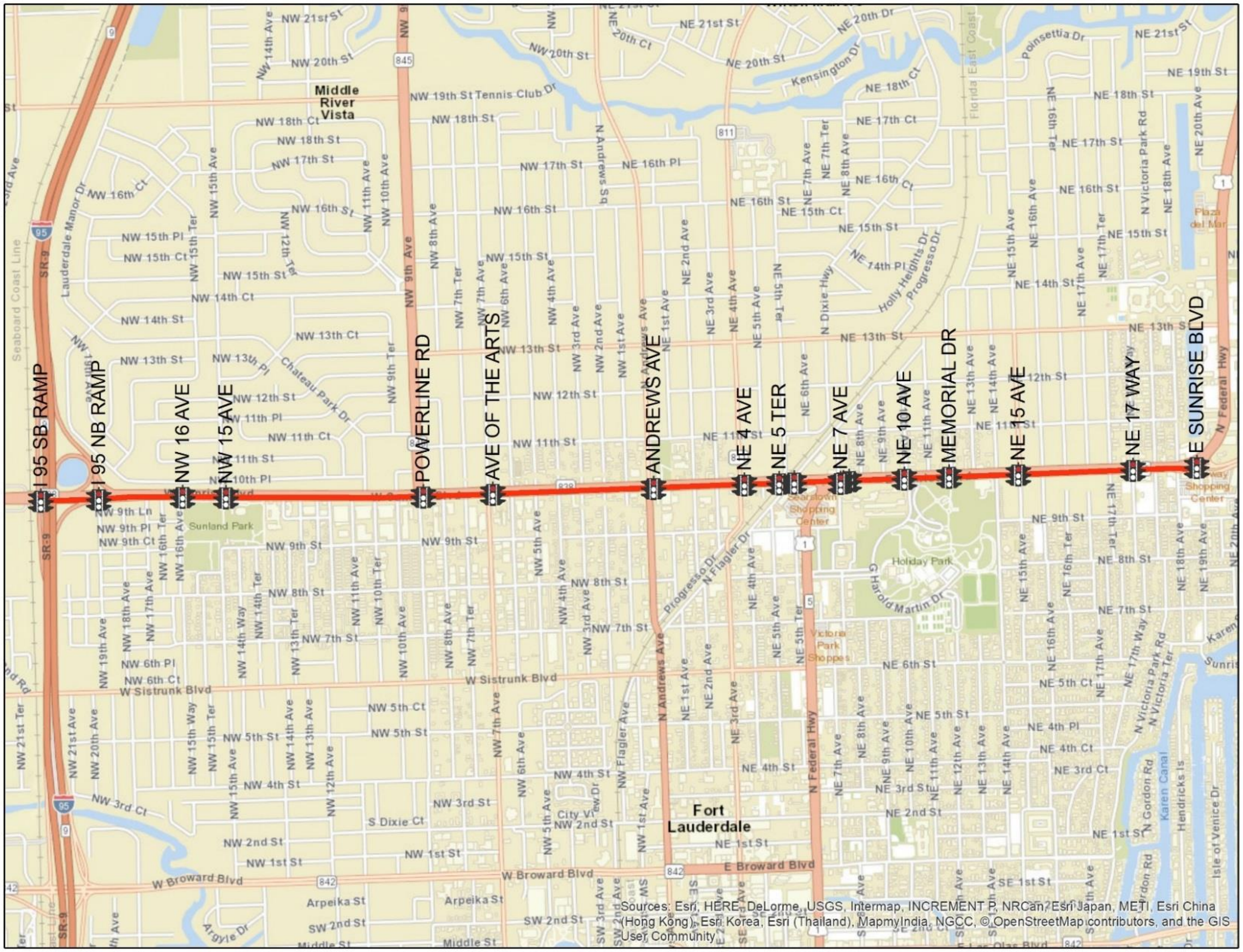


Figure 24. Site 2 – E Sunrise Blvd Study Segment (I-95 to US-1)

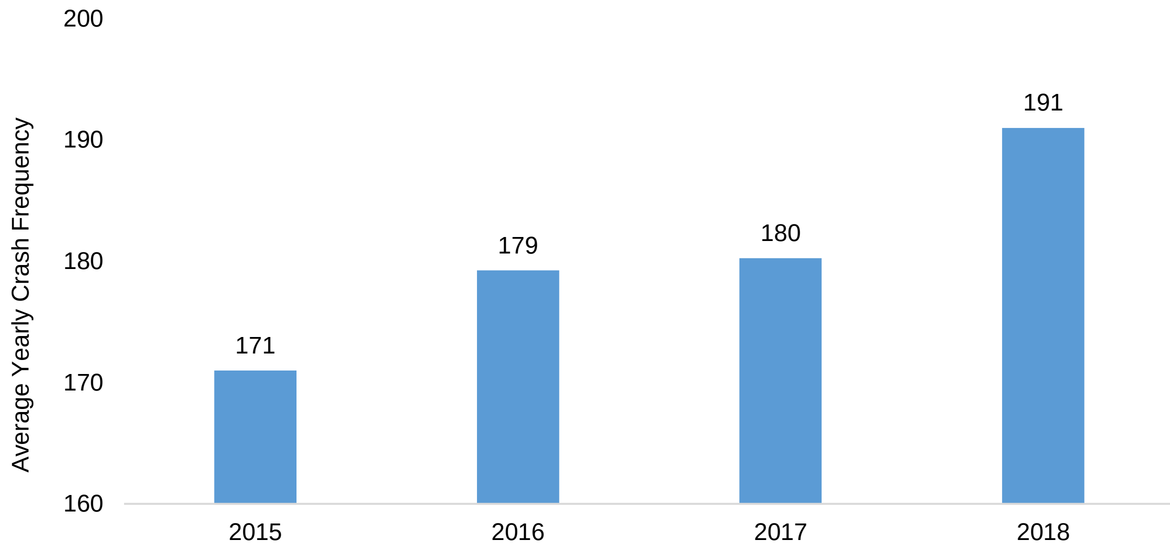


Figure 25. Average Yearly Crash Frequency, E Sunrise Blvd Study Site

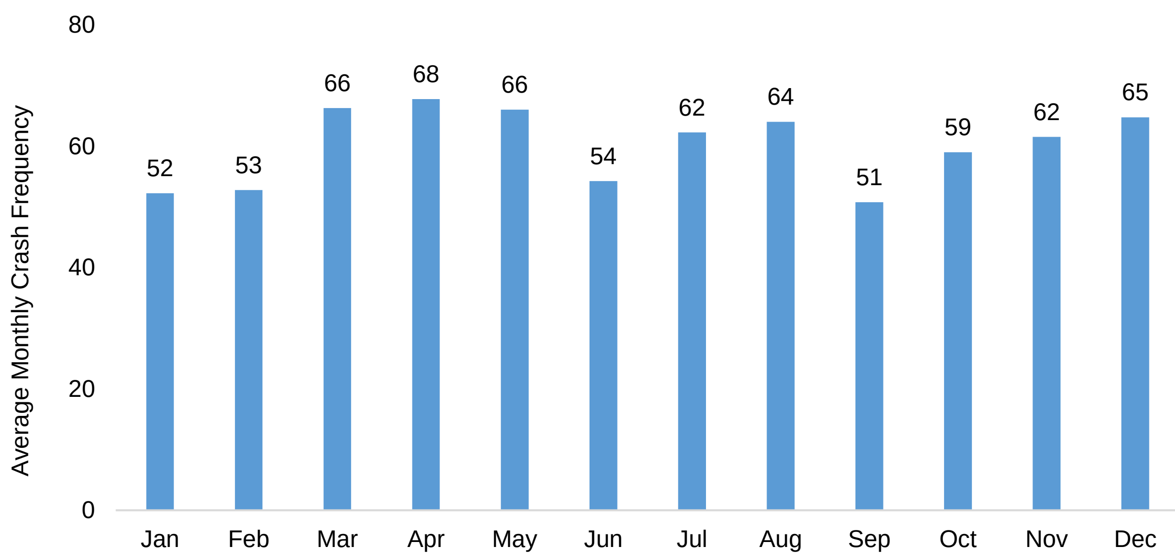


Figure 26. Average Monthly Crash Frequency, E Sunrise Blvd Study Site

Figure 27. Top Crash Intersections, E Sunrise Blvd Study Site

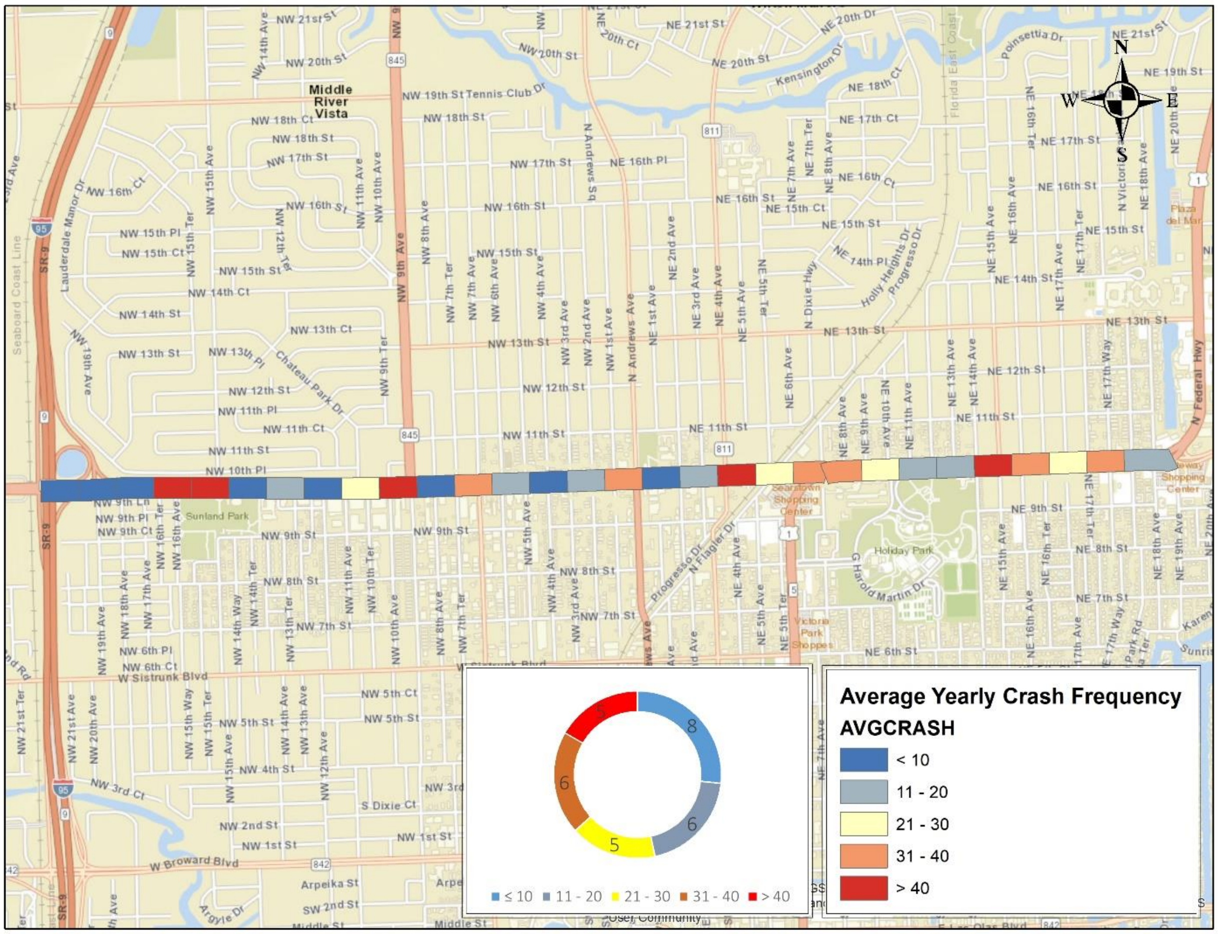


Figure 28. Crash Density, E Sunrise Blvd Study Site

4.2.4 Local Data Sensors

In total, 5 MVDS devices and 7 Bluetooth devices are installed on the E Sunrise Blvd corridor for traffic data collection. The device locations are shown in Table 16 and Figure 29.

Table 16. Summary of MVDS and Bluetooth Devices, Sunrise Blvd Study Site

SunGuide ID	Type	Roadway	Cross Street
M-11	MVDS	SR-838/E Sunrise Blvd	NW 17th Ave
M-12	MVDS	SR-838/E Sunrise Blvd	NW 12th Ave
M-13	MVDS	SR-838/E Sunrise Blvd	NW 4th Ave
M-14	MVDS	SR-838/E Sunrise Blvd	NE 8th Ave
M-15	MVDS	SR-838/E Sunrise Blvd	NE 17th Ave
B-15	Bluetooth	SR-838/E Sunrise Blvd	I-95
B-16	Bluetooth	SR-838/E Sunrise Blvd	NW 17th Ave
B-17	Bluetooth	SR-838/E Sunrise Blvd	NW 12th Ave
B-18	Bluetooth	SR-838/E Sunrise Blvd	NW 9th Ave
B-19	Bluetooth	SR-838/E Sunrise Blvd	Andrews Ave
B-20	Bluetooth	SR-838/E Sunrise Blvd	NE 17th Ave
B-21	Bluetooth	SR-838/E Sunrise Blvd	SR-5/Federal Hwy/US-1

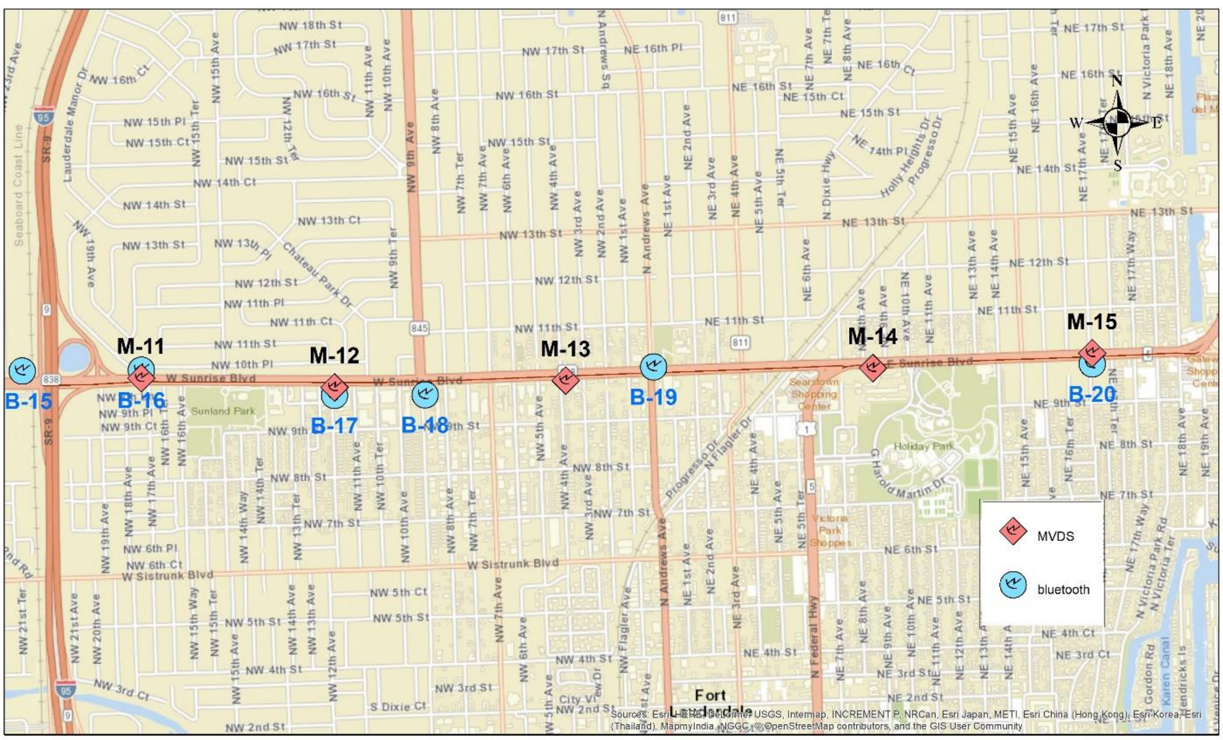


Figure 29. Locations of MVDS and Bluetooth Devices, E Sunrise Blvd Study Site

There are eight CCTV devices on the E Sunrise corridor to monitor traffic operations and incidents, as shown in Table 17 and Figure 30.

Table 17. Summary of CCTV Devices, E Sunrise Blvd Study Site

SunGuide ID	Roadway	Cross Street
C-19	SR-838/E Sunrise Blvd	I-95
C-20	SR-838/E Sunrise Blvd	NW 9th Ave
C-21	SR-838/E Sunrise Blvd	Andrews Ave
C-22	SR-838/E Sunrise Blvd	NE 4th Ave
C-23	SR-838/E Sunrise Blvd	Flagler Dr
C-24	SR-838/E Sunrise Blvd	NE 8th Ave
C-25	SR-838/E Sunrise Blvd	NE 15th Ave
C-26	SR-838/E Sunrise Blvd	SR-5/Federal Hwy/US-1

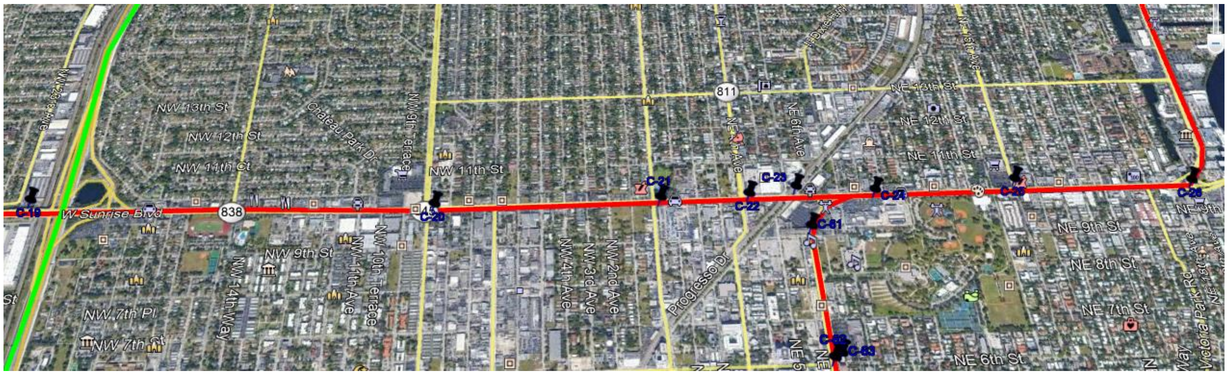


Figure 30. Locations of CCTV devices, E Sunrise Blvd Study Site

4.2.4.1 DMS

There are two DMS devices on the E Sunrise corridor, as shown in Table 18. The locations of DMS devices are shown in Figure 31.

Table 18: Attributes of DMS Devices, E Sunrise Blvd Study Site

SunGuide ID	Direction	Cross Street	Text Capacity
D-03	EB	Before Powerline Rd	2 lines, 13 characters
D-04	WB	Beyond NE 17th Way	2 lines, 13 characters

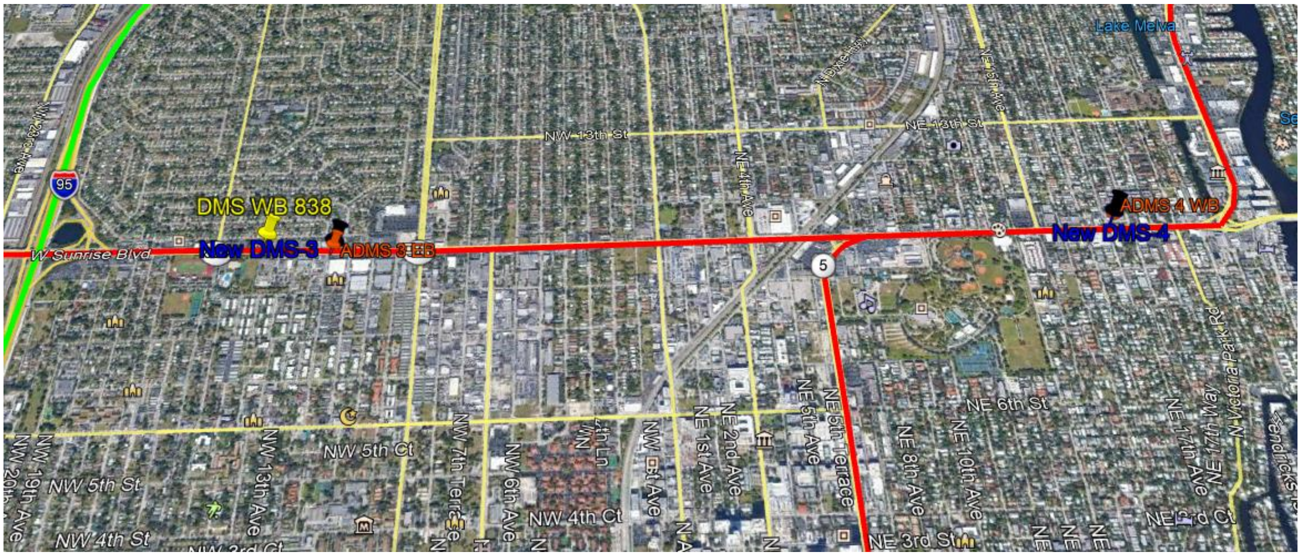


Figure 31: Locations of DMS devices, E Sunrise Blvd Study Site

4.2.5 Site 3 – PGA Blvd

4.2.5.1 Overview

The third site is a principal arterial segment on PGA Blvd between a Florida's Turnpike SB off-ramp and Prosperity Farms Rd in Palm Beach County. This site includes five signalized intersections, including three top crash intersections in West Palm Beach County. Site characteristics are presented in Table 19, and the layout of Site 3 is shown in Figure 32.

Table 19. Summary of Characteristics, PGA Blvd Study Site

Category	Characteristics	Value
Geometry	Boundary	Florida's Turnpike–Prosperity Farms Rd
	Facility type	Principal arterial
	Length	3.865 mi
	Lane configuration	3–4 per direction
	Number of signals	13
	Median Attributes	Raised median, full/directional openings
	Speed Limit	45 mph
Traffic	AADT	38500–75000

4.2.5.2 Historical Crash Data

Spatial analysis of crash over segments between two signals indicates 11 sub-segments experiencing 30 or more crashes per year, as shown in Figure 33. Yearly crash frequencies on the PGA Blvd segment were 371, 482, and 376 crashes per year for 2016, 2017, and 2018, respectively, as shown in Figure 34. The monthly trend (Figure 35) shows that each month experienced 27 or more crashes on this segment. Based on Signal Four Analytics data, the PGA Blvd study site includes three intersections among the top 100 highest crash intersections in West Palm Beach County for 2016–2018.

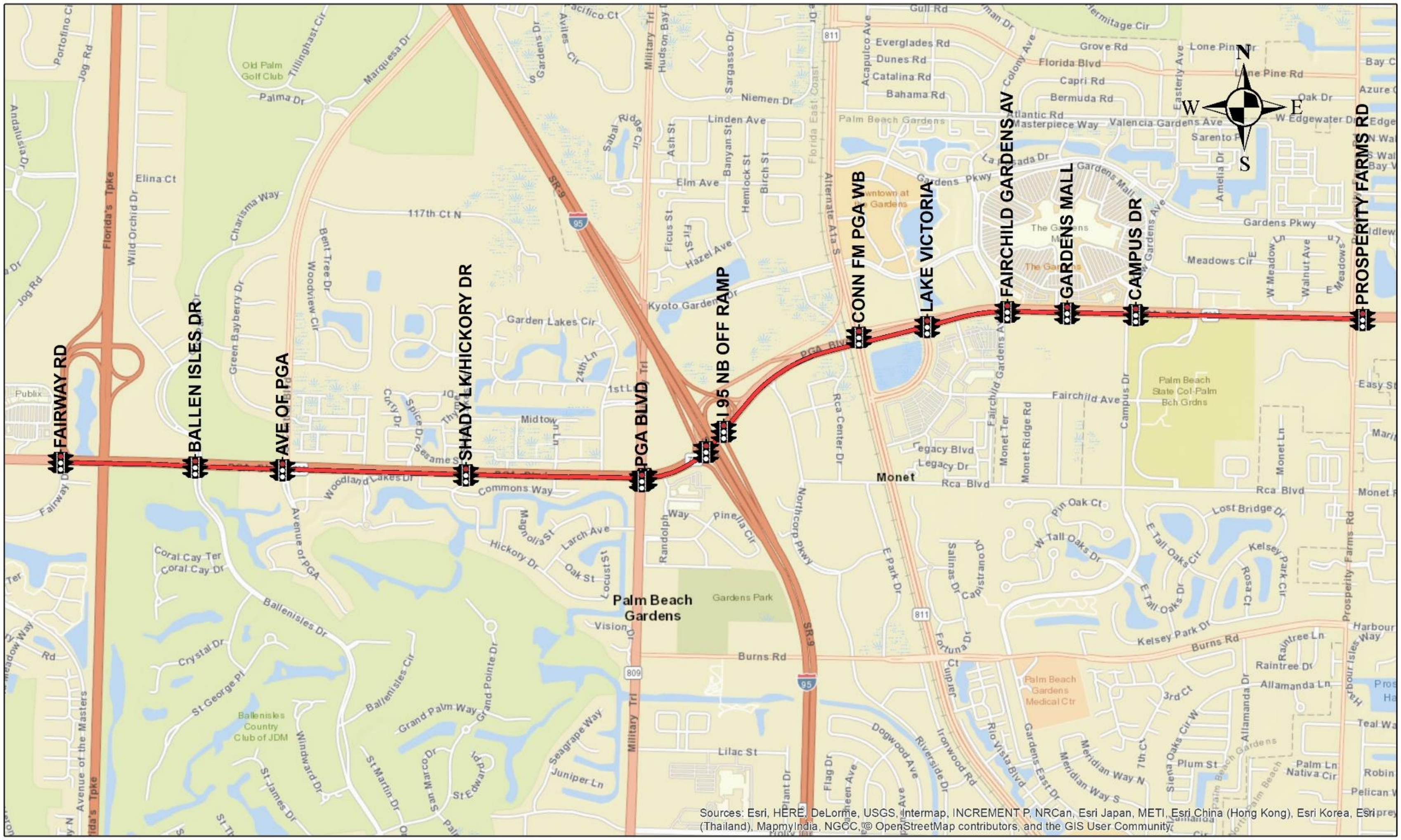


Figure 32. Site 3 – PGA Blvd (Florida's Turnpike–Prosperity Farms Rd) Study Site

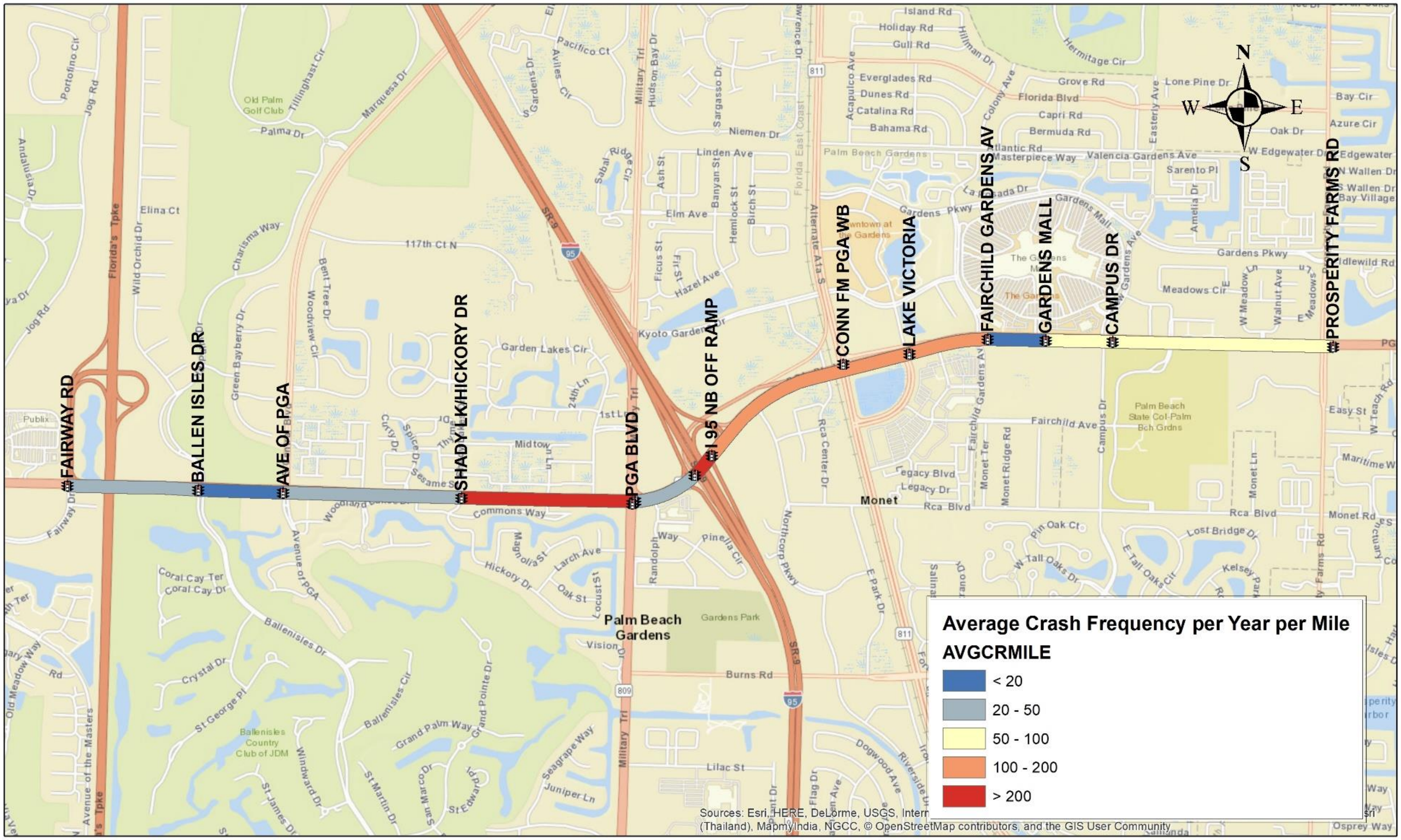


Figure 33. Crash Density, PGA Blvd Study Site

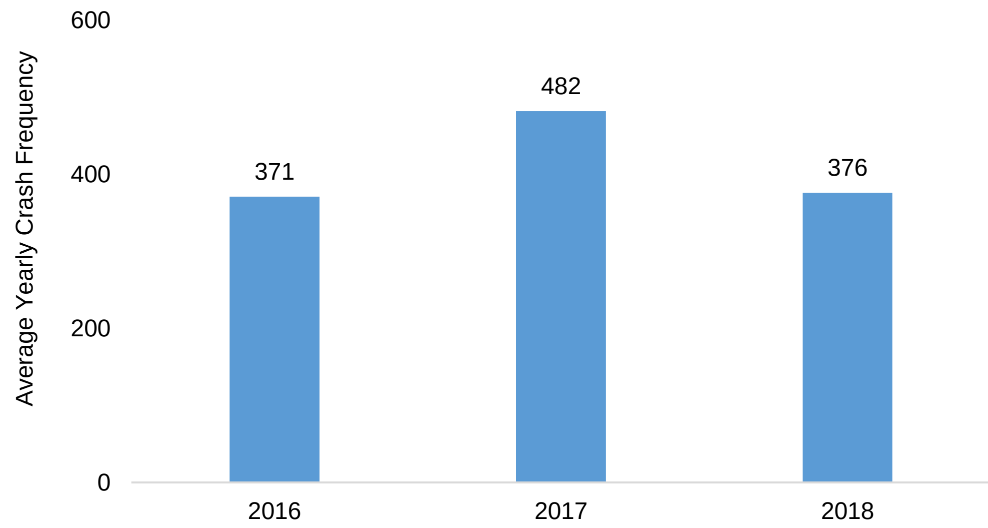


Figure 34. Average Yearly Crash Frequency, PGA Blvd Study Site

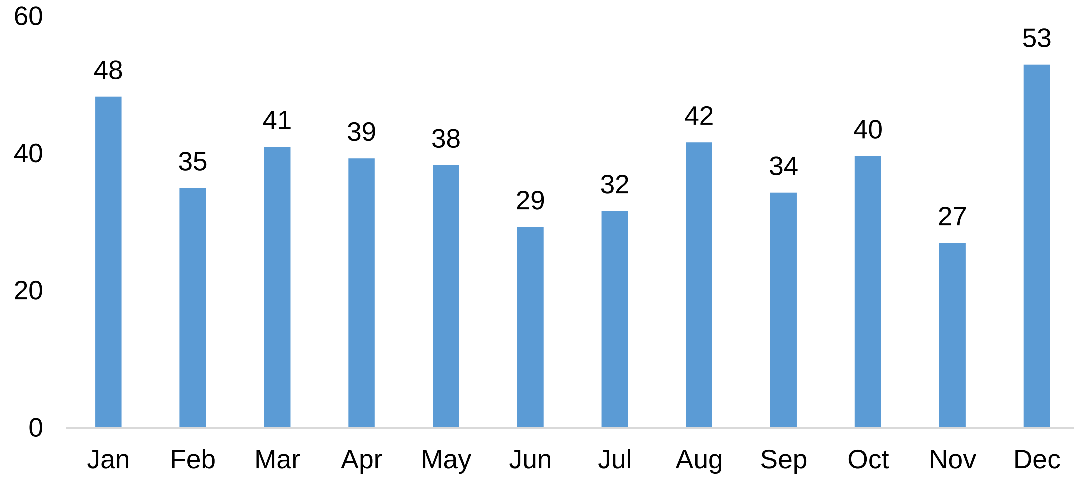


Figure 35. Average Monthly Crash Frequency, PGA Blvd Study Site

4.2.5.3 Data Sensors

The PGA corridor includes six portable traffic-monitoring stations, five Bluetooth devices, and four MVDS devices for traffic data collection. Device locations are shown in

Table 20.

Table 20. Summary of Traffic Sensor Locations, PGA Blvd Study Site

Device ID	Type	Roadway	Location
930072	Portable	PGA Blvd	W of SR-91/FL Turnpike
930073	Portable	PGA Blvd	E of SR-91/FL Turnpike
930074	Portable	PGA Blvd	W of SR-9/I-95
935300	Portable	PGA Blvd	E
935402	Portable	PGA Blvd	E of SR-811/Alt-A1A
930712	Portable	PGA Blvd	E of Prosperity Farms Rd
1*	Bluetooth	PGA Blvd	Turnpike
2*	Bluetooth	PGA Blvd	Central Blvd
3*	Bluetooth	PGA Blvd	Military Trail
4*	Bluetooth	PGA Blvd	Garden Mall
5*	Bluetooth	PGA Blvd	Prosperity Farms Rd
1*	MVDS	PGA Blvd	FL Turnpike to Ballenisles Dr
2*	MVDS	PGA Blvd	Shady Lakes Dr to Military Trail
3*	MVDS	PGA Blvd	I-95 to RCA Blvd
4*	MVDS	PGA Blvd	Campus Dr to Prosperity Farms Rd

*Not official number

Three CCTV devices are available on PGA Blvd for incident management:

- PGA Blvd at Florida's Turnpike
- PGA Blvd at I-95
- PGA Blvd at Gardens Mall

4.2.5.4 DMS

Three DMS devices are installed on PGA Blvd:

- EB – PGA Blvd, W of Military Trail
- WB – PGA Blvd, W of Fairchild Gardens
- EB – PGA Blvd, W of Prosperity Farms Rd

4.3 Data Preparation

4.3.1 Data Collection

The research team collect data at Sites 1 and 2 in two stages: (1) collecting historical data for five years (2015–2019) for model calibration purposes and (2) collecting latest data in 2020 for offline testing. Historical data were also collected at Site 3; however, Site 3 was not included in the offline test since its testing data were unavailable. Data collection for the three sites is summarized in Table 21.

Table 21. Summary of Data Collection

Characteristics		Site 1: I-95	Site 2: Sunrise Blvd	Site 3: PGA Blvd[3]
Facility type		Interstate	Principal Arterial	Principal Arterial
Length		8.516 mi	3.023 mi	3.865 mi
Calibration Data				
Time Frame		2015 - 2019	2015 - 2019	2015 – 2019
Traffic Data	Source	RITIS[1]	D4 TMC	Here
	Items	Volume, speed, occupancy	Volume, speed, occupancy	Volume, speed, occupancy
	Spatial Resolution	By lane	By lane	By segment
	Sampling Rate	20 sec	1 min	1 min
Crash Data	Source	SignalFour[2]	SignalFour[2]	SignalFour[2]
	Items	Date, time, direction	Date, time, direction	Date, time, direction
Offline Testing Data				
Time Frame		Jan, Feb, Jul in 2020	Jan, Feb, Jul in 2020	N/A
Traffic Data	Source	RITIS[1]	D4 TMC	
	Items	Volume, speed, occupancy	Volume, speed, occupancy	
	Spatial Resolution	By lane	By lane	
	Sampling Rate	20 sec	1 min	
Crash Data	Source	SignalFour[2]	SignalFour[2]	
	Items	Date, time, direction	Date, time, direction	

[1] Regional Integrated Transportation Information System, https://ritis.org/.
[2] https://s4.geoplan.ufl.edu/.
[3] Historical data collected for Site 3, but site not tested as offline testing data unavailable.

Traffic information such as speed, volume, and occupancy was collected from different sources. For the I-95 site, detector data were downloaded from the RITIS website (https://ritis.org/). The interface of the RITIS detector tool is shown in Figure 36. Because there were no detector data available for the E Sunrise Blvd site in the RITIS database, traffic data collected via the MVDS were requested from FDOT District 4. Both datasets were lane-by-lane raw count data.

Crash data were also collected, including information such as crash time and crash location. Crash data for both study sites were downloaded via the SignalFour Analytics website developed by the GeoPlan Center at the University of Florida (https://s4.geoplan.ufl.edu/analytics/). The interface of the SignalFour Analytics web application is shown in Figure 37.

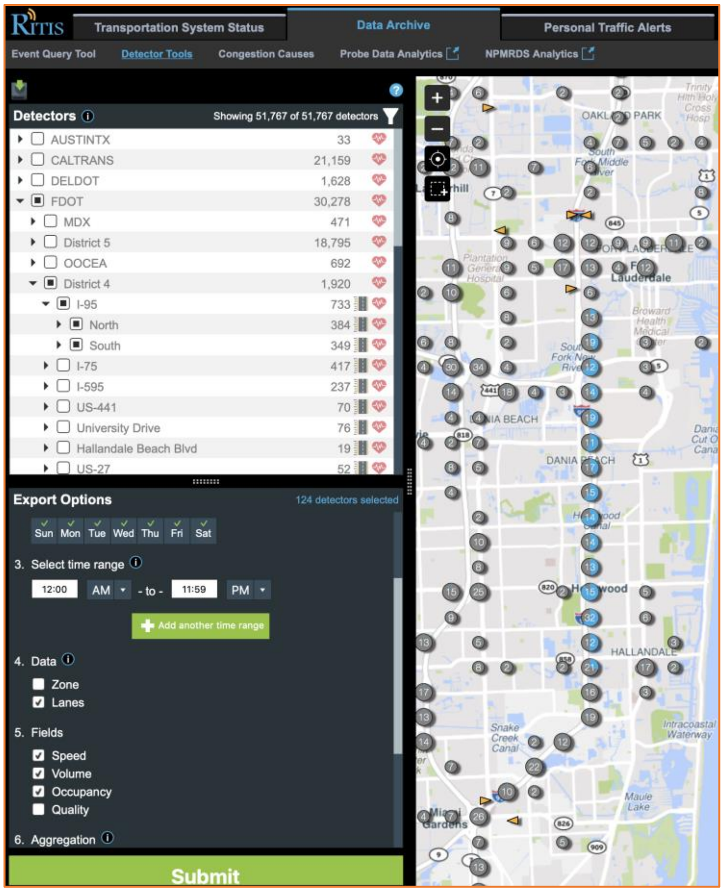


Figure 36. Interface of RITIS Tools

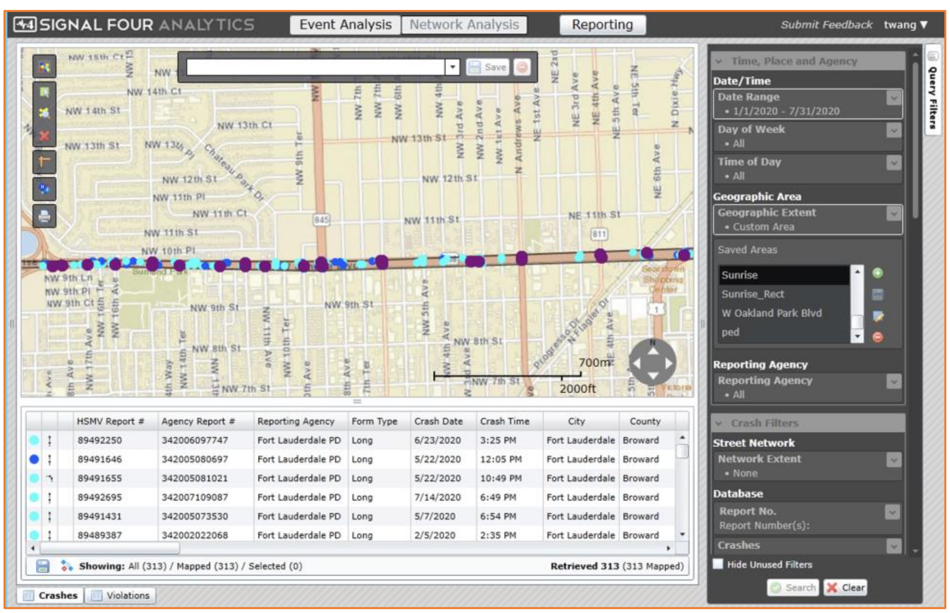


Figure 37. Interface of SignalFour Tools

4.3.2 Data Processing

Raw calibration data (traffic and crash data for 2015–2019) were directly provided to WayCare for model calibration. Testing data for January, February, and July 2020 were processed by the research team to generate testing datasets. The data process procedure is described as follows.

Step 1: Split the Data by Segment – The WayCare model predicts crash risk for sub-zones rather than for whole corridors. The sub-zones used for I-95 and E Sunrise Blvd are shown in Figure 38 and Figure 39, respectively. The research team grouped traffic data and crash data by sub-zone.

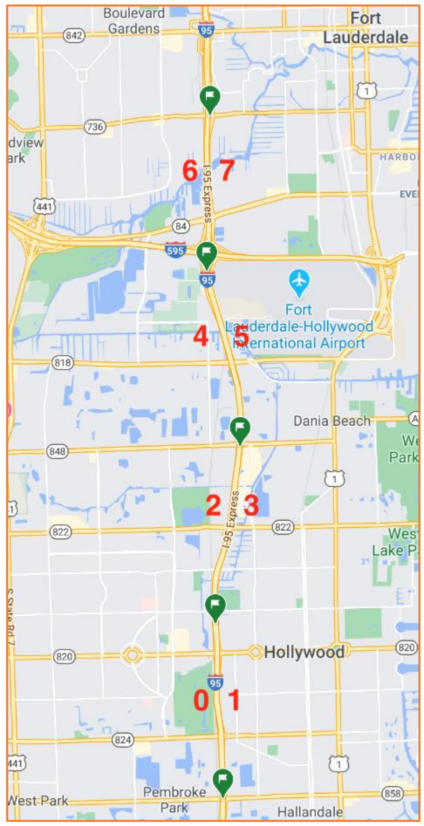


Figure 38. Sub-Zones at I-95 Site

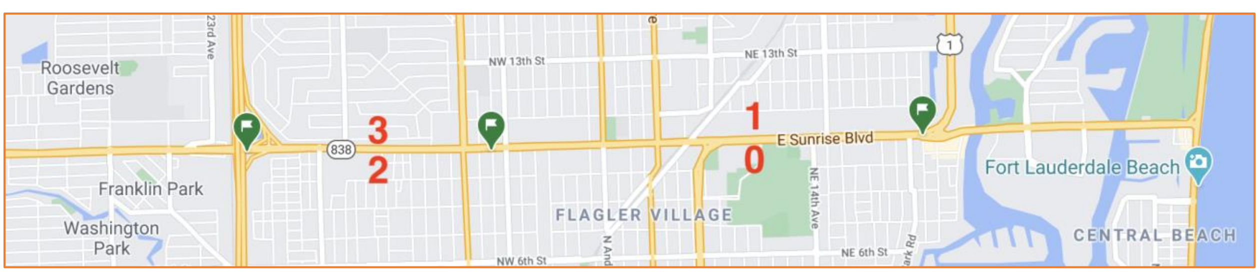


Figure 39. Sub-Zones at E Sunrise Blvd Site

Step 2: Split Data by Time – The WayCare model predicts the crash risk for the next three hours for a sub-zone based on the previous nine-hour traffic and crash data. The research team divided the whole day into four periods, as shown in Table 22, and grouped the traffic and crash data by the four time periods for each sub-zone.

Table 22. Time Periods Used for Prediction Input and Output

Time Periods	Input Period (9 hrs before prediction)	Prediction Period (3 hrs)
AM (morning)	9:00 PM (previous day)–6:00 AM	6:00–9:00 AM
MD (mid-day)	3:00 AM–12:00 PM	12:00–3:00 PM
PM (afternoon)	6:00 AM–3:00 PM	3:00–6:00 PM
Night	12:00 PM–9:00 PM	9:00 PM–12:00 AM

Step 3: Filter Data – The raw traffic dataset on the I-95 sites contained some errors, such as extreme values, missing data, or incorrect codes. These error data were removed to avoid their impact on the prediction performance. Data filtering conditions were as follows:

- Speed – > 0 mph and < 100 mph
- Volume – > 0 vehicle per lane per 20 sec and < 50 vehicles per lane per 20 sec
- Occupancy – > 0% and < 80%

Upon completion of the three steps, the research team generated the model inputs, including traffic and crash data. The data description of the model inputs is given in Table 23.

Table 23. Data Fields for Model Inputs

Field	Description/Format
Traffic Data	
Time	Time in 24-hour format – HH:MM:SS.S
Detector ID	Unique number for traffic sensors
Lane ID	Integer number indicating a lane
Direction	N/S for I-95, E/W for Sunrise Blvd
Volume	Number of vehicles per lane per 20 sec (I-95 site) Number of vehicles per lane per one min (Sunrise Blvd)
Occupancy	%
Speed	Miles per hour
Crash Data	
Datetime	Crash date time – MM/DD/YYYY HH:MM
Latitude	Latitude of crash location, decimal degree
Longitude	Longitude of crash location, decimal degree
Direction	N/S for I-95, E/W for E Sunrise Blvd

4.4 Model Calibration

The WayCare team calibrated its prediction models based on the five-year historical data (2015–2019) for the two study sites. A machine learning methodology was used to build the connection between the input traffic/crash characteristics and the output crash risk. WayCare randomly selected samples from the calibration data for model training and evaluation, as shown in Figure 40.

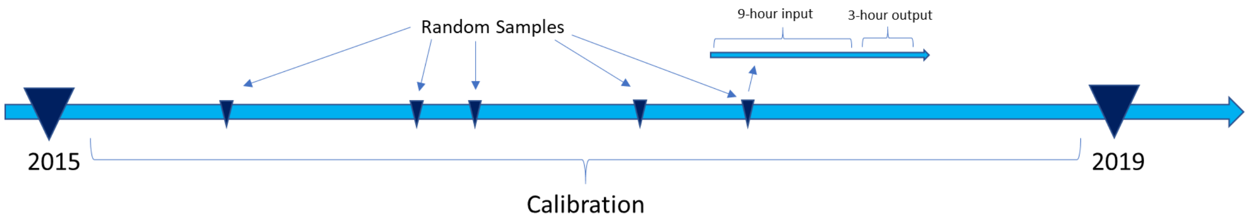


Figure 40. WayCare Model Training and Evaluation

The model was incorporated in a simple tool for offline model testing. With this tool, the users upload the model inputs—nine-hour traffic data and crash data—through a webpage (Figure 41) and downloads the prediction results using the same webpage (Figure 42).

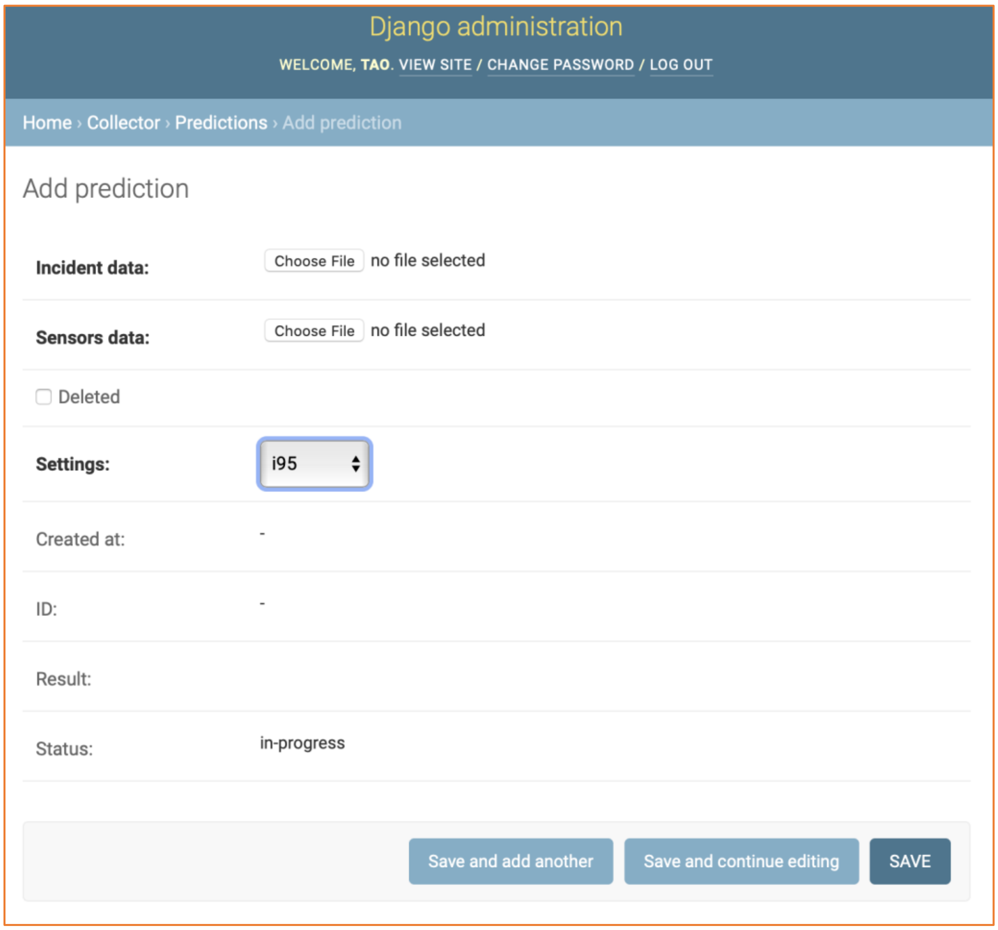


Figure 41. Offline Prediction Interface – Upload Traffic and Crash Data

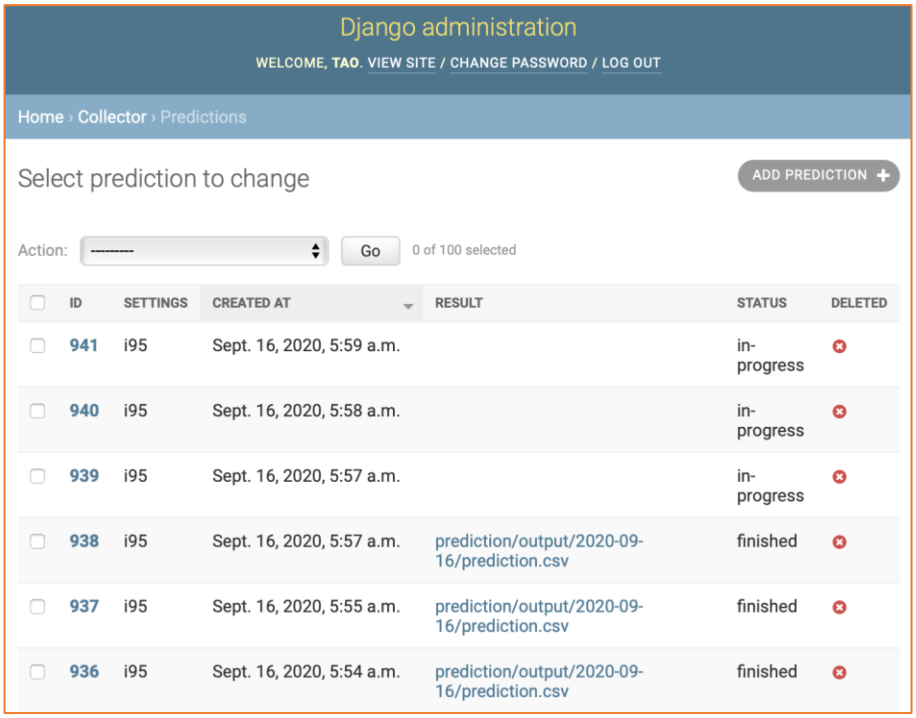


Figure 42: Offline Prediction Interface – Download Prediction Results

The detailed model calibration procedure and results are given in Appendix A and Appendix B, respectively.

4.5 Evaluation Methods

The research team conducted an offline test on the calibrated WayCare Model using the 2020 dataset. The offline test was independent of WayCare's evaluation and assessed the performance of the dynamic crash prediction methodology in the Florida roadway environment. The evaluation criterions and procedure are given below.

4.5.1 Performance Measures

The prediction results of the model were compared to the archived crash events that corresponded to the road segment and time period. Four assessment types were used in the evaluation of prediction quality, as shown in Table 24:

- True Positive (TP) – Assesses the degree to which there were crashes; the model successfully predicted them and thus triggered true alarms.
- False Positive (FP) – Assesses the degree to which there were no crashes; the model predicted this incorrectly and triggered false alarms. This is also called Type I error.
- True Negative (TN) – Assesses the degree to which there were no crashes; the model predicted them correctly and did not give alarms.

- False Negative (FN) – Assesses the degree to which there were crashes; the model predicted them incorrectly and did not give alarms. This is also called Type II error.

Table 24. Concepts of Prediction Performance Metrics

	Crash Cases	**No-Crash Cases**
Alarm (predicted crash)	TP (correctly predicted crash events)	FP (Type I error)
No alarm (predicted no crash)	FN (Type II error)	TN (correctly predicted non-crash events)

With the assessment types listed above, the following performance metrics (measures) were calculated for the evaluation:

- Precision – probability of true alarms; that is, the percentage of true alarms that correctly predicted the crash cases, calculated as the number of true alarms divided by the total number of alarms.

$$\text{Precision} = \frac{\text{Correctly Predicted Crashes}}{\text{Total Alarms}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

- False Alarm Rate – probability of false alarms; that is, the percentage of false alarms that an alarm is generated but no "real" crash event occurs.

$$\text{False Alarm Rate} = 1 - \text{Precision} = \frac{\text{Alarms w/o Crash Occurring}}{\text{Total Alarms}} = \frac{\text{FP}}{\text{TP} + \text{FP}} \tag{2}$$

- Recall – Probability of crash detection; that is, the percentage of crash cases successfully predicted, calculated as the number of crash cases predicted correctly divided by the total number of crash cases.

$$\text{Recall} = \frac{\text{Correctly Predicted Crashes}}{\text{Total Crashes}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

- F1-score – Harmonic mean of precision and recall; the highest possible value of F-score is 1 (or 100%), indicating perfect precision and recall, and the lowest possible value of F-score is 0, if either precision or recall is zero.

$$\text{F1} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}} \tag{4}$$

- Accuracy – Percentage of true predictions including both true positive and true negative prediction, calculated as the sum of TP and TN divided by the total population.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{5}$$

It is worth noting that Precision (or FAR assessment), Recall, and F-score were more critical measures than accuracy in this study. This is because, as stated earlier, accuracy is the percentage of true predictions. A large proportion of true predictions used in accuracy calculation is predicting no crashes (true negative predictions meaning no alarms triggered when there are no crashes). This accuracy prediction of no crashes dilutes the value of the accuracy measure since this is not the objective of implementing the product.

4.5.2 Evaluation Procedure

Figure 43 presents the procedure for evaluating the selected dynamic crash prediction system based on the performance metrics listed in the previous section. Crash status (Crash/No Crash) was indicated for each road segment and for each testing three-hour period (AM, MD, PM, Night). Traffic (sensor data) and crash (SignalFour data) input files were prepared for each nine-hour period before each three-hour testing period. Each pair of input files was uploaded to the WayCare offline model to create a prediction, and a prediction output file was then downloaded to get the prediction results for a three-hour period at a study site. The steps of creating a prediction and downloading the output file were repeated until all prediction output files were retrieved. The prediction results were then compared with the crash statuses identified, as noted previously. The comparison results were used to assess the prediction quality based on the four assessment types (TP, FP, FN, TN), as reflected by the used performance metrics (Precision, Recall, F-score, Accuracy), as defined in the previous section.

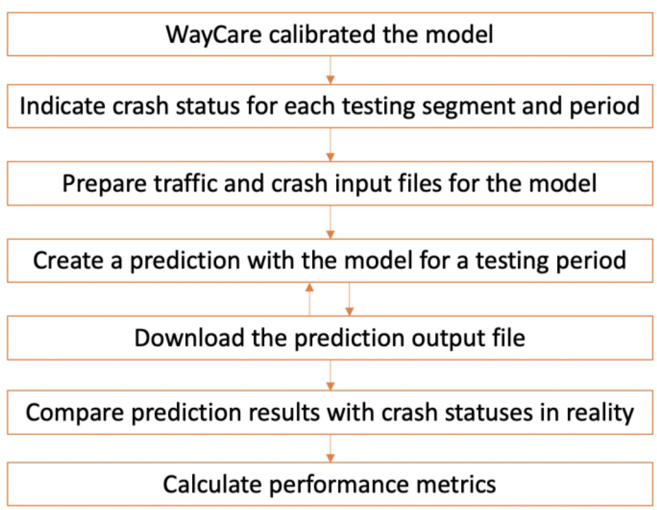


Figure 43. Flow Chart of Evaluation Procedure

The model was first evaluated with two months of data (January and February 2020). To determine if the model was overfitted to the calibration, the test was also done for a month (November 2017) included in the provided calibration data. Another month (July 2020) was also tested to check the effect of the COVID-19 pandemic on the evaluation. As there was no significant difference in performance between the first two months (January and February 2020) and July 2020, they were combined to get larger samples, and the results of the combined three months (January, February, July 2020) are also presented in this study. Therefore, the following four test periods were used in the analysis:

- January and February 2020 – Test using data not included in the calibration without pandemic effect
- November 2017 – Test using data already used in calibration of model
- July 2020 – Test using data from a post-pandemic month
- January, February, July 2020 – Test using combined data not used in calibration

To evaluate the performance of the model during different periods in a day, the analysis was also performed for the following periods:

- AM: 6:00 AM–9:00 AM
- MD: 12:00 PM–3:00 PM
- PM: 3:00 PM–6:00 PM
- Night: 9:00 PM–12:00 AM
- ALL: combination of the four time periods listed above

4.6 Offline Test Results for I-95

Figure 44 to Figure 47 show the evaluation results for the I-95 site for the four test periods—January and February 2020, November 2017, July 2020, and January, February, July 2020. Detailed information is given in Table 25. It is worth noting that the model could not produce all the prediction output files successfully (for all tested three-hour periods) for the I-95 site. The missing output files could not be downloaded because they were always in "in-progress" status instead of "finished." For example, for the combined three-month (January, February, July 2020), the model run only 42.2% (1229 of 2912) of the prediction time intervals successfully; however, the samples were enough for evaluation of the I-95 site. It is not clear why the model was not able to produce the prediction for all test intervals and this may be a software issue.

The overall F-score (for time period ALL) for January and February 2020 was 23% with a Precision of 17% and Recall of 36%, as shown in Figure 44. The value based on the month used in calibration (November 2017) was 28% with a Precision of 29% and Recall of 26%, as shown in Figure 45. This indicates that the performance with the data used in the calibration was not significantly better than the performance with the data not used in the calibration, indicating that there was no overfitting issue. The overall F-score for July 2020 was 18% with Precision of 18% and Recall of 18%, as shown in Figure 46. This performance was lower than that of January and February 2020. Finally, the combined three-month results, as shown in Figure 49, indicate an overall F-score of 20% with Precision of 17% and Recall 25%; that is, 17% of the alarms triggered by the model were true alarms (there were crashes in reality), and there was a successful prediction of 25% of the true crashes.

The model had poor performance for the AM, MD, and Night periods but much better results for the PM peak. For example, during the combined three-month period (January, February, July 2020), the number of crash cases tested for AM, MD, and Night were 26, 22, and 17, respectively, but the model produced only one alarm for the AM peak (which was incorrect) and another alarm for MD, which made the Precision, Recall, and F-score all 0%. However, for the PM peak, the F-score was 27%, and the Recall was as high as 55%. This indicates that the model possibly over-fit the PM crash data. It may be useful to produce different models for different times of day.

It is worth noting that the results showed high accuracy of predictions (more than 85%) for AM, MD, and Night in all four test periods despite poor Precision, Recall, and F-score. This proved that the accuracy of predictions is not an important metric for this research, as noted earlier.

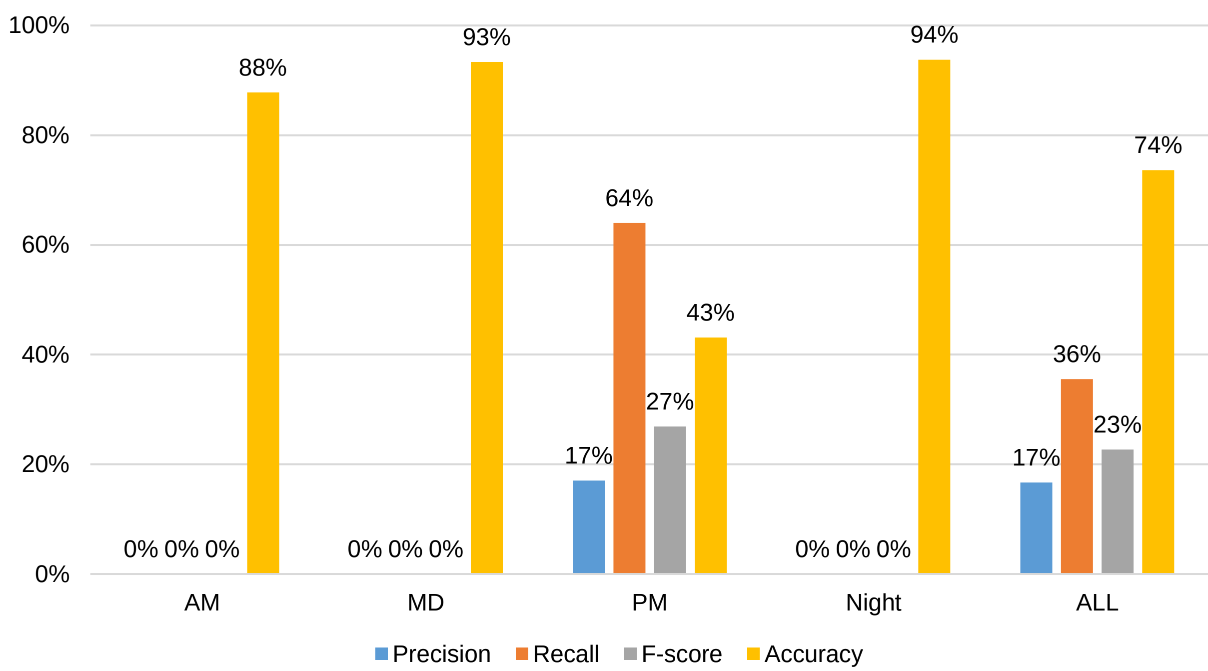


Figure 44. Evaluation Results for January and February 2020 at I-95 Site

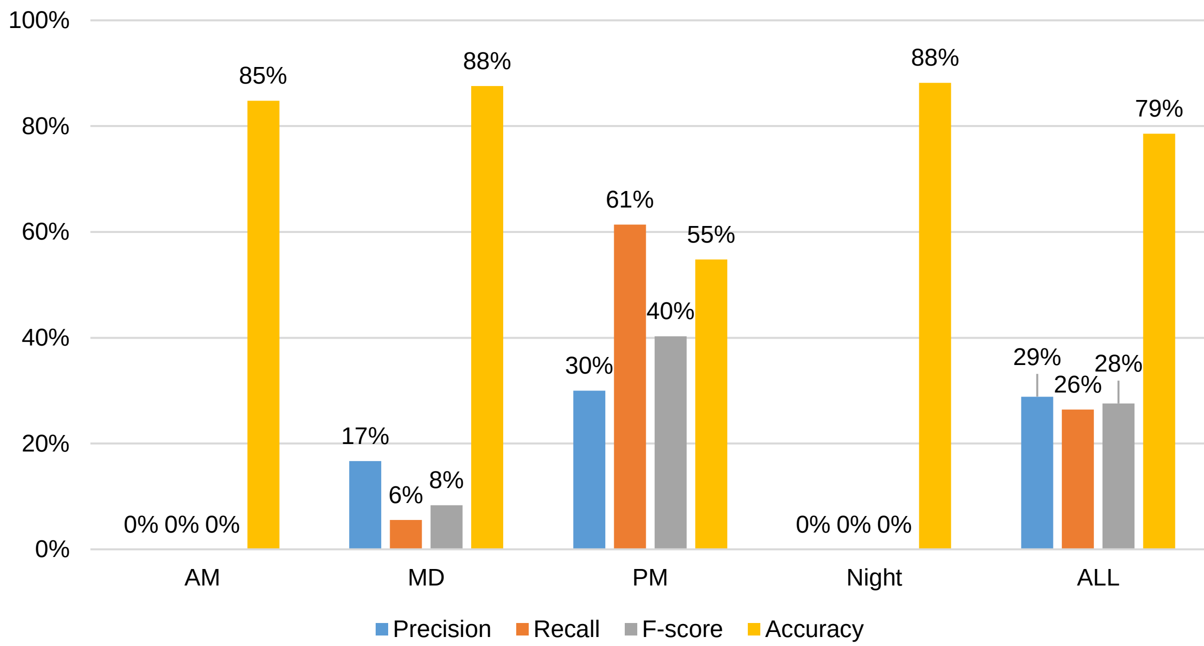


Figure 45. Evaluation Results for November 2017 at I-95 Site

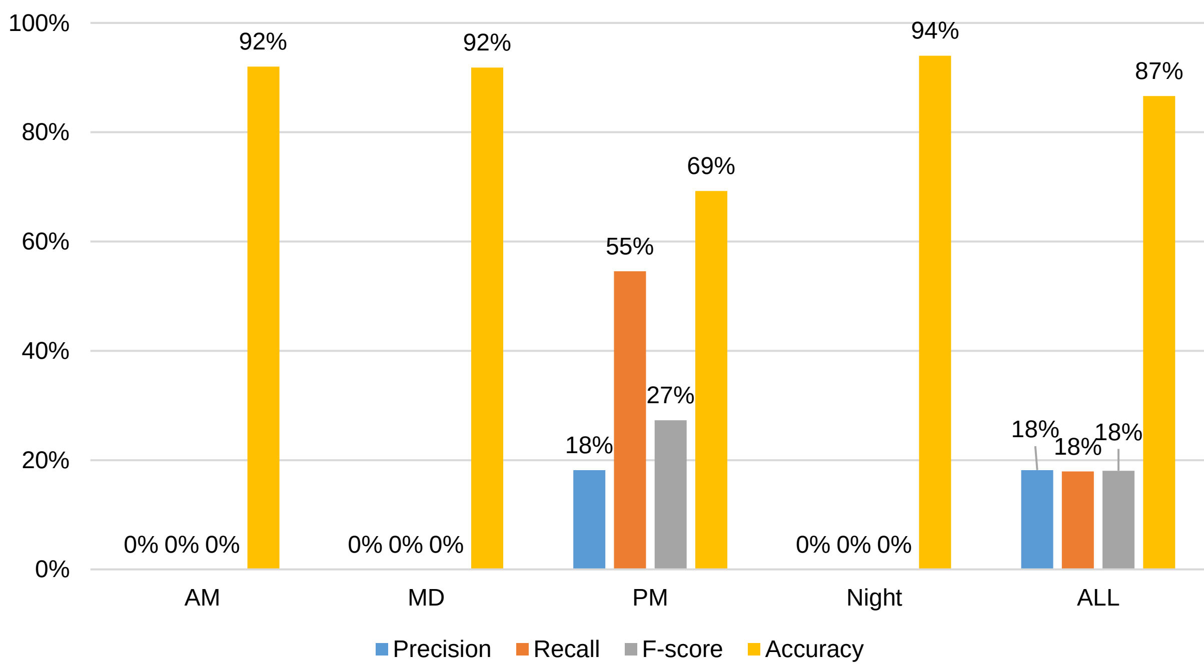


Figure 46. Evaluation Results for July 2020 at I-95 Site

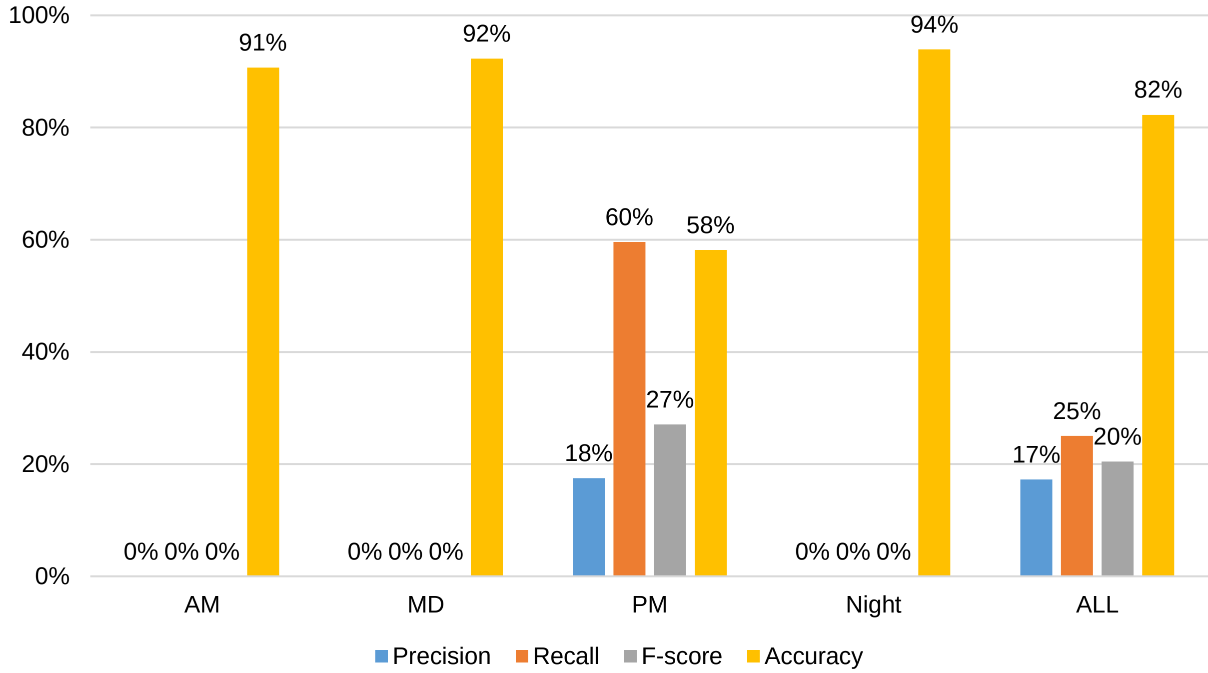


Figure 47. Evaluation Results for January, February, and July 2020 at I-95 Site

Table 25. Summary of Offline Test Results for I-95 Site

Test Period	Time Period	Precision	FAR*	Recall	F-score	Accuracy	TP	FP	TN	FN
November 2017	AM	0%	-	0%	0%	85%	0	0	145	26
	MD	17%	83%	6%	8%	88%	1	5	154	17
	PM	30%	**70%**	**61%**	40%	55%	27	63	70	17
	Night	0%	100%	0%	0%	88%	0	1	142	18
	ALL	**29%**	71%	**26%**	**28%**	**79%**	**28**	**69**	**511**	**78**
January and February 2020	AM	0%	100%	0%	0%	88%	0	1	79	10
	MD	0%	100%	0%	0%	93%	0	1	84	5
	PM	17%	**83%**	**64%**	27%	43%	16	78	50	9
	Night	0%	-	0%	0%	94%	0	0	75	5
	ALL	**17%**	83%	**36%**	**23%**	**74%**	**16**	**80**	**288**	**29**
July 2020	AM	0%	-	0%	0%	92%	0	0	184	16
	MD	0%	-	0%	0%	92%	0	0	191	17
	PM	18%	**82%**	**55%**	27%	69%	12	54	132	10
	Night	0%	-	0%	0%	94%	0	0	188	12
	ALL	**18%**	82%	**18%**	**18%**	**87%**	**12**	**54**	**695**	**55**
January, February, July 2020	AM	0%	100%	0%	0%	91%	0	1	263	26
	MD	0%	100%	0%	0%	92%	0	1	275	22
	PM	18%	**82%**	**60%**	27%	58%	28	132	182	19
	Night	0%	-	0%	0%	94%	0	0	263	17
	ALL	**17%**	**83%**	**25%**	**20%**	**82%**	**28**	**134**	**983**	**84**

*FAR (False Alarm Rate) = 1 – Precision

4.7 Offline Test Results for E Sunrise Blvd

Figure 48 to Figure 51 show the evaluation results of the E Sunrise Blvd site for the four test periods, respectively. Detailed information is given in

Table 26. Different from the I-95 site, the model produced all attempted 1,456 predictions successfully for the E Sunrise Blvd site.

The overall F-score for the January and February 2020 test was only 5% with Precision of 5% and Recall of 6%, as shown in Figure 48. The results based on the data used in the calibration, as shown in Figure 49, indicate a much higher F-score at 14% with Precision of 10% and Recall of 20%. The overall F-score for July 2020 was 13% with Precision of 10% and Recall of 19%, as shown in Figure 50, which proved that the model was not overfitted to the calibration data. Finally, the combined three-month test results shown in Figure 51 indicate an overall F-score of 9% with Precision of 7% and Recall of 11%; that is, 7% of the alarms triggered by the model were true alarms and 11% of the crashes were predicted.

Similar to the results of I-95 site, the E Sunrise Blvd site also had poor performance for the AM and Night periods, although it had much better results for MD and PM peak, especially MD. During the combined three-month period (January, February, July 2020), the number of crash cases tested for the AM and Night periods were 10 and 14, respectively, but the model produced only one alarm for the Night period, which was incorrect, making Precision, Recall, and F-score all 0%. However, for the MD and PM periods, the values of the F-score were 15% and 6%, respectively.

The E Sunrise Blvd site also had a high accuracy of predictions, especially for the AM and Night periods, but it was also because most of the true predictions were true negative predictions.

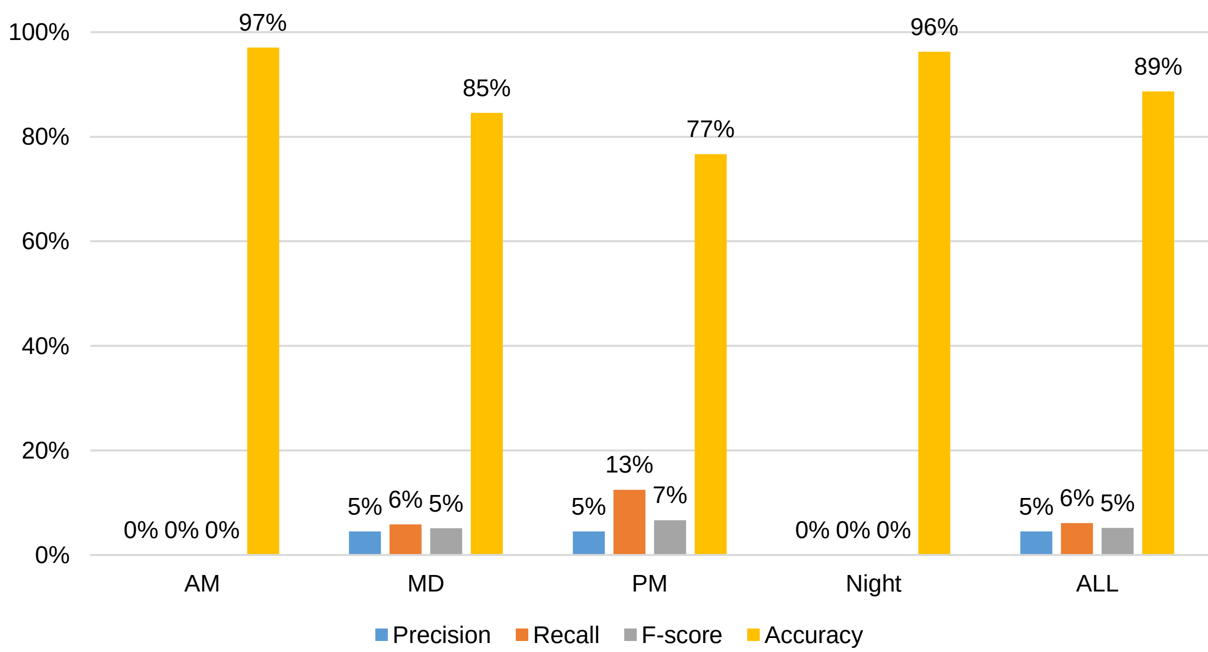


Figure 48. Evaluation Results for January and February 2020 at E Sunrise Blvd Site

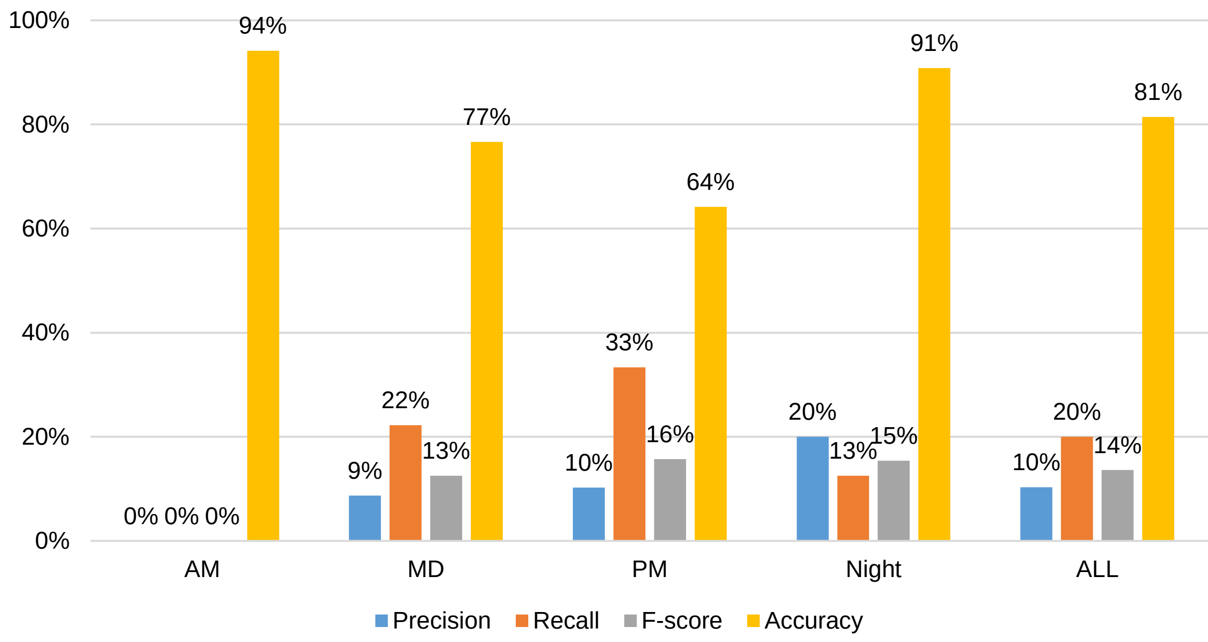


Figure 49. Evaluation Results for November 2017 at E Sunrise Blvd Site

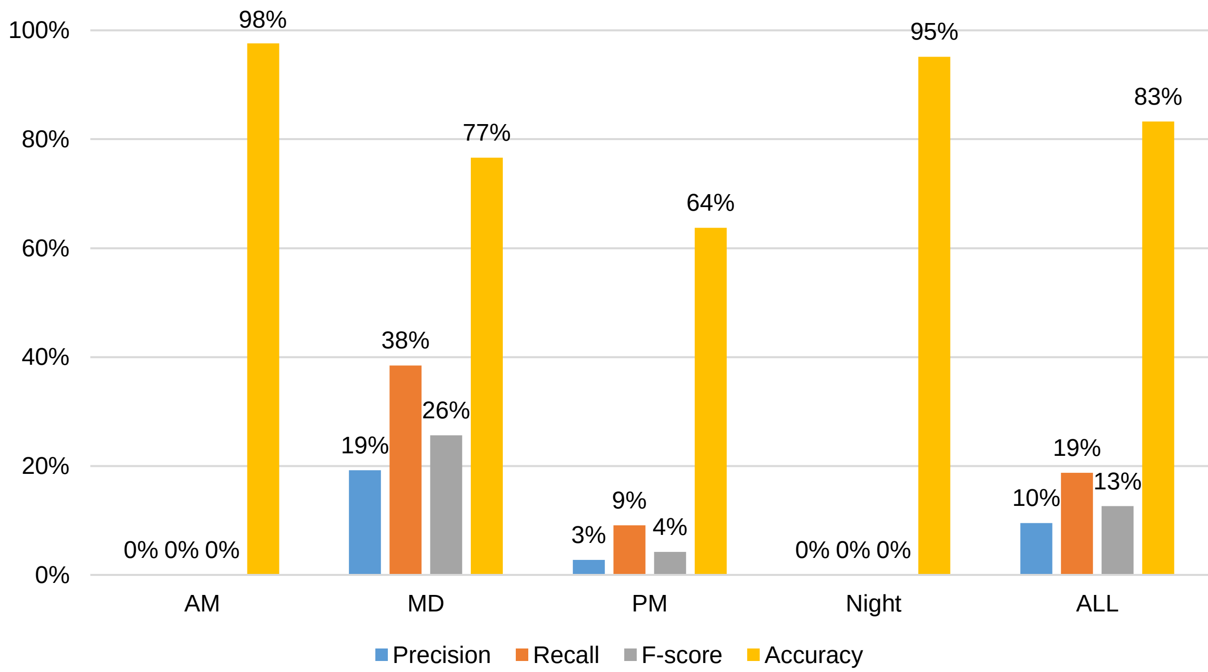


Figure 50. Evaluation Results for July 2020 at E Sunrise Blvd Site

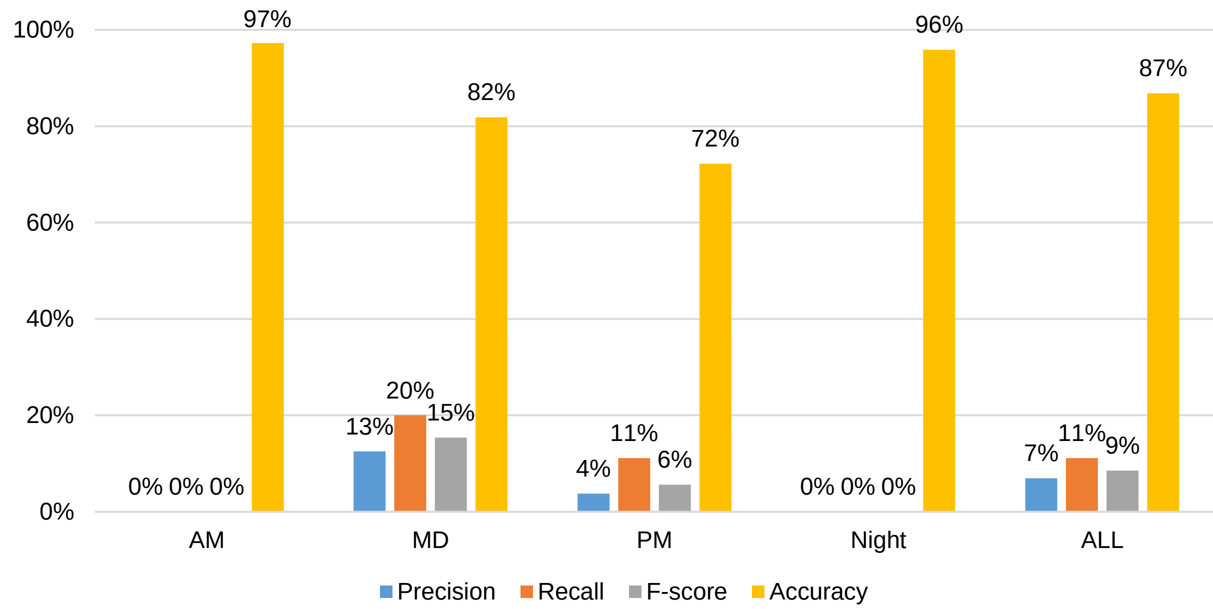


Figure 51. Evaluation Results for January, February, July 2020 at E Sunrise Blvd Site

Table 26. Summary of Offline Test Results for E Sunrise Blvd Site

Test Period	Time Period	Precision	FAR*	Recall	F-score	Accuracy	TP	FP	TN	FN
November 2017	AM	0%	100%	0%	0%	94%	0	1	113	6
	MD	9%	**91%**	**22%**	13%	77%	2	21	90	7
	PM	10%	**90%**	**33%**	16%	64%	4	35	73	8
	Night	20%	80%	13%	15%	91%	1	4	108	7
	ALL	**10%**	**90%**	**20%**	**14%**	**81%**	**7**	**61**	**384**	**28**
January and February 2020	AM	0%	-	0%	0%	97%	0	0	233	7
	MD	5%	95%	6%	5%	85%	1	21	202	16
	PM	5%	95%	13%	7%	77%	2	42	182	14
	Night	0%	-	0%	0%	96%	0	0	231	9
	ALL	**5%**	**95%**	**6%**	**5%**	**89%**	**3**	**63**	**848**	**46**
July 2020	AM	0%	-	0%	0%	98%	0	0	121	3
	MD	19%	81%	**38%**	26%	77%	5	21	90	8
	PM	3%	97%	9%	4%	64%	1	35	78	10
	Night	0%	100%	0%	0%	95%	0	1	118	5
	ALL	**10%**	**90%**	**19%**	**13%**	**83%**	**6**	**57**	**407**	**26**
January, February, July 2020	AM	0%	-	0%	0%	97%	0	0	354	10
	MD	13%	87%	**20%**	15%	82%	6	42	292	24
	PM	4%	96%	11%	6%	72%	3	77	260	24
	Night	0%	100%	0%	0%	96%	0	1	349	14
	ALL	**7%**	**93%**	**11%**	**9%**	**87%**	**9**	**120**	**1255**	**72**

*FAR (False Alarm Rate) = 1 – Precision

4.8 Discussions and Conclusions

To facilitate discussion in this section, the research team focused on using two major performance measures to evaluate the WayCare model:

- Recall – Percentage of crash events that can be predicted by the model; higher is better.
- Precision – Percentage of alarms (crash prediction) that are true; higher is better. False Alarm Rate (FAR) is the flip side of Precision (=1-Precision), which is defined as the percentage of alarms that are false; lower is better.

These two measures are institutive and straightforward to non-machine learning engineers. To keep constant with machine-learning terms, the research team also provided F-score and Accuracy for reference. Based on the offline testing results, findings were as follows:

- The WayCare model presented better performance for the I-95 site than the E Sunrise Blvd site for Recall (25% vs. 11%) and FAR (83% vs. 93%). The high number of crashes and relative simplicity of traffic patterns on the freeway may explain why the WayCare model worked better on I-95.
- The WayCare model presents varying Recall performances by period for both I-95 and E Sunrise Blvd.

   - On the I-95 sites, the WayCare model presented “good” performance for the PM period (3:00 PM–6:00 PM); 55–65% of crashes could be predicted for different months. These performances were close to WayCare’s evaluation based on historical data for 2015–2019 (54% of crashes could be predicted for I-95, on average, without distinguishing periods), as shown in Appendix A.
   - The WayCare model had “poor” performance on I-95 for the MD and Night periods. The model outputs could not predict any crashes in most scenarios for these periods except for the MD period in July 2017 (6% of crashes can be predicted).
   - It is worth noting that the Recall performance for the PM periods in July 2020 (55% of crashes predicted) was lower than those for January and February 2020 (64% of crashes predicted) and July 2017 (61%). This comparison may imply that the COVID-19 pandemic event had an impact on model performance (Recall reduction of 6–9%) on I-95.
   - For E Sunrise Blvd., the WayCare model presented relatively “better” performance for MD (12:00 PM–3:00 PM) and PM (3:00 PM–6:00 PM). Based on 2020 data, an average 20% of MD crashes and 11% of PM crashes could be predicted. It was interesting to find that the model had better performance in July than in January or February 2020, which is the opposite of the finding for I-95.
- FARs were relatively high ($\geqslant$ 70%) across scenarios (Precision was relatively low, $\leq$ 30%). This implies that 70% (or higher) of alarms were not actually associated with a crash. The possible causes are:
   - Underreported crashes – some minor crashes tend to not be reported to police and thus are not included in the crash database but can be predicted by the WayCare model.
   - Near-crash events – some near-crash events, such as serious conflicts, are high-risk events but do not necessarily result in crashes. Prediction of these near-crash events are useful to apply actions to prevent risky situations.

As no data for underreported crashes and near-crash events were available, it was impossible to estimate a “true” false alarm rate. However, WayCare reported that, typically, in-vehicle data show that the WayCare model can predict 20–30% more crashes that are not documented.

5 Recommendations

Based on the pilot study results, the research team developed recommendations for implementing dynamic crash predictions in Florida as well as implementation recommendations and action plan.

5.1 Implementation Recommendations

Based on the pilot study results, the following recommendations for implementing dynamic crash prediction were developed:

- Implement the dynamic prediction model preferably on freeways but work with WayCare to improve model performance for periods other than the PM period considering the following:
  - Model produced good performance for the PM period (3:00–6:00 PM) on the tested freeway section (correctly predicted 60% of crash cases).
  - Local resources for model data input and crash prevention (i.e., traffic sensors, ITS/T&SMO actions, etc.) are plentiful on interstates.
  - Freeways experience high traffic volumes and excessive crash frequencies compared to other road facilities; implementation of dynamic crash prediction could bring significant safety and mobility benefits.
- Consider implementation of the dynamic prediction model on arterials but work with WayCare to improve model performance for periods other than the MD and PM periods, if traffic agencies have a high need for arterial safety management, considering the following:
  - Model showed "positive" performance for the two periods on arterials (correctly predicting 11–20% of crash cases).
  - Relatively high volumes and crash frequencies on major arterials introduce the need for dynamic crash prediction and prevention; arterials have more complex traffic patterns.
  - Traffic agencies should decide on implementation based on their arterial safety management goals and needs.
- Real-time implementation of the model at TMCs will require maintaining traffic and crash/incident data for the past nine hours to predict crash rates for the next three-hour prediction window. The time interval of traffic sensor data is suggested to be 20 sec or 1 min. Longer time intervals can be applied; however, they may reduce prediction performance. The protocol for data transfer between TMC SunGuide software and databases and the WayCare web platform needs to be addressed.

5.2 Suggested Crash Prevention Actions

The WayCare model outputs an alarm if it predicts a high crash risk for a three-hour time window. The alarm allows TMCs to activate actions to reduce crash risk prior to crash occurrence within these time windows. As discussed in Chapters 2 and 3, previous studies indicate that crashes are highly related to driving speed and speed variation. Thus, an effective strategy for preventing crashes is to reduce and homogenize running speeds on roadway segments with predicted crash alarms. In addition, sharing the predicted crash risk with incident response services (i.e., Road Rangers) can reduce response time to potential crash events and, consequently, mitigate the risk of secondary crashes and non-recurring congestion. The research team proposes three actions to respond to predicted crash alarms considering the following factors:

- *Effectiveness* – Crash prevention should realize any one of the two safety strategies—speed management or information-sharing—and can theoretically improve safety and mobility.
- *Availability* – Crash prevention actions should be widely implemented on Florida interstates and arterials.
- *Experience* – Crash prevention actions should be tested with the dynamic crash prediction in pilot studies; suggestions from vendors and/or current users are also considered.
- *Proved benefits* – Qualified and/or quantitative safety and mobility benefits of the proposed prevention actions can be found in the literature.

5.2.1 Dynamic Message Signs (DMS)

DMS are widely implemented ITS/T&SMO devices on Florida interstates and major arterials. They display dynamic messages to warn drivers about special events such as traffic congestion, crashes, incidents, AMBER/Silver/Blue alerts, or work zones. As the operation cost of DMSs is relatively low (TMCs directly operate DMSs), it is suggested to display safety messages on DMSs upon receiving a predicted alarm for a three-hour time window. The suggested warning message would be "Reduce Your Speed." If law enforcement action is activated, the warning messages would be "Reduce Your Speed" + "Police Ahead."

5.2.2 Law Enforcement

Law enforcement activities regulate driving behaviors, especially driver speed choice, and implementation requires cooperation from law enforcement agencies (Florida Highway Patrol on interstates and County Sheriff on major arterials). Suggested law enforcement actions include a stationary police car on the roadside with flashing blue lights or a patrolling police car along the alarmed segment. Considering that the FAR of the prediction model is relatively high ($\geq$ 80%), the following factors could be considered to decide whether to apply law enforcement actions and their duration:

- If there are crashes occurring within the nine hours prior to the three-hour time window, apply and keep law enforcement activation for three hours.
- If there are no crashes occurring within nine hours prior to the three-hour time window, apply law enforcement as an option or shorten its activation duration.

5.2.3 *Incident Response Vehicles*

Increasing the patrolling frequency of incident response vehicles on alarmed segments can reduce reaction time. TMCs should share a predicted alarm with Road Rangers or other incident response services. However, a study of optimal patrolling scheduling is needed.

Table 27. Summary of Recommended Crash Prevention Actions

	DMS	Law Enforcement	Incident Management
Actions	• If no law enforcement, "Reduce Your Speed" • If with law enforcement, "Reduce Your Speed" +"Police Ahead"	• Stationary police cars with blue lights on roadside, or • Patrolling police cars	• Increase patrolling frequencies of incident response vehicles
Activation Criterions	• Prediction alarm	• Prediction Alarm, or • Prediction Alarm + Crash records for past nine hours	• Prediction alarm
Activation Duration	• Three-hour time window	• Three-hour time window or less	• Three-hour time window or less
Agencies	• TMC	• FHP (for interstates) • County Sheriff (for arterials)	• Road Ranger (for interstates)
Qualified Safety Benefits	• Alert drivers to reduce speed • Prevent primary and secondary crashes	• Alert drivers to reduce speed • Prevent primary and secondary crashes	• Reduce emergency vehicle response time • Prevent secondary crashes
Qualified Mobility benefits	• Reduce risk of non-recurring congestion caused by crashes	• Reduce risk of non-recurring congestions caused by crashes	• Reduce risk of non-recurring congestion caused by crashes
Quantitative Safety Benefits	• 17% crash reduction on interstates combing DMS and stationary police cars[1]	• 9% of crash reduction (presence of stationary police car)[2] • 17% crash reduction on freeway combining DMS and stationary cars[1]	• Unavailable

[1] WayCare report for Las Vegas pilot study.
[2] Sarit Weisburd, "The Effect of Police Patrol on Car Accidents," Master's thesis, 2013.

6 Summary and Conclusions

6.1 Summary

FDOT's Safety and TSM&O programs have been collecting, archiving, and analyzing a wide range of traffic, crash, event, and other data to improve congestion and safety on the SHS. Dynamic crash prediction, a proactive safety management strategy, predicts crash risk based on prevailing traffic conditions and applies crash prevention actions to prevent crashes before occurrence. As an innovative technology, dynamic crash prediction provides a potential way for FDOT to take advantage of information provided by ITS devices and other sources, combined with increasingly available big data/data analytics to effectively prevent crash occurrence and improve safety and mobility on Florida roadway systems. Although many Florida traffic agencies have shown interest in dynamic crash prediction methods and have plans to implement them, there is no clear understanding on the applicability of dynamic crash prediction in performance, implementability, integrability, and impacts.

This project aimed to evaluate existing dynamic crash prediction methods and practices related to accuracy and timeliness, use in TMCs, and impacts on safety and mobility for implementing a proactive safety strategy in Florida. To achieve this goal, the following tasks were completed:

- A comprehensive literature review was conducted to summarize previous studies on the following:
    - Theoretical framework of dynamic crash prediction
    - Data needs for dynamic crash prediction
    - Spatial and temporal resolution for dynamic crash prediction
    - Modeling methodologies
    - Accuracy and timeliness
- Existing vendors providing dynamic crash prediction functions and existing users that have implemented the functions were identified. Through interviews and document review, the research team developed an understanding of the art-of-practice of dynamic crash prediction.
- The research team evaluated the identified dynamic crash prediction technologies/ platforms in functionality, performance and impacts, data and local resource needs, usability, and maturity. Based on the evaluation, one technology (WayCare) was selected for the pilot study.
- The research team conducted a pilot study with the selected technology in FDOT District 4. First, the research team collected historical traffic and crash data for five years (2015–2019) on two study sites (I-95 and E Sunrise Blvd). This dataset was provided to WayCare for model calibration. With the calibrated model, the research team conducted offline tests using 2020 data for three months (January, February, July). The performance

of dynamic crash prediction in the Florida roadway environment, including interstates and major arterials, was evaluated from the offline test.

- The research team developed recommendations for implementing dynamic crash predictions in Florida based on pilot study results and vendor evaluation. Crash prevention actions were also suggested.

6.2 Conclusions

Major conclusions from this study are as follows:

- Although many academic papers have explored dynamic crash prediction, including data needs, sampling methods, algorithm/models, and performance, only a limited number of vendors and technologies were found on the current market. Only one vendor (WayCare) was found to provide relatively mature and integral commercial solutions.
- No current users were found to implement dynamic crash prediction systems, although many agencies have shown interest. Only the Nevada Department of Transportation (NDOT) tested the functions in a pilot study. Two local agencies in Florida (City of Tampa, Pinellas County) have implemented the WayCare system; however, they do not apply the dynamic crash prediction function.
- Only one pilot study (NDOT) was found to test the dynamic crash prediction system in a real roadway environment. It produced some preliminary evaluation results and proved the concepts of dynamic crash predictions; however, it lacked an independent assessment from a third party and a comprehensive evaluation report.
- With Florida data, the WayCare technology can predict crash risk for a three-hour time window based on nine-hour traffic and crash information prior to the prediction. The traffic data include speed, volume, and occupancy. If the predicted crash risk is higher than a threshold, an alarm will be produced. However, the threshold is non-configurable on the WayCare platform.
- The pilot study showed that the WayCare model presents better prediction performance on freeway segments than on arterials due to the relative simplicity of traffic patterns on freeways. The WayCare model calibrated in this study exhibited various performances by time windows for either freeway segments or arterials.
- It is suggested to implement the dynamic prediction model preferably on freeways because the model produced good performance for the PM period (3:00–6:00 PM) on the tested freeway section (correctly predicted 60% of crash cases).
- It is suggested to work with WayCare to improve its model performance for periods other than PM on freeways and all periods on arterials. As a data-driven method, dynamic crash prediction requires more data to upgrade prediction ability.

- Three crash prediction actions—DMS safety messages, stationary police cars with flashing lights, and advance warning to Road Rangers—were proposed based on WayCare's experience and the availability of TSM&O applications in FDOT District 4. A further study is needed to address the safety and mobility of these crash prediction actions.
- The WayCare system (and other systems) is hosted on a cloud platform that does not need special implementation. However, a data connection is needed to feed real-time data from TMCs to the WayCare platform. This data connection should be addressed in follow-up studies considering security and reliability.

References

1. FDOT. *RCI Features & Characteristics Handbook*. 2017.

2. FHWA Road Safety Audit Guidelines. https://safety.fhwa.dot.gov/rsa/guidelines/appendix_a.htm. Accessed Jan. 2, 2018.

3. Lord, D., A. Manar, and A. Vizioli. Modeling Crash-Flow-Density and Crash-Flow-V/C Ratio Relationships for Rural and Urban Freeway Segments. *Accident Analysis & Prevention*, Vol. 37, No. 1, 2005, pp. 185–199.

4. Wang, X., Q. Zhou, M. Quddus, and T. Fan. Speed, Speed Variation and Crash Relationships for Urban Arterials. *Accident Analysis & Prevention*, Vol. 113, 2018, pp. 236–243.

5. Elvik, R. A Re-Parameterisation of the Power Model of the Relationship between the Speed of Traffic and the Number of Accidents and Accident Victims. *Accident Analysis & Prevention*, Vol. 50, 2013, pp. 854–860.

6. Nilsson, G. The Effects of Speed Limits on Traffic Crashes in Sweden. *Proceedings of the International Symposium on the Effects of Speed Limits on Traffic Crashes and Fuel Consumption*, Dublin. Paris: Organisation for Economy, Co-Operation, and Development (OECD). 1982.

7. Finch, D. J., P. Kompfner, C. R. Lockwood, and G. Maycock. Speed, Speed Limits and Crashes. *Project Record S211G/RB/Project Report PR*, Vol. 58, 1994.

8. AASHTO. *Highway Safety Manual*. Washington, DC, 2010.

9. Garber, N. J., and R. Gadiraju. Factors Affecting Speed Variance and Its Influence on Accidents. *Transportation Research Record*, Vol. 1213, 1989, pp. 64–71.

10. Oh, C., J.-S. Oh, S. Ritchie, and M. Chang. Real-Time Estimation of Freeway Accident Likelihood. 2001.

11. Quddus, M. Exploring the Relationship between Average Speed, Speed Variation, and Accident Rates Using Spatial Statistical Models and GIS. *Journal of Transportation Safety & Security*, Vol. 5, No. 1, 2013, pp. 27–45.

12. Lum, H., and J. A. Reagan. Interactive Highway Safety Design Model: Accident Predictive Module. *Public Roads*, Vol. 58, No. 3, 1995.

13. Roshandel, S., Z. Zheng, and S. Washington. Impact of Real-Time Traffic Characteristics on Freeway Crash Occurrence: Systematic Review and Meta-Analysis. *Accident Analysis & Prevention*, Vol. 79, 2015, pp. 198–211.

14. Basso, F., L. J. Basso, F. Bravo, and R. Pezoa. Real-Time Crash Prediction in an Urban Expressway Using Disaggregated Data. *Transportation Research Part C: Emerging Technologies*, Vol. 86, 2018, pp. 202–219.

15. Hossain, M., and Y. Muromachi. A Real-Time Crash Prediction Model for the Ramp Vicinities of Urban Expressways. *IATSS Research*, Vol. 37, No. 1, 2013, pp. 68–79.

16. Yu, R., and M. Abdel-Aty. Multi-Level Bayesian Analyses for Single- and Multi-Vehicle Freeway Crashes. *Accident Analysis & Prevention*, Vol. 58, 2013, pp. 97–105. https://doi.org/10.1016/j.aap.2013.04.025.

17. Sun, J., and J. Sun. A Dynamic Bayesian Network Model for Real-Time Crash Prediction Using Traffic Speed Conditions Data. *Transportation Research Part C: Emerging Technologies*, Vol. 54, 2015, pp. 176–186.

18. Hossain, M., and Y. Muromachi. A Bayesian Network Based Framework for Real-Time Crash Prediction on the Basic Freeway Segments of Urban Expressways. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 373–381.

19. Lee, C., B. Hellinga, and F. Saccomanno. Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic. *Transportation Research Record*, No. 1840, 2003, pp. 67–77.

20. Oh, J.-S., C. Oh, S. G. Ritchie, and M. Chang. Real-Time Estimation of Accident Likelihood for Safety Enhancement. *Journal of Transportation Engineering*, Vol. 131, No. 5, 2005, pp. 358–363.

21. Xu, C., W. Wang, P. Liu, and F. Zhang. Development of a Real-Time Crash Risk Prediction Model Incorporating the Various Crash Mechanisms across Different Traffic States. *Traffic Injury Prevention*, Vol. 16, No. 1, 2015, pp. 28–35.

22. Xu, C., W. Wang, and P. Liu. A Genetic Programming Model for Real-Time Crash Prediction on Freeways. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 2, 2013, pp. 574–586.

23. Sun, J., J. Sun, and P. Chen. Use of Support Vector Machine Models for Real-Time Prediction of Crash Risk on Urban Expressways. *Transportation Research Record*, No. 2432, 2014, pp. 91–98.

24. Pande, A., and M. Abdel-Aty. Comprehensive Analysis of the Relationship between Real-Time Traffic Surveillance Data and Rear-End Crashes on Freeways. *Transportation Research Record*, No. 1953, 2006, pp. 31–40.

25. Ahmed, M. M., and M. A. Abdel-Aty. The Viability of Using Automatic Vehicle Identification Data for Real-Time Crash Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 2, 2012, pp. 459–468.

26. Abdel-Aty, M. A., H. M. Hassan, M. Ahmed, and A. S. Al-Ghamdi. Real-Time Prediction of Visibility Related Crashes. *Transportation Research Part C: Emerging Technologies*, Vol. 24, 2012, pp. 288–298.

27. Zhang, K., and M. A. P. Taylor. Towards Universal Freeway Incident Detection Algorithms. *Transportation Research Part C: Emerging Technologies*, Vol. 14, No. 2, 2006, pp. 68–80.

28. Golob, T. F., and W. W. Recker. A Method for Relating Type of Crash to Traffic Flow Characteristics on Urban Freeways. *Transportation Research Part A: Policy and Practice*, Vol. 38, No. 1, 2004, pp. 53–80.

29. Golob, T. F., W. Recker, and Y. Pavlis. Probabilistic Models of Freeway Safety Performance Using Traffic Flow Data as Predictors. *Safety Science*, Vol. 46, No. 9, 2008, pp. 1306–1333.

30. Oh, C., J.-S. Oh, and S. G. Ritchie. Real-Time Hazardous Traffic Condition Warning System: Framework and Evaluation. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 6, No. 3, 2005, pp. 265–272.

31. Pande, A., and M. Abdel-Aty. Multiple-Model Framework for Assessment of Real-Time Crash Risk. *Transportation Research Record*, No. 2019, 2007, pp. 99–107.

32. Abdel-Aty, M., N. Uddin, and A. Pande. Split Models for Predicting Multivehicle Crashes during High-Speed and Low-Speed Operating Conditions on Freeways. *Transportation Research Record*, No. 1908, 2005, pp. 51–58.

33. Abdel-Aty, M., R. Cunningham, V. Gayah, and L. Hsia. Dynamic Variable Speed Limit Strategies for Real-Time Crash Risk Reduction on Freeways. *Transportation Research Record*, No. 2078, 2008, pp. 108–116.

34. Pande, A., A. Das, M. Abdel-Aty, and H. Hassan. Estimation of Real-Time Crash Risk: Are All Freeways Created Equal? *Transportation Research Record*, No. 2237, 2011, pp. 60–66.

35. Abdel-Aty, M. A., and R. Pemmanaboina. Calibrating a Real-Time Traffic Crash-Prediction Model Using Archived Weather and ITS Traffic Data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7, No. 2, 2006, pp. 167–174.

36. Abdel-Aty, M., and A. Pande. Identifying Crash Propensity Using Specific Traffic Speed Conditions. *Journal of Safety Research*, Vol. 36, No. 1, 2005, pp. 97–108.

37. Al-Deek, H., C. Venkata, and S. Ravi Chandra. New Algorithms for Filtering and Imputation of Real-Time and Archived Dual-Loop Detector Data in I-4 Data Warehouse. *Transportation Research Record*, No. 1867, 2004, pp. 116–126.

38. Xu, C., P. Liu, W. Wang, and X. Jiang. Development of a Crash Risk Index to Identify Real Time Crash Risks on Freeways. *KSCE Journal of Civil Engineering*, Vol. 17, No. 7, 2013, pp. 1788–1797.

39. Mussone, L., A. Ferrari, and M. Oneta. An Analysis of Urban Collisions Using an Artificial Intelligence Model. *Accident Analysis & Prevention*, Vol. 31, No. 6, 1999, pp. 705–718.

40. Abdel-Aty, M., A. Pande, C. Lee, V. Gayah, and C. Dos Santos. Crash Risk Assessment Using Intelligent Transportation Systems Data and Real-Time Intervention Strategies to Improve Safety on Freeways. *Journal of Intelligent Transportation Systems*, Vol. 11, No. 3, 2007, pp. 107–120.

41. Abdel-Aty, M., and F. Abdalla. Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes: Generalized Estimating Equations for Correlated Data. *Transportation Research Record*, No. 1897, 2004, pp. 106–115.

42. Shi, Q., and M. Abdel-Aty. Big Data Applications in Real-Time Traffic Operation and Safety Monitoring and Improvement on Urban Expressways. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 380–394.

43. Lee, C., M. Abdel-Aty, and L. Hsia. Potential Real-Time Indicators of Sideswipe Crashes on Freeways. *Transportation Research Record*, No. 1953, 2006, pp. 41–49.

44. Quddus, M. A., C. Wang, and S. G. Ison. Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models. *Journal of Transportation Engineering*, Vol. 136, No. 5, 2009, pp. 424–435.

45. Amini, S., I. Gerostathopoulos, and C. Prehofer. Big Data Analytics Architecture for Real-Time Traffic Control. 2017.

46. St-Aubin, P., N. Saunier, and L. Miranda-Moreno. Large-Scale Automated Proactive Road Safety Analysis Using Video Data. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 363–379.

47. Ahmed, M., M. Abdel-Aty, and R. Yu. Bayesian Updating Approach for Real-Time Safety Evaluation with Automatic Vehicle Identification Data. *Transportation Research Record*, No. 2280, 2012, pp. 60–67.

48. Pirdavani, A., M. Magis, E. De Pauw, and S. Daniels. Real-Time Crash Risk Prediction Models Using Loop Detector Data for Dynamic Safety Management System Applications. 2014.

49. Xu, C., A. P. Tarko, W. Wang, and P. Liu. Predicting Crash Likelihood and Severity on Freeways with Real-Time Loop Detector Data. *Accident Analysis & Prevention*, Vol. 57, 2013, pp. 30–39.

50. You, J., J. Wang, and J. Guo. Real-Time Crash Prediction on Freeways Using Data Mining and Emerging Techniques. *Journal of Modern Transportation*, Vol. 25, No. 2, 2017, pp. 116–123.

Appendix A: Literature Review Matrix

Ref.	Title	Input Parameters & Aggregation Time	Methodology and Model	Accuracy	Time Lines	Sensor Type	Test Bed	Facility Type
(*14*)	Real-Time Crash Prediction in Urban Expressway Using Disaggregated Data	Traffic data • Flow of vehicles • Density of vehicles • Mean speed • Std deviation speed	• Support Vector Machines + Logistic Regression • Random Forest (Parameter Selection)	Detection Rate • 67.89% False • 20.94%	5 min	Automatic Vehicle Identification	Simulation	• Freeway • Santiago, Chile
(*17*)	A Dynamic Bayesian Network Model for Real-Time Crash Prediction Using Traffic Speed Conditions Data	Traffic data • Flow of vehicles • Input uses 9 different combinations of mean speed	• Dynamic Bayesian Network	Success Rate • 76.40% False • 23.70%	5–20 min	Dual Loop Detectors	Simulation	• Freeway • Shanghai
(*18*)	A Bayesian Network-Based Framework for Real-Time Crash Prediction on the Basic Freeway Segments Of Urban Expressways	Traffic data • No. of heavy vehicle count • Speed • Avg speed • Avg occupancy	• Bayesian Belief Net (BBN) • Random multinomial Logit (RMNL)	Success Rate • 66% False • 20%	4–9 min	Loop Detectors	Simulation	Type of Road • Freeway w/o ramps • 250 meters section Place • Tokyo
(*19*)	Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic	Traffic data • Traffic density • Speed • Avg variation of speed on each lane • Avg variation of speed difference across adjacent lanes	• Aggregated Log Linear model		20 min	Loop Detectors	Simulation	Type of Road • Freeway Place • Toronto
(*20*)	Real-Time Estimation of Accident Likelihood for Safety Enhancement	Traffic data • Flow • Occupancy • Speed	• Epanechnikov Kernel Function	Success Rate • 65.4%	5 min	Loop Detectors	Simulation	Type of Road • Freeway Place • California I-880

(*21*)	Development of a Real-Time Crash Risk Prediction Model Incorporating Various Crash Mechanisms Across Different Traffic States	Traffic Data • Flow • Occupancy • Speed • Types of congestion	• Random Parameter Models • Fixed Parameter Values	• Different rates	5-10 min	Loop Detector	Simulation	Type of Road • Freeway Place • California I-880
(*22*)	A Genetic Programming Model for Real-Time Crash Prediction on Freeways	Traffic data • Flow • Occupancy • Speed • Crash type	• RF Model • Genetic Programming	Success Rate • 75.4%	5-10 min	Loop Detectors	Simulation	Type of Road • Freeway Place • California I-880
(*23*)	Use of Support Vector Machine Models for Real-Time Prediction of Crash Risk on Urban Expressways	Date and time of crash Crash type Weather conditions Traffic data • Vehicle count • Avg speed • Avg occupancy	• SVM	Success Rate • 80%	5-10 min	Loop Detector	Simulation	Type of Road • Freeway Place • Shanghai
(*24*)	Comprehensive Analysis of the Relationship Between Real-Time Traffic Surveillance Data and Rear-End Crashes on Freeways	Traffic data • Avg speed crash type	• Kohonen Clustering Algorithm • Multilayer Perception • Normalized Radial Basis Function • Neural Network	Success Rate • 75%	5-10 min	Loop Detector	Simulation	Type of Road • Freeway Place • Orlando, I-4
(*15*)	A Real-Time Crash Prediction Model for the Ramp Vicinities of Urban Expressways	Upstream • Ramp flow • Flow • Congestion index downstream • Flow • Speed	• Bayesian Belief Net	Success Rate • 55% False • 10%	5 min	Loop Detectors	Simulation	Type of Road • Freeway without Ramps Place • Tokyo

(*25*)	The Viability of Using Automatic Vehicle Identification Data for Real-Time Crash Prediction	Traffic data • Speed travel time	• Random Forest for Variable Selection • Stratified Matched Case Control	Success Rate • 70%	30 min	AVI	Simulation	Type of Road • Freeway Place • Orlando, I-4
(*26*)	Real-time Prediction of Visibility Related Crashes	Traffic data • Mean speed Weather Crash type	• Bayesian Matched Case Control Logistic Regression Model	Success Rate • 73%	5-10 min	AVI	Simulation	Type of Road • Freeway Place • Orlando, I-4, I-95
(*27*)	Towards Universal Freeway Incident Success Algorithms	Traffic Data • Std deviation of speed • Avg volume	• Comparison of Different Algorithms	Success Rate • 71.4%	5-10 min	Loop Detectors	Simulation	Type of Road • Freeway Place • Melbourne, Australia
(*28*)	A Method for Relating Type of Crash to Traffic Flow Characteristics on Urban Freeways	Crash type Traffic data • Flow Highway geometry Weather conditions Visibility	• Principal Components Analysis • Cluster Analysis		5-10 min	Loop Detectors	Case Study	Type of Road • Freeway Place • California
(*29*)	Probabilistic Models of Freeway Safety Performance Using Traffic Flow Data as Predictors	Accident data Crash type No. of vehicles involved Traffic data • Movement of vehicles • Volume • Occupancy	• Statistical Summarization	Successful Alarm Rate • 92.00%	5 min	TASAS Database Loop Detectors	Case Study	Type of Road • Freeway Place • California
(*30*)	Real-Time Hazardous Traffic Condition Warning System: Framework and Evaluation	Traffic Data • Avg flow, speed, occupancy • Std deviation of speed, flow, occupancy	• Probabilistic Neural Network • Bayesian Network	Comparison of 2 Models	5 min	Loop Detector	Simulation	Type of Road • Freeway Place • California

(*31*)	Multiple-Model Framework for Assessment of Real-Time Crash Risk	• Crash type	• Bayesian Network	Model Comparison	5 min	Loop Detectors	Simulation	Type of Road • Freeway Place • Orlando, I-4
(*32*)	Split Models for Predicting Multivehicle Crashes During High-Speed and Low-Speed Operating Conditions on Freeways	Traffic data • Speed • Vehicle count • Occupancy	• Stratum Case Control Logistic Regression	Successful Alarm Rate • 89.30%	5 min	Loop Detector	Simulation	Type of Road • Freeway Place • Orlando, I-4
(*33*)	Dynamic Variable Speed Limit Strategies for Real-Time Crash Risk Reduction on Freeways	• Crash type Traffic data • Flow • Speed difference	• PARAMICS Microsimulation • (Software)	Successful Alarm Rate • 85.00%	5 min	Loop Detectors	Simulation	Type of Road • Freeway Place • Orlando, I-4
(*34*)	Estimation of Real-Time Crash Risk: Are All Freeways Created Equal?	Traffic data • Avg speed • Avg volume • Avg occupancy	• Multi-Layer Perceptron Neural Network	Success Rate • 79%, 77%, 70%, 70%	15-20 min	Loop Detector	Simulation	Type of Road • Freeway Place • Orlando, I-4, I-95
(*35*)	Calibrating a Real-Time Traffic Crash-Prediction Model Using Archived Weather and ITS Traffic Data	Weather Data Traffic data • Speed variance • Avg occupancy	• Logistic Regression	Success Rate • 59%	5 min	Loop Detector Weather Data	Simulation	Type of Road • Freeway Place • Orlando, I-4
(*36*)	Identifying Crash Propensity Using Specific Traffic Speed Conditions	Traffic data • Volume • Occupancy • Avg speed	• Probabilistic Neural Network	Success Rate • 70%	10-15 min	Loop Detector	Simulation	Type of Road • Freeway Place • Orlando, I-4

(*37*)	New Algorithms for Filtering and Imputation of Real-Time and Archived Dual-Loop Detector Data in I-4 Data Warehouse	Traffic data • Volume • Occupancy • Speed	• Pairwise Model		5 min	Loop Detector	Simulation	Type of Road • Freeway Place • Orlando, I-4
(*38*)	Development of a Crash Risk Index to Identify Real Time Crash Risks on Freeways	Crash data Traffic data • Speed • Volume • Occupancy	• Fisher Discriminant Analysis • Conditional Logistic Regression	Success Rate • 65.7	0-30 min	Loop Detector	Simulation	Type of Road • Freeway Place • California
(*39*)	An Analysis of Urban Collisions Using An Artificial Intelligence Model	Crash type Time of day Weather condition Crash data	• Artificial Neural Network	Success Rate • 58.33%	5 min	City Police Records TABTOT	Simulation	Type of Road • Arterial Place • Milan
(*40*)	Crash Risk Assessment Using Intelligent Transportation Systems Data and Real-Time Intervention Strategies to Improve Safety on Freeways	Traffic data • Variation of speed • Avg occupancy • Std deviation of volume	• Matched Case Control Logistic Regression	Success Rate • 72.50%	5 min	Loop Detector	Simulation	Type of Road • Freeway Place • Orlando
(*41*)	Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes Generalized Estimating Equations for Correlated Data	Traffic data • Avg speed • Volume • Occupancy	• Generalized Estimating Equation	Success Rate • 78.34%	0-15 min	Loop Detector	Simulation	Type of Road • Freeway Place • Orlando

(*42*)	Big Data Applications in Real-Time Traffic Operation and Safety Monitoring and Improvement on Urban Expressways	Traffic data • Avg speed • Volume • Occupancy • Congestion index	• Bayesian Logit Model • First Order Reliability Analysis	Success Rate • 65.70%	5-10 min	Microwave Vehicle Detection System	Simulation	Type of Road • Freeway Place • Orlando
(*43*)	Potential Real-Time Indicators of Sideswipe Crashes on Freeways	Traffic data • Avg speed • Avg flow • Avg occupancy • Crash type	• Overall Avg Flor Rate • (Modification of Parameters)		5-10 min	Loop Detector	Simulation	Type of Road • Freeway
(*44*)	Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models	Traffic data • Avg speed • Avg flow • Avg occupancy • Crash severity	• Ordered Response Model	Successful Alarm Rate • 78.34%	5-10 min	STATS19 UK road crash data	Case Study	Type of Road • Freeway Place • UK
(*45*)	Big Data Analytics Architecture for Real-Time Traffic Control	Traffic data • Speed • Position • Travel time • Volume • Obstacle • Occupancy	• Kafka • SUMO	Successful Alarm Rate • 60.32%	5-10 min	Video Loop Detector	Simulation	Type of Road • Freeway Place • Munich
(*46*)	Large-scale Automated Proactive Road Safety Analysis Using Video Data	Traffic data • Avg speed • Volume count	• Motion Prediction • Measurement of Tracking Accuracy	Success Rate • 94%	5-10 min	CCTV	Before and After Study	
(*47*)	Bayesian Updating Approach for Real-Time Safety Evaluation with Automatic Vehicle Identification Data	Traffic data • Avg speed • Std deviation of speed • Coefficient of variation of speed	• Naïve Bayesian	Successful Alarm Rate • 75.93%	5-10 min	Loop Detector	Simulation	Type of Road • Freeway

(*48*)	Real-time Crash Risk Prediction Models Using Loop Detector Data for Dynamic Safety Management System Applications	Traffic data • Std deviation of speed • Coefficient of variation of speed • Avg density	• Binary Logit	Successful Alarm Rate • 60.26%	0-5 min	Loop detector	Simulation	Type of Road • Freeway
(*49*)	Predicting Crash Likelihood and Severity on Freeways with Real-Time Loop Detector Data	Traffic data • Std deviation of speed • Avg density • Avg volume	• Binary Logit	Successful Alarm Rate • 91.40%	5-10 min	Loop detector	Simulation	Type of Road • Freeway
(*50*)	Real-time Crash Prediction on Freeways Using Data Mining and Emerging Techniques	Traffic data • Avg speed • Std deviation of speed • Avg density • Density variation • Avg volume • Std deviation of volume	• C-SVM (Support Vector Machine)	Successful Alarm Rate • 84.34%	5-10 min	Loop detector	Simulation	Type of Road • Freeway

Appendix B: Summary of Potential Vendors

Vendor	Link	Description	System Function	Dynamic Crash Prediction
WayCare	http://waycaretech.com/	Platform using in-vehicle information and municipal traffic data for predictive insights and proactive traffic management optimization.	• ***Dynamic crash prediction (mutual)*** • Proactive traffic management optimization • On-board automated incident detection and management • Data-driven decision for road safety improvement, and traffic flow and road design assessment	**Yes**
OpenDataNation	https://visionzeronetwork.org/resources/vision-zero-cities/	Cloud-based, smart city, machine learning engine and enterprise platform that brings together all data available to predict greatest risks of life.	• ***Dynamic crash prediction (in developing)*** • Connected cars • Data-driven decision making	In development
Brisksynergies	https://brisksynergies.com/	Uses AI and Deep Learning to evaluate video traffic interactions to understand road user behavior via cloud-based platform. Platform captures line pattern of each vehicle and predicts to reduce collisions.	• Crash detection based on trajectories (from videos)	No
Waze	https://www.waze.com/ccp	Free two-way data exchange empowering decisions to achieve concrete community impact.	• Data exchange	No
Data4democracy	https://github.com/Data4Democracy/crash-model	Open source application to build crash prediction modeling application that leverages multiple data sources to generate set of dynamic predictions to identify potential trouble spots and direct timely safety interventions.	• Crash prediction	?

UrbanLogiq	https://www.urbanlogiq.com/traffic	Platform to analyze traffic data and predict behavior based on historical data and real-time data.	• ***Dynamic crash prediction (in developing)*** • Smart traffic management • Integration of various data resources	In development
MioVision	https://miovision.com/	Solutions to help improve mobility and livability in cities of all sizes.	• Crash detection based on trajectories (from videos)	No
GreenRoad	https://greenroad.com/	Platform to provide real-time driver behavior data and give alerts to drivers and managers of vehicle fleet.	• Fleet management • Process driver behavior data and give alerts to drivers	No
TTC Driverprotect	https://www.ttc-driverprotect.com/	Division of TTC Group dedicated to delivering driver risk management and work-related road safety. End-to-end managed service is committed to minimizing workplace road safety risk and optimizing driver-related business performance.	• Fleet management • Driver risk management	No
Mojio connected car and Motion	https://www.moj.io/connected-car-platform/	Provides real-time GPS and behavior data for connected customer cars to help to shape clear understanding of driver behavior	• Fleet management • Process driver behavior data	No
Numina	http://www.numina.co/	Deploy-anywhere sensor solution that gives cities unprecedented traffic data.	• Crash detection based on pedestrian trajectories from videos	No

Appendix C: WayCare and Pilot Study in Las Vegas

C.1 Introduction to WayCare

WayCare is a start-up company headquartered in Tel Aviv, Israel, with offices in the U.S. It provides cloud-based solutions to shape future city mobility by using in-vehicle information and public traffic data for predictive insights and proactive traffic management optimization, including:

- Crash and incident identification and prediction
- Traffic management operations
- Dynamic traffic flow characteristic optimization
- Law enforcement & emergency services
- Roadway & safety service patrol
- Traffic engineering assessment for roadway safety

To archive these functions, WayCare's systems integrate real-time data from various resources beyond the existing roadway infrastructure, such as:

- TMC traffic monitoring data (loop, Bluetooth, etc.)
- Roadway camera feeds
- In-vehicle data (OBD II, navigation apps, telematics, Waze, etc.)
- Localized weather data
- Events (construction, lane closures, concerts, sports, etc.)
- Public transit
- Historical crash/incident data

WayCare integrates an implementable and tested dynamic crash prediction function in its system. To understand the features of the WayCare system, the research team conducted three interviews with WayCare staff and also searched news reports, webpages, and technical reports related to WayCare, especially dynamic crash prediction and prevention. WayCare has launched projects in Las Vegas, Nevada; Tampa and Pinellas County, Florida; and agencies in Delaware.

Based on the collected information, the dynamic crash prediction functions of the WayCare system are summarized below.

Platform and Deployment

The WayCare system is a cloud-based system that does not require deployment of specific hardware or software packages in local TMCs. A web-based user interface (UI) running on an Internet-connected computer software allows TMC users to access system functions, including monitoring traffic operations on target roadways, receiving crash risk warning information, configuring system settings, etc. An example of the WayCare interface for incident monitoring is given in Figure 52.

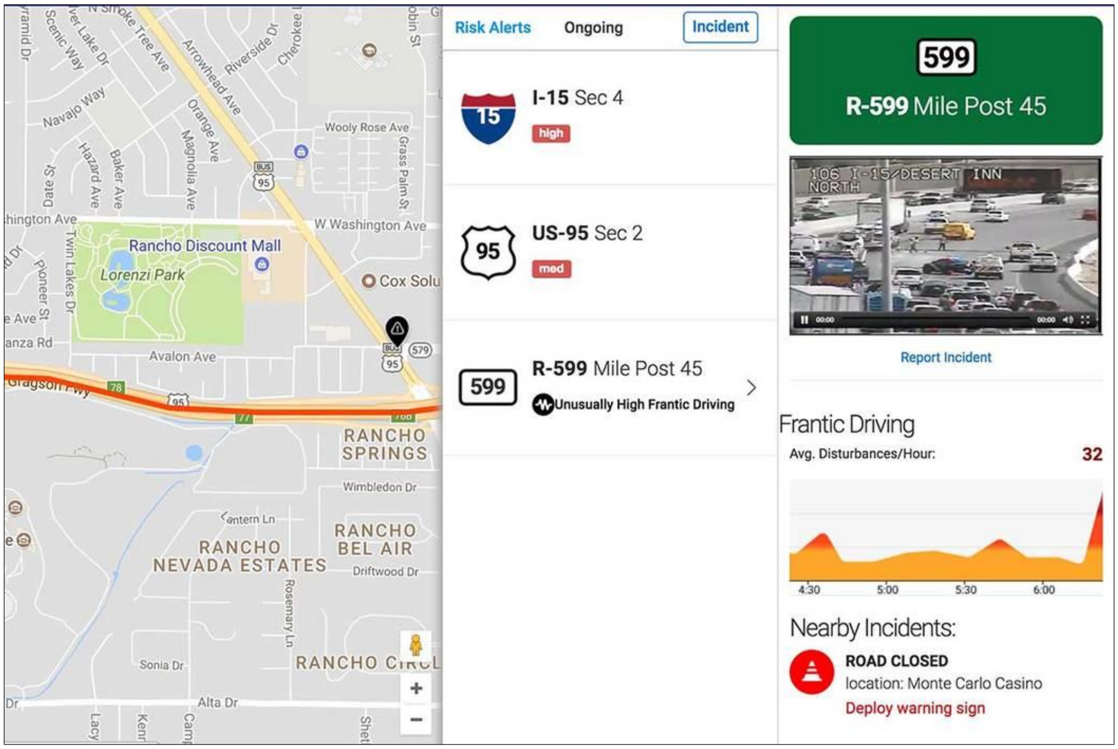


Figure 52. Web-based Interface of WayCare System Implemented in Nevada

(Source: Provided by WayCare)

The deployment of the WayCare system requires two stages: on-boarding and go-live.

- *On-boarding* (Calibration) – WayCare collects historical and real-time data from both external and internal data resources. WayCare then customizes the platform based on user needs and trains the prediction model based on the collected data.
- *Go-live* (Operation) – Once model training is completed, the system is activated online to monitor traffic conditions and predict crash risk. WayCare provides training and ongoing technical support at this stage.

System Functions

The current version of the WayCare system provides three major functions—dynamic crash prediction, incident detection, and reactivation of safety analysis.

- Dynamic crash prediction—The WayCare system can monitor real-time traffic conditions on target roadway segments. A machine-learning model continuously predicts crash risk based on the traffic conditions. The UI displays the predicted risk on a map with colors or texts to indicate the risk level. Once the crash risk is higher than a predefined threshold (i.e., 83% used in the Nevada pilot study), the system sends a warning message to the TMC and other involved agencies.

- Incident detection—The WayCare system provides a function to detect an incident/crash event after its occurrence. The integrated CCTV can display the field conditions to TMC staff.
- Reactivation of safety management—The WayCare system can collect historical crash data. Based on historical crash data, the system identifies the segments with a high crash risk and displays it on maps.

Roadway Facility Type

The dynamic crash prediction function of the WayCare system can be implemented in various roadway facility types, such as basic freeway segments and merging and diverging segments near interchanges. It tested the dynamic crash prediction function on two types of facilities in the Nevada pilot study. It also states that dynamic crash prediction can be applied on arterial corridors. Testing results for this facility type are not available.

Data Needs

Data needs are different for calibration (on-boarding) and operation (go-live). The system calibration needs historical data for model training, which includes:

- *Historical crash data* – Historical crash data are required from TMCs. The minimum requirement is one-year of historical data with location and direction information. However, multiple years of historical crash data are suggested; more crash data allows better training performance.
- *Historical traffic data* – Traffic conditions associated with identified crash events are an optional request from TMCs. WayCare can retrieve traffic data from external data resources (e.g., Waze). However, high-resolution traffic data from a TMC is suggested, as the data can significantly improve training performance.
- *Historical weather data* – Weather conditions associated with identified crash events are an optional request from TMCs. WayCare can retrieve weather conditions from external resources.
- *Historical event data* – Event information (e.g., construction, sports, concert, etc.) that associates with identified crash events is required from TMCs.

The data needs for online operations include:

- *Real-time traffic data* – WayCare can obtain real-time traffic conditions from external data resources (such as Waze) as traffic inputs for crash risk prediction. This means that even without traffic data from TMCs, the WayCare system can predict crash risk based on third-party data. This feature is beneficial for implementing the system in areas without local traffic data resources. However, high-resolution traffic data from TMCs are suggested. The prediction performance of the WayCare system can be improved significantly by using these high-quality data.

- *Event data* – WayCare needs event information (e.g., construction, concert, etc.) from TMCs or local agencies.
- *Weather data* – WayCare can obtain weather data from external resources. However, real-time weather conditions from local weather sensors is beneficial to improve prediction performance.

In summary, WayCare has its own data resources for model training and system operations, including floating car data from navigation apps (e.g., Waze) and weather condition information providers. In future, it plans to integrate individual vehicle sensor data (such as braking, speed, acceleration, etc.) from vehicle manufacturers. In addition, it expects high-resolution traffic data and multiple-year crash data to improve crash risk prediction performance. The accuracy and efficiency of dynamic crash prediction depends on data quality and quantity.

Crash-Prevention Actions

The WayCare system can send a warning message to TMCs, law enforcement, and emergency response departments when a predicted crash risk is higher than a predefined threshold. For instance, the pilot study in Nevada adopted 83% as the threshold. With warning information, local agencies may apply several control actions to prevent crash occurrence. The WayCare system tested three crash-prevention actions in the pilot study in Nevada:

- *Police high-visibility presence* – A police car with lights presents in a high-risk segment. Drivers slow down their speed and, consequently, reduce crash occurrence speed.
- *Dynamic message system* – Warning messages display on a DMS board in the upstream of the high-risk segment to notify drivers to pay attention to potential risk and slow down their speed.
- *Emergency service* – A warning message of potential crash risk is sent to emergency rescue services so reaction time to incidents can be significantly reduced.

In addition to crash-prevention actions that have been tested, WayCare is planning and developing innovative control strategies. An example is that WayCare can set up two-way communication between the TMC and drivers using the Sirius XM radio system. Drivers can receive notice of potential crash risk through two-way communication.

Performance and Impacts

The WayCare system tested its dynamic crash prediction and prevention in Nevada. According to its self-report, the performance and impacts of dynamic crash prediction are as follows:

- *Accuracy* – can correctly predict 56% of crash events
- *Timelines* – can provide a crash risk warning in advance, up to two hours
- *System latency* – nearly real-time (from time of data input to time of crash prediction output)
- *Crash reduction* – around 17% crash reduction and 23% secondary crash reduction after implementation of system
- *Speed reduction* – 91% of drivers reduce their speed below 65 mph in the high-risk segments when a police car presents with flashing lights
- *Emergency response time saving* – can reduce Highway Patrol response time by 12%

Implementation

The Regional Transportation Commission (RTC) of Southern Nevada, in cooperation with WayCare, completed a pilot study of dynamic crash prediction and prevention in July 2017. In the first phase, the pilot study was along US-95 in Las Vegas. In Florida, Tampa and Pinellas County also implemented the WayCare system. To date, efforts are focusing on incident detection rather crash prediction, although they plan to implement the latter.

C.2 Pilot Study in Las Vegas

Overview

WayCare, with the Regional Transportation Commission (RTC) of Southern Nevada, Nevada Highway Patrol (NHP), and the Nevada Department of Transportation (NDOT), conducted a year-long pilot study that began in September 2017. This is the first pilot project of an AI-based prediction and prevention of the crash risk based on real-time data in the US. The project has successfully demonstrated that its AI—paired with specific responses from law enforcement and transportation officials—can reduce highway collisions.

Period and Location

The first stage of the pilot study began in September 2017 and ended in September 2018. The first stage included assessment of available external and internal data sources, historical and real-time data collection, system customization, and a two-month go-live test (August–September 2018). The testing bed was a 5.4-mi interstate corridor, I-95 from Russell Rd to Charleston Blvd in Las Vegas. As shown in Figure 53, the features of the interstate segment are as follows:

- 4–7 through lanes on each direction
- 6 interchanges
- 2 horizontally curved sections
- Concrete barrier median
- Speed limit of 70 mph
- Dynamic Message System (DMS) in both directions

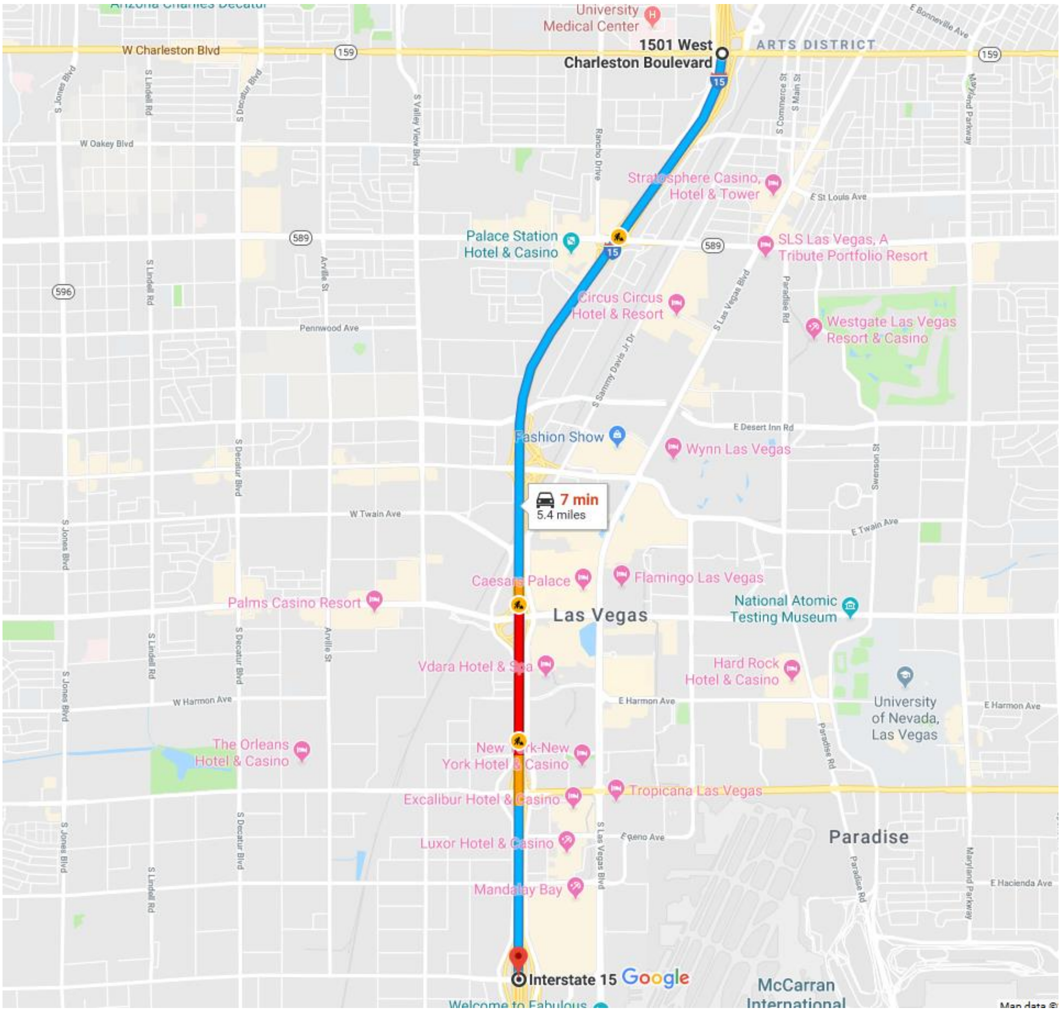


Figure 53. Test Bed of Nevada Pilot Study

(I-15, Russell Rd to Charleston Blvd, Las Vegas)

Data Needs

The real-time traffic data in this pilot study was collected mainly from crowdsourced apps such as Waze, a Google-owned company with the largest online community of motorists in the world. The system refined and synthesized real-time information including speed, braking, and acceleration. The pilot study also collected information on infrastructure, construction activities, weather conditions, and special events (sports or concerts) from local agencies. The WayCare system integrated these datasets and predicted potentially high crash-risk spots on the test bed. It also connected to cameras along the test corridor for incident verification.

Crash Prediction and Prevention

Combining all kinds of the data, the WayCare platform continuously predicts when and where crashes are likely to happen. In the Nevada pilot study, the warning-trigger threshold was set as

83%. If the predicted crash risk was higher than this value, the WayCare system sent a crash warning to TMC operators by popping up a warning window on the web-based UI.

With a crash risk alert, the WayCare system activated two crash-prevention actions, as shown in Figure 54, in the Nevada pilot study.

- *Police proactively positioned* – Stationed police vehicles presented in the high-risk segments with lights flashing. The high-visibility police vehicle increases driver attention and reduces their speed.
- *DMS with preliminary warning* – Two kinds of warning messages were displayed on DMS boards to notify upstream drivers of potential crash risk and encourage them to reduce speed—"Reduce Speed" and "Police Monitoring Ahead."

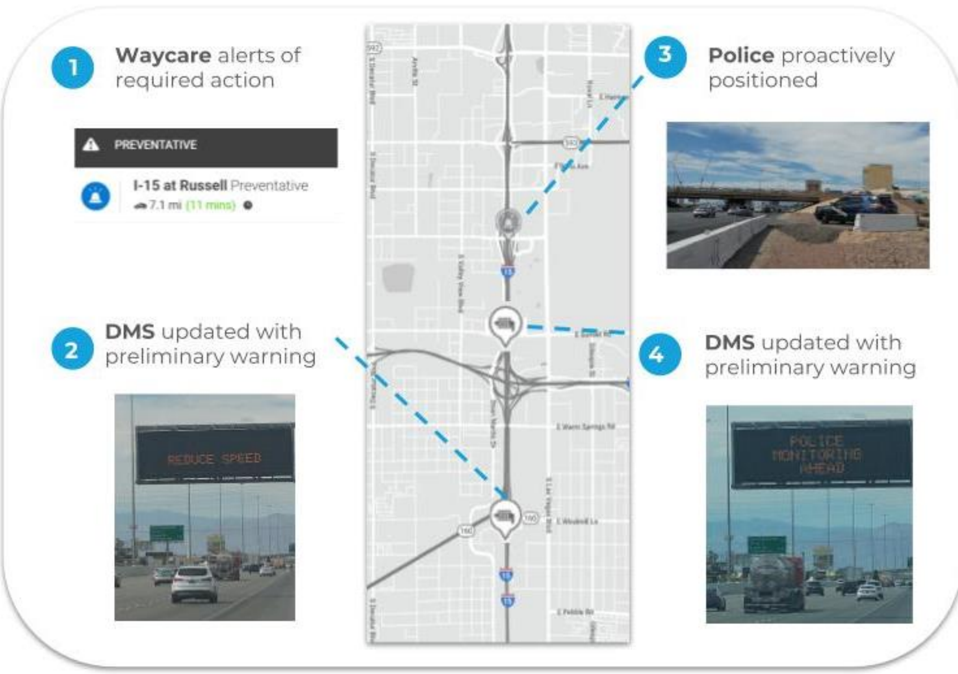


Figure 54. Crash Prevention Actions in the Nevada Pilot Study

(Source: Provided by WayCare)

Incident Response Solution

In addition to dynamic crash prediction, the WayCare system also tested incident response in the Nevada pilot study. The procedure of incident response in the pilot study was as follows:

- The WayCare platform synthesized information from social media to crowdsourcing apps (Waze) to identify crashes, near-crashes, or congestion events.
- Once the system identified an incident, the platform suggested potential problem areas via pop-up windows.

- TMC staff used those cameras more effectively, zooming out as far as a mile, then sending confirmed incident reports with geotags and 30-second video clips to officers en route.
- By receiving an advance or instant warning on an incident with detailed information, officers could rapidly determine the exact location and detailed status of the incident and quickly respond to the incident.

Effectiveness

The pilot study results showed that around 56% of crashes and incidents could be predicted. The advanced crash-risk warning time was up to two hours. The impacts of the WayCare system were identified as follows.

Speed Reduction

It was observed that 91% of drivers reduced their speed to lower than 65mph in the risk segments where police vehicles presented with flashing lights. Because speed is a predominant factor contributing to traffic fatalities, speed reduction is likely to decrease fatalities in traffic crashes.

Primary Crash Reduction

A before-after study was conducted to compare primary crash frequency before (without) and after (with) the implementation of the WayCare system. The comparison is given in Table 28.

Table 28. Before-After Comparison of Primary Crash Frequency in Nevada Pilot Study

Stage	Period	Number of Days	Number of Crashes	Crashes per Day
Before	May–July 2018	92	57	0.62
After	August–September 2018	29	15	0.52
Crash Reduction Rate				17%

Response Time and Secondary Crash Reduction

The probability of secondary collision rises more than 2.5% for every minute a travel lane is blocked. NHP officials estimated that with the WayCare system, there was a 12% improvement in NHP response time to an incident and a 23% reduction in secondary crashes, which are often more serious than primary crashes due to quicker incident clearance.

Appendix D: WayCare Calibration Report

waycare

Florida Freeway and Arterial Crash Prediction

May 1, 2020

Report for:
Center for Urban Transportation Research at
University of South Florida

Introduction:

Identifying dangerous road segments in real time carries significant value to traffic operators and Highway patrol agencies, allowing them to take proactive measures to allocate resources and increase safety. In doing so, lives may be saved, and much time, effort and money lost in the clearance of such events, unexpected traffic they cause and other inconveniences may be spared.

Usually the information available to law enforcement agencies about such events is static and statistical in nature, giving visibility only into "what historically took place" at given locations. Waycare developed (and continues refining) a machine learning (ML) approach based on historical data from various sources to greatly improve on those static methods.

In this project with the **Center for Urban Transportation Research** in the University of South Florida we were asked to implement and test such a model on three different road segments, namely, portions of Oakland Boulevard, Sunset Boulevard and Interstate 95. We were given historical traffic accident data, speed sensor measurements of different resolutions and construction data. In this paper, we describe in general terms the existing approaches (Literature overview), our approach and results.

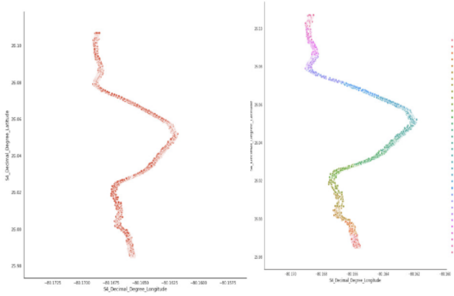
Literature Overview:

While in no way exhaustive, we mention here several published articles pertaining to the topic of crash prediction. Existing research typically either focuses on determining the outcome of crashes [1] or very coarse grained [2], predicting a year's crash total as its goal. A relatively recent model made 24-hour predictions over a 5km x 5km cells [8]. Some models are ML-based [1-3], others are parametric [4,5]. For a comprehensive review on the topic, see cf. [6,7] and references therein. We have yet to encounter a model offering operative time frames and targets for prediction.

Our Approach:

We chose to proceed with ML methodology as our main tool. Best practices in terms of data cleaning and feature engineering were employed and will not be detailed. The first question we address is what to set as the target for our prediction. Since both temporal and spatial characteristics are continuous, we cannot directly predict the exact time and place of a potential incident. In order for predictions to be of use as well as have some predictive power, we need to bin both time and space. All of the three road segments under investigation were processed in a similar manner with minor changes between them. We demonstrate the general approach on the I-95 segment. Fig. 1 shows the spatial distribution of incidents and the grouping derived from it. Resulting groups are roughly half a mile in length (of course, different directions are treated separately, i.e. Northbound incidents are grouped separately from Southbound).

Figure 1



Left: Overall incident distribution on the I-95 segment **Right:** Spatially grouped incidents

Several spatial grouping methods were tested, including distance-based, density-based, and clustering. Results did not show significant difference in performance so we chose a relatively uniform spacing that is easy to implement and comprehend.

Next we address the question of temporal binning. Our time frame needs to be such that it is operationally useful, while at the same time providing non-trivial information about the world. The first requirement would prevent us from using a very short time frame. Given, for example, near-real time speed measurements, we could attempt to detect minor perturbations in the steady flow to try and predict incidents minutes before they take place. This may be mathematically easier than longer-term predictions but would not be very useful from the perspective or traffic managers of Highway Patrol. The second requirement makes sure we avoid longer-term time frames, such as days or weeks, converging to statistical averages. Our time frame of choice after some experimentation was three-hour time windows.

Having made these decisions we construct the target data set to match our desired targets, i.e. grouping events by spatio-temporal bins.

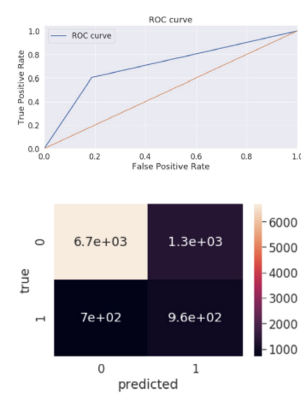
Data Processing:

Once our desired targets were set, all other datasets were processed to match. Since we are binning large segments of sensor data (as an example) into one "event", we extract statistical descriptors of each such block to make sure we retain as much useful information as possible.

Model Construction:

After data had been properly cleaned, imputed and binned, we proceeded to construct the actual model. One of the main difficulties of the data is (thankfully) the relative scarcity of "true" samples compared to "false" ones. That is, there are many more three-hour periods on each road segment where no events took place compared to periods with one or more incidents. Since imbalanced data sets are notoriously difficult for machine learning models, we used standard and some proprietary techniques to handle the imbalance. A first-pass model produces results shown in Fig. 2 (again, demonstrated on I-95).

Figure 2



Top: The ROC curve

Bottom: Confusion matrix for the model

As we can see from the confusion matrix, there are roughly 8000 false events and about 1600 true ones. For the next step we employ model stacking where several principally different learners are combined using a meta learner to produce the final outcome.

Remarks:

Parameter tuning: due to various constraints, we did not perform full parameter optimization for the stacking approach. Best-practice values were used. That means that in principle, better performance may be achieved given more effort.

Segment size: Our experiments with segment sizes show that increasing the segment size typically helps decrease false positive predictions. In case false positives are more harmful than the spatial granularity is helpful, the model may be optimized for usability.

Current conditions: While almost certainly self evident, it should be noted that the current conditions on the road due to restrictions imposed as a result of the COVID-19 epidemic are markedly different than those the model was trained on. Due to that we may expect deterioration in performance.

Conclusion:

We presented here a machine learning model to predict traffic accidents in the near future with relatively fine spatial granularity. The model shows good performance compared to baseline benchmarks (not discussed) such as trivial frequency-based prediction and space-time aggregated frequency prediction. In addition to the parameter optimization mentioned above, the model may be improved by incorporating additional data sources such as weather conditions\forecast, ad-hoc road conditions such as hazards, closures, etc and incorporating better granularity of vehicles on the road (proportion of trucks, for instance). As it stands, however, it presents a major source of previously unavailable information that may benefit traffic managers and law enforcement in their day-to-day operations.

Bibliography:

[1] Dong, C., Shao, C., Li, J. and Xiong, Z., 2018. An improved deep learning model for traffic crash prediction. Journal of Advanced Transportation, 2018.

[2] Jadaan, K.S., Al-Fayyad, M. and Gammoh, H.F., 2014. Prediction of road traffic accidents in jordan using artificial neural network (ANN). Journal of Traffic and Logistics Engineering, 2(2).

[3] Yasin Çodur, M. and Tortum, A., 2015. An artificial neural network model for highway accident prediction: A case study of Erzurum, Turkey. PROMET-Traffic&Transportation, 27(3), pp.217-225.

[4] Caliendo, C., Guida, M. and Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis & Prevention, 39(4), pp.657-670.

[5] Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention, 38(4), pp.751-766.

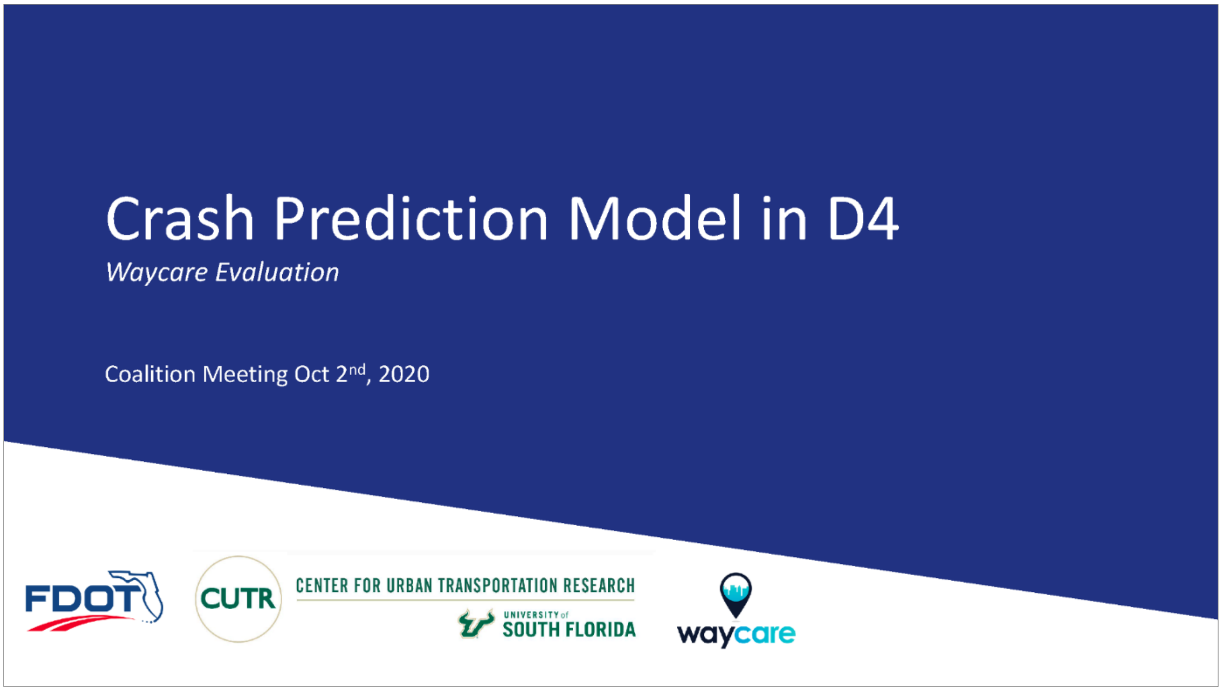
[6] Yannis, G., Dragomanovits, A., Laiou, A., La Torre, F., Domenichini, L., Richter, T., Ruhl, S., Graham, D. and Karathodorou, N., 2017, October. Road traffic accident prediction modelling: a literature review. In Proceedings of the Institution of Civil Engineers-Transport (Vol. 170, No. 5, pp. 245-254). Thomas Telford Ltd.

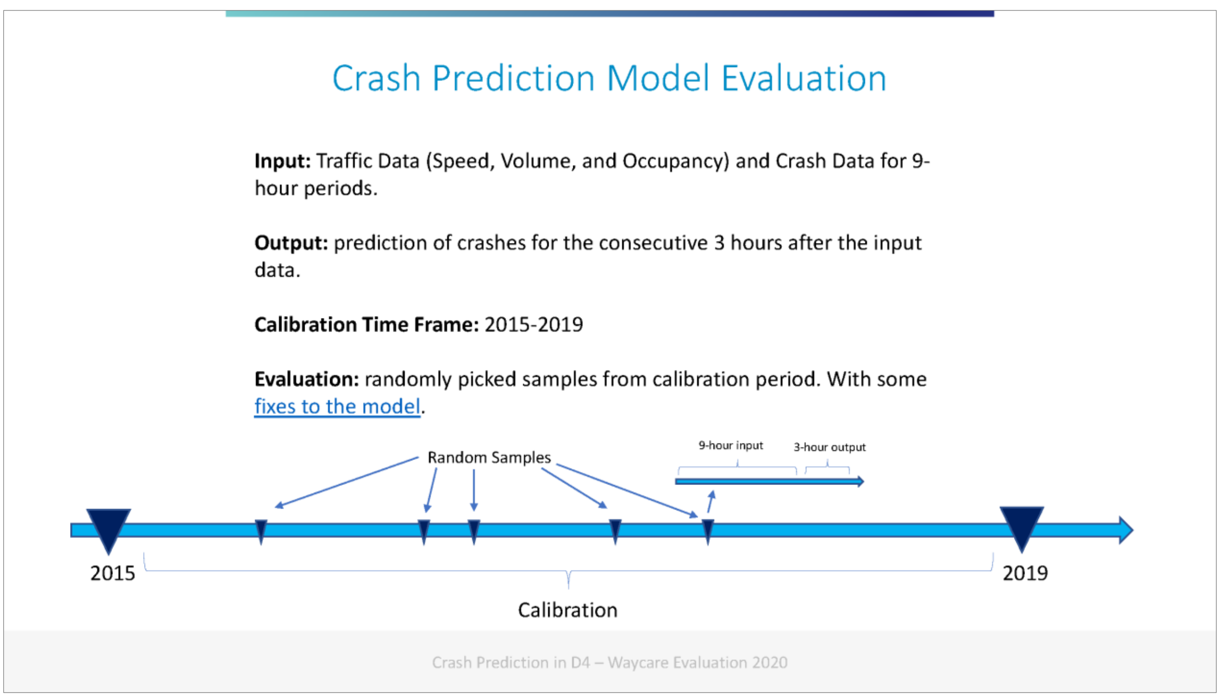
[7] Pradhan, B. and Sameen, M.I., 2020. Review of traffic accident predictions with neural networks. In Laser scanning systems in highway and safety assessment (pp. 97-109). Springer, Cham.

[8]Yuan, Z., Zhou, X. and Yang, T., 2018, July. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 984-992).

Appendix B: WayCare Evaluation Results

WayCare's evaluation was based on the randomly-selected samples on I-95 for 2015–2019. Evaluation results for Sunrise Blvd are not included.



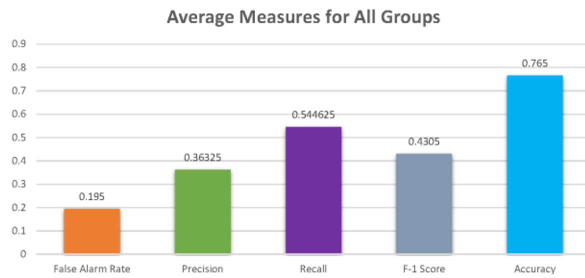


Crash Prediction Model Evaluation

Results Summary

$$\text{False Alarm Rate} = \frac{\text{False alarms}}{\text{Non Crashes}}$$

Rate of predictions the model was wrong out of non crash cases.



$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{Crashes + Non Crashes}}$$

Rate of cases the model was correct.

$$\text{F1 Score} = \frac{2(\text{Recall+Precision})}{\text{Recall+ Precision}}$$

Weighted average of precision and recall.

$$\text{Precision} = \frac{\text{crashes predicted}}{\text{crashes predicted + false alarms}}$$

Percentage of predictions the model was correct out of all predictions for crashes. The higher the precision- the lower the false positives.

$$\text{Recall} = \frac{\text{crashes predicted}}{\text{Crashes}}$$

Percentage of cases the model was correct out of crash cases. The higher the recall- the lower the false negatives.

Crash Prediction Model Evaluation

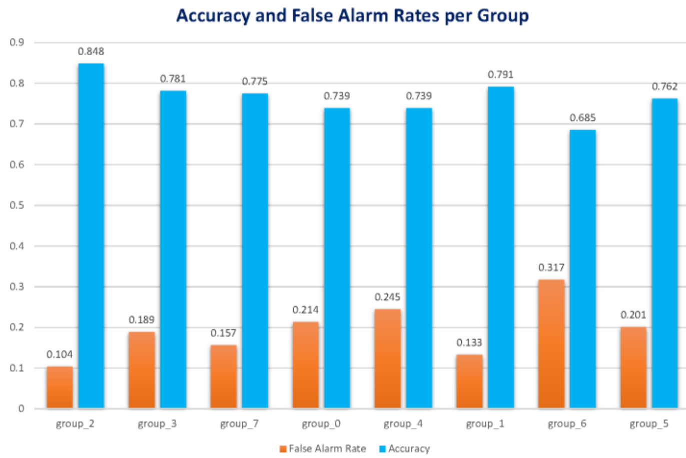
Caveats

- Given some of the false positives were near-miss crashes, this can possibly explain high numbers of false positives.
- Unreported Crashes- The true number of crashes is unknown; therefore number of reported crashes could be smaller than reality.
  - Typically in-vehicle data sources for Waycare account 20%-30% more incidents detected.

Limitations

- Off-line model developed solely using speed, volume, and occupancy as real-time input. On-line model with live feeds ensure increased data points and accuracy.
- Initial rendering of model underwent manual fixes while on-line model would benefits from deep learning capabilities and real-time input.

Crash Prediction Model Evaluation



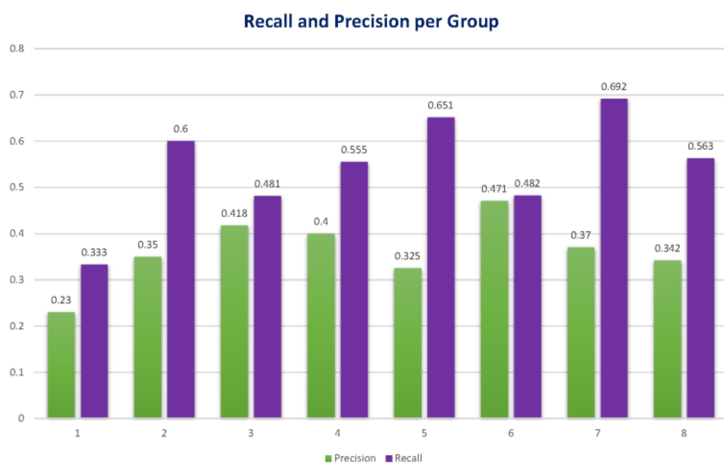
$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{Crashes + Non Crashes}}$$

Rate of cases the model was correct.

$$\text{False Alarm Rate} = \frac{\text{False alarms}}{\text{Non Crashes}}$$

Rate of predictions the model was wrong out of non crash cases.

Crash Prediction Model Evaluation



$$\text{Recall} = \frac{\text{crashes predicted}}{\text{Crashes}}$$

Percentage of cases the model was correct out of crash cases. **The higher the recall- the lower the false negatives.**

$$\text{Precision} = \frac{\text{crashes predicted}}{\text{crashes predicted + false alarms}}$$

Percentage of predictions the model was correct out of all predictions for crashes. **The higher the precision- the lower the false positives.**

Crash Prediction Model Evaluation

Assessing the model

In order to understand how useful a model is, we need to understand why using it is better than not using it.
For this example, we will take group number 7:

	Ground Truth		
	CRASH	NON-CRASH	
ALARM	143	199	342
NO ALARM	154	1072	1226
	297	1271	1568

False Alarm Rate	0.157
Precision	0.418
Recall	0.481
F-1 Score	0.448
Accuracy	0.775

Probability for a Crash: $\frac{\text{Number of Crashes}}{\text{Crashes + Non Crashes}} = \frac{297}{297 + 1271} = \mathbf{0.1894}$

If we try to build a basic prediction model, where:
probability for an alarm = probability for crash, and given number of cases=1480
then we will get these numbers:

	CRASH	NON-CRASH
ALARM	0.19*0.19*1481= **53.13**	0.19*(1-0.19)*1481= **227.38**
NO ALARM	0.19*(1-0.19)*1481= **227.38**	(1-0.19)*(1-0.19)*1481= **973.09**

Crash Prediction Model Evaluation

Assessing the model

Waycare's machine learning model:

	Ground Truth		
	CRASH	NON-CRASH	
ALARM	143	199	342
NO ALARM	154	1072	1226
	297	1271	1568

Basic probability- based model:

	CRASH	NON-CRASH
ALARM	0.19*0.19*1568= **56.25**	0.19*(1-0.19)*1568= **240.74**
NO ALARM	0.19*(1-0.19)*1568= **240.74**	(1-0.19)*(1-0.19)*1568= **1030.25**

Waycare's off-line model is better than a basic crash probability-based model.
It can be **2.5 times more** accurate in correctly predicting a crash (143 in compared to 56), can reduce **false alarms by 20%** (199 compared to 240), and can reduce **false negatives by 35%.**

Crash Prediction Model Evaluation

Off-Line Testing Discrepancy

- Off-line models and testing in not standard Waycare offering. Given nature of project the team was tasked with creating new tool which didn't go through the regular development, QA & iterations.
- Issues with the data processing in the real-time model- e.g. speed and occupancy levels were filtered incorrectly.
- Some issues with the incidents assigned (specifically affecting January and February-Incidents 2018 were assigned to segments using an incorrect field.
- Less than half (42%) of rendering were successful which indicates some sort of error (testing or code).
- July 2020 as one of the three tested month may be use given Covd-19 impact on traffic.
- **To achieve similar or better results to the crash prediction model mentioned previously necessary development would be required or testing should be conducted on live platform with real-time data.**