2019

The use of real-time connected vehicles and HERE data in developing an automated freeway incident detection algorithm

Hendry Nyanza Imani
University of North Florida, h.imani@unf.edu

THE USE OF REAL-TIME CONNECTED VEHICLES AND HERE DATA IN DEVELOPING AN AUTOMATED FREEWAY INCIDENT DETECTION ALGORITHM

By

Hendry Nyanza Imani

A thesis submitted to the School Engineering

In partial fulfillment of the requirements for the degree of

Master of Science in Civil Engineering

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF COMPUTING, ENGINEERING, AND CONSTRUCTION

December 2019

The thesis "The use of Real-time Connected Vehicles and HERE data in Developing an Automated Freeway Incident Detection algorithm" submitted by Hendry Imani in partial fulfillment of the requirements for the degree of Master of Science in Civil Engineering has been

Approved by the thesis committee: **Date:**

__________________________________ ____________________

Dr. Thobias Sando,

Thesis Co-Advisor and Committee Chairperson

__________________________________ ____________________

Dr. Priyanka Alluri,

Thesis Co-Advisor and Committee Member

__________________________________ ____________________

Dr. Cigdem Akan ,

Committee Member

DEDICATION

I dedicate this thesis to the Almighty God for his fruitful help to finish this thesis, as well as my friends, and also to myself.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Thobias Sando, for his immense support. He spent much of his time to instruct, assist, encourage and advise me. His help has been a great support towards the completion of this thesis work. It is great honor to work under his supervision.

Special thanks also go to my committee members, Drs. Priyanka Alluri and Akan Cigdem for their creative ideas and guidance throughout my thesis preparation.

Lastly, I would like to appreciate supports from ALL UNF Transportation Laboratory members of the year 2019 and everyone else who helped contribute to this thesis.

Thanks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AID	Automatic Incident Detection
ATMS	Advanced Traffic Management System
AVI	Automatic Vehicle Identification
BSM	Basic Safety Message
CCTV	Closed-circuit Television
CV	Connected Vehicle
DMS	Dynamic Message Signs
DR	Detection Rate
DSRC	Dedicated Short Range Communication
FAR	False Alarm Rate
FDOT	Florida Department of Transportation
ITS	Intelligent Transportation Systems
MTTD	Mean Time to Detect
OBU	On-board Unit
PI	Primary Incident
RSU	Road Side Unit
SC	Secondary Crash

SPaT	Signal Phase and Timing
SPMD	Safety Pilot Model Deployment
THEA	Tampa Hillsborough Expressway Authority
TIM	Traffic Incident Management
TMC	Traffic Management Center
USDOT	US Department of Transportation
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
VANET	Vehicle Ad-hoc Network
VSL	Vehicle Service Locator

ABSTRACT

Traffic incidents cause severe problems on roadways. About 6.3 million highway crashes are reported annually only in the United States, among which more than 32,000 are fatal crashes. Reducing the risk of traffic incidents is key to effective traffic incident management (TIM). Quick detection of unexpected traffic incidents on roadways contribute to quick clearance and hence improve safety. Existing techniques for the detection of freeway incidents are not reliable.

This study focuses on exploring the potential of emerging connected vehicles (CV) technology in automated freeway incident detection in mixed traffic environment. The study aims at developing an automated freeway incident detection algorithm that will take advantage of the CV technology in providing fast and reliable incident detection. Lee Roy Selmon Expressway was chosen for this study because of the THEA CV data availability.

Findings of the study show that the emerging CV technology generates data that are useful for automated freeway incident detection, although the market penetration rate was low (6.46%). The algorithm performance in terms of detection rate (DR) and false alarm rate (FAR) indicated that CV data resulted into 31.71% DR and zero FAR while HERE yielded a 70.95% DR and 9.02% FAR. Based on Pearson’s correlation analysis, the incidents detected by the CV data were found to be similar to the ones detected by the HERE data. The statistical comparison by ANOVA shows that there is a difference in the algorithm’s detection time when using CV data and HERE data. 17.07% of all incidents were detected quicker when using CV data compared to HERE data, while 7.32% were detected quicker when using HERE data compared to CV data.

Keywords: Automated Freeway Incident Detection, Connected Vehicles, Traffic Incident Management, Detection Rate, False Alarm Rate, Mean Time to Detect.

CHAPTER 1 INTRODUCTION

Traffic incidents such as crashes, vehicle breakdowns and spilled loads cause severe problems on roadways. They are a source of traffic delays, traffic congestion, environmental pollution, vehicular damage, personal injury, and secondary accidents (Yang et al., 2017). They deteriorate traffic safety and operational conditions. About 6.3 million highway crashes are reported annually only in the United States, among which more than 32,000 are fatal crashes (NHTSA, 2016). In 2014, motorists spent about 6.9 billion hours and 3.1 billion gallons of fuel, equivalent to approximately $160 billion, because of traffic congestion in the United States (Schrank et al., 2015). Traffic incidents also account for more than a half of all urban traffic delays and almost all rural traffic delays (Baykal-Gürsoy et al., 2009). Furthermore, traffic incidents increase the chance of occurrence of secondary crashes (SCs) (Karlaftis et al., 1999). For every additional minute an incident is present on the freeway during a peak travel period, the chance of secondary crash occurrence increases by 2.8% and result into a 4-minute delay to the traffic using the freeway (Owens et al., 2010).

Reducing the risk of traffic incidents is key to effective traffic incident management (TIM) (Owens et al., 2009). There has been a growing interest by transportation agencies in addressing traffic incidents once they occur. In practice, many TIM agencies have adopted quick clearance policy for minimizing the traffic flow disruption and reduce the potential for secondary incidents (Xie et al., 2018). An aggressive clearance strategy plays a big role on incident management and incident duration. However, implementation of this strategy depends on how fast the incidents are detected. Delay in detection increases traffic congestion and the chance of another crash hence making hard for the responders to reach the incident scene. Quick detection of unexpected traffic incidents on roadways contribute largely to quick clearance and hence reduce traffic delays and increase safety. Fast and reliable freeway incident detection guide the traffic flow toward smooth operation by communicating relevant information to travelers entering the freeway upstream of the incident (Stephanedes et al., 1992).

Different techniques have been used to detect freeway incidents as quick as they occur. These include cellular telephone call-ins, navigation applications such as WAZE, closed-circuit television (CCTV) cameras, road sensors/detectors, automatic freeway incident detection (AID) algorithms and freeway service patrol such as road rangers. After the initial detection, the incident is verified using the traffic monitoring video cameras, or a freeway service patrol in areas that do not have video coverage, or areas outside of the planned coverage.

Existing techniques for the detection of freeway incidents do not provide the necessary reliability for freeway operations. For instance, telephone call-in is not applicable in some condition as the use of telephone while driving is against the law, navigation application tends to produce a lot of false alarms and location inaccuracy. Majority of AID algorithms generates high level of false alarms and poor detection rate that operators are unable to rely on. Lastly CCTV cameras, sensors and highway service patrol cover only some areas and hence make freeway incident detection unreliable in many areas.

Recent technology advancements, especially in the areas of information and communication, enables vehicles to sense their environment, communicate with other entities, such as other vehicles, drivers, pedestrians, and roadside units, and eventually to be self-navigating, without human input. Connected Vehicles (CVs) have potential of improving transportation incident management, given the capability of CVs to communicate important information between themselves and the surrounding infrastructure (Iqbal et al., 2018). This potential, coupled with roadside equipment, can subsequently provide the ability to alert drivers about downstream incidents, which can lead to enhanced safety and mobility. A notable benefit of such technology could be an automated freeway incident detection.

Study Objective

Although the CV technology seems to have potential of improving transportation incident management, its usability in an automated freeway incidents detection in mixed traffic environment is not yet investigated. This study focuses on exploring the potential of the emerging CV technology in automated freeway incident detection in mixed traffic environment. The study aims at developing an automated freeway incident detection algorithm that will use the CV technology in providing fast and reliable freeway incident detection.

Thesis Organization

This thesis contains six chapters. Chapter 1 gives the general introduction and study objective. Chapter 2 provides an extensive literature review on the CV technology and automated freeway detection algorithms. Chapter 3 discusses the framework of the study methodology and the proposed algorithm. Chapter 4 describes the study site and the data collection effort. Chapter 5 discusses the analysis and results, and finally, Chapter 6 outlines the conclusions and recommendations for future work.

CHAPTER 2 LITERATURE REVIEW

This chapter provides a summary of an extensive literature review on automated incident detection (AID) algorithms and connected vehicles (CVs) technology. A critical review of the most widely accepted conventional incident detection algorithms are presented and their limitations discussed. Some emerging incidents detection techniques are also reviewed and lastly the CV technology is reviewed and its potential for use in automated incident detection.

2.1 Automated Incident Detection Algorithms

The automated incident detection (AID) is a method for quickly detecting potential incidents once they occur (Li et al., 2013). AID has been researched, and tested since the 1970s (Martin et al., 2001). During that time, many incident detection methods and algorithms were developed. There are several comprehensive studies focusing on or containing a state-of-the-art literature review of existing incident detection algorithms (Parkany et al., 2005; Stephanedes et al., 1992b; Martin et al., 2001). Each of the above studies had slightly different evaluation scopes, classification schemes, and performance results. Although the algorithms have different structural complexities, data requirements, calibration and implementation methods, most of them use data from fixed in-road or roadside sensors on freeways and their performance is heavily affected by the data quality, which is a function of the detection accuracy and reliability of sensors. Based on the data collection techniques used, the automated incident detection algorithms can be classified into; roadway-based algorithm, probe-based algorithms and video image processing. The following section discusses the mentioned algorithm categories in detail.

Roadway-Based Algorithms

Roadway-based algorithms relies on the traffic data that are usually collected from roadway-based sensors such as inductive loop detectors, magnetic sensors, microwave sensors, infrared sensors, ultrasonic sensors, acoustic sensors and laser sensors. Studies by Parkany et al., (2005) and Martin et al., (2001) categorized roadway-based incident detection algorithms into four classes: (1) pattern recognition, (2) catastrophe theory, (3) statistical, and (4) artificial intelligence. While pattern and statistical based algorithms were first created in the 1970s, artificial intelligence is the latest and least mature of the four. The brief summary of these categories and their corresponding algorithms are given in the following section followed by a brief summary of their performance.

Currently, pattern-based algorithms are the most commonly used. They use occupancy, traffic volume, and traffic flow data. Potential incidents are recognized by identifying abnormal patterns in the data for a stretch of a roadway. This method requires preset thresholds that define normal interrupted flow. Anything outside of 'normal flow' should set off an alarm. Setting these thresholds is difficult and time consuming as they play an important role in the algorithm performance. Algorithms that fit into this category include the California Algorithm, Traffic Services Corporation (TSC) Algorithm 7, Traffic Services Corporation (TSC) Algorithm 8, All Purpose Incident Detection (APID), and Pattern Recognition Algorithm (PATREG).

The catastrophe theory adopts its name from sudden discrete changes that occur in one variable of interest while other related variables exhibit smooth and continuous change. These variables are speed, flow, and occupancy. When speed drops dramatically without a corresponding increase in occupancy and flow, an alarm sounds. Catastrophe algorithms can differentiate between incidents and recurring congestion, congestion builds up slowly, while incidents cause a sudden queue to develop and drastic changes in speed to occur. The only type of algorithm that fits into this classification is the McMaster algorithm.

Statistical methods compare real-time traffic data with data forecasts. These algorithms model the actual traffic patterns, using time series data, and create a forecasted range of values. Any unexpected changes in traffic are compared to the forecasted traffic flows for use in classifying incidents. The advantage to this method is that it does not require large amounts of data. Algorithms that fit into this category are a High Occupancy (HIOCC) algorithm, a Double Exponential Smoothing (DES) algorithm, a Bayesian algorithm, Auto-Regressive Integrated Moving-Average (ARIMA) time series algorithm, Single-Station Incident Detection (SSID) algorithm, Standard Normal Deviates (SND) and filtering models.

Artificial Intelligence (AI) is a recent development of AID algorithms. These algorithms detect incidents by either a rule-based algorithm or an algorithm that has learned to recognize incident patterns. Neural Networks and a Fuzzy Set Logic are the main AI applications that have been applied to AID.

Table 2.1 gives a summary of the existing AID algorithms, traffic parameters used and number of detector stations. Traffic variables used most frequently are speed, volume, and occupancy. Single-station algorithms are based on traffic data obtained from only one traffic detector located either upstream or downstream of a traffic incident. Dual-station algorithms adopt traffic data obtained from detectors located at both the upstream and downstream of a traffic incident. Detection rate

(DR), false alarm rate (FAR), and mean time to detect (MTTD) are the three parameters that have been mostly used to measure the performance of AID algorithms.

Table 2.1: The reported performance of existing incident detection algorithms (Source: Parkany & Xie, 2005, Martin et al., 2001)

Type of Algorithm	Algorithm	Traffic Variables			Number of Stations		Measure of Performance		
		Volume	Occupancy	Speed	Single	Dual	DR (%)	FAR (%)	MTTD (min)
Pattern Recognition	California Algorithm		*			*	82.00	1.73	0.85
	TSC Algorithm 7		*			*	67.00	0.13	2.91
	TSC Algorithm 8		*			*	68.00	0.18	3.04
	APID Algorithm		*			*	89.00	0.05	2.50
Catastrophe Theory	McMaster Algorithm	*	*	*	*	*	100.00	0.04	1.50
Statistical	HIOCC Algorithm		*		*		96.00	2.60	2.50
	ARIMA Algorithm		*		*		100.00	1.50	0.40
	SND Algorithm		*			*	92.00	1.30	1.10
	DES Algorithm	*	*	*		*	92.00	1.87	0.70
	Filtering Models		*			*	80.00	0.35	3.00
	Bayesian Algorithm		*			*	100.00	0.00	4.00
	SSID Algorithm		*		*		100.00	0.20	3.00
Artificial Intelligence	Neural Networks	*	*	*	*	*	89.00	0.92	0.96
	Fuzzy Set Algorithm	*	*	*	*	*	N.A	N.A	N.A

Based on literature, when most of these traditional AID systems were installed, the number of false alarms became a problem causing the transportation agencies to stop using them. Moreover, other systems have a poor detection rate that operators were unable to rely on them as the primary method of incident detection. The tendency of frequently malfunction of the majority of roadway-based detectors makes traffic data unreliable, and hence lead to poor performance of the algorithms. In additional to that, the installation and maintenance of roadway sensors is costly hence only available in a few targeted areas.

Probe-Based Algorithms

Probe-based algorithms derive traffic data from probe-based sensors. Probe-based sensors refer to vehicle-mounted sensors that have positioning or identification functions and have the capability of transmitting real-time individual probe data to roadside readers or to a remote base station through wireless communication. Compared to roadway-based sensors, this type of sensor is mobile and hence can sense the spatial variation of traffic flow over a wide area (Parkany & Xie, 2005). Probe-based sensor technologies include automatic vehicle location (AVL), cellular positioning systems, and automatic vehicle identification (AVI). Automatic vehicle location (AVL) systems are designed to determine the location of a vehicle at a particular point in time using long-range communication (Hellinga & Knapp, 2000). The most widely used AVL technology is the global positioning system (GPS). Automatic Vehicle Identification (AVI) is designed to identify a vehicle that is situated at a specific location at a specific time using short-range communication (Bernstein & Kanaan, 1993). AVI systems have two primary components, the transponder and the reader, with a wireless communication link between them. Some of the applications of the AVI technology include electronic toll collection (ETC), electronic congestion pricing (ECP), and fleet control. The following algorithms belong to probe-based group; ADVANCE algorithm, The Texas Transportation Institute (TTI) algorithm, TRANSMIT algorithm and Waterloo algorithm. Most of these algorithms make use of travel time to detect traffic variations in order to judge whether an incident has occurred or not. ADVANCE and Waterloo algorithms were based on simulation experiments while TTI and TRANSMIT were based on the field data obtained from Texas and New York City (NYC) respectively. Probe-based algorithms are highly controlled by the penetration rate of vehicles equipped with sensors in a traffic network, the lower the penetration the lower the performance of these algorithms. Moreover, GPS-based incident detection scheme may suffer from weakening or even blockage of the GPS satellite signals caused by high buildings in the area.

Video Image Processing

Video Image Processors (VIP) employ machine vision techniques to automatically analyze traffic data collected with Closed Circuit Television (CCTV) systems or other video cameras (Martin et al., 2001). A VIP system consists of one or more video cameras, a microprocessor-based computer for digitizing and processing the video imagery, and software for interpreting the images and converting them into traffic flow data (Parkany & Xie, 2005). Video image technology can provide information about traffic flow at a higher level, it can measure travel times, average speed, and detect stalled or stopped vehicles within the detection zone. The image-processing program interprets the entire video image to find stationary or slow-moving vehicles, to detect an incident. An Autoscope incident detection algorithm (AIDA) is an example of the image processing algorithms. The AIDA algorithm takes advantage of temporal variations of traffic characteristics in addition to spatial ones, it looks for rapid traffic breakdowns, comparing speed and occupancy with the preset thresholds for determining congestion levels (Mahmassani et al., 1998). AIDA has been reported to have DR of 80% with only a 3% FAR and MTTD of 63 sec (Martin et al., 2001). One of advantages of the image processing-based incident detection technique is that a detected incident in the field of view of a video camera can be verified visually in a short time. It is also capable of monitoring traffic and detecting incidents outside of through lanes, e.g., shoulders, intersections, or ramps, and under both low and high-volume traffic conditions (Parkany & Xie, 2005). Although the image processing technology shows promising results, it only applicable within the range of the camera's vision. Video image processing algorithms also degrade as visibility decreases. It is highly affected by weather and lighting conditions.

2.2 Application of Smartphones

Freeway incidents have been detected by cellular telephone for decades using emergency calls from drivers. The emerging of smartphones lead to the rise of navigation applications such as WAZE. WAZE is a navigation application that leverages crowdsourced user reports for providing service. This system allows users to create and send highway advisory messages from their smartphone at the incident scene. The use of this communication technology helps motorists know what is happening on the road, alerts them instantly about the traffic, incidents, police, construction, and even detour suggestions to save time (Waze, 2006).

A study by Amin-Naseri et al., (2017) evaluated the reliability, coverage, and added value of crowdsourced traffic incident reports from WAZE in Iowa. The study concluded that the crowdsourced data stream from WAZE is an invaluable source of information for broad coverage traffic monitoring, covering 43.2% of Iowa's Advanced Traffic Management System (ATMS) crash and congestion reports. The WAZE application also offered timely reporting, 9.8 minutes earlier than the probe-based alternative, on average, and with reasonable geographic accuracy. WAZE reports currently make significant contributions to incident detection and were found to have potential to further complement the ATMS coverage.

Although crowdsourced data are usually relatively inexpensive, there are challenges in understanding and interpreting this type of data. Crowdsourced data may be slightly inaccurate in time and/or location of an incident. For example, for users traveling on a freeway at 60 miles per hour, a 30 second delay in reporting an incident will result in a 0.5-mile error in the incident location. Moreover, users may assume irregular congestion may be due to a crash event and report a crash when none exists. For less crowded hours of the day (12 a.m. to 6 a.m.), WAZE reports are not a reliable source for monitoring road conditions.

2.3 Traffic Managing Software

The traffic managing software such as SunGuide™ offers a comprehensive set of tools to the traffic management centers (TMCs), including managing Intelligent Transportation Systems (ITS) devices, incident detection, and assisting with event management. TMCs manage traffic flow along the roadway network by monitoring closed-circuit television (CCTV) cameras, road sensors/detectors, video wall, and other ITS devices, and coordinating with law enforcement agencies, and Road Ranger service patrols. Since SunGuide™ is an open architecture-based software, it enables users to manage multiple subsystems. Operators can use the software to perform incident management tasks, obtain real-time traffic data from vehicle detection systems, and display videos from roadside cameras, and then alert motorists using DMS messages and the highway advisory radio.

These activities are fairly time consuming. If the roadway network has too few ITS devices, TMC operators have to depend heavily on law enforcement officers and Road Rangers, spending a significant portion of their time in coordination. However, if the roadway network has a significant penetration of ITS devices, then TMC operators could be overwhelmed with information from each of these devices. Installing and managing this system is costly hence only covers some targeted areas and make freeway incident detection unreliable in many areas.

2.4 Connected Vehicle Technology

Recent technology advancements, especially in the areas of information and communication enable vehicles to sense their environment, communicate with other entities, such as other vehicles, pedestrians, and roadside units, and eventually to be self-navigating, without human input. CVs have the potential of improving transportation incident management, given the capability of CVs to communicate important information between themselves and the surrounding infrastructure (Iqbal et al., 2018). This potential, coupled with roadside equipment, can subsequently provide the ability to alert drivers of downstream incidents, which can lead to enhanced safety and mobility. A notable benefit of such technology could be the automated freeway incident detection on mixed traffic environment.

Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications are key aspects of CV technology. V2V communication allows CVs to essentially “speak” with each other, by transmitting basic safety messages (BSMs) through Vehicle Ad-hoc Network (VANETs) and most often using the Dedicated Short-Range Communication (DSRC) protocol (Yang et al., 2017). DSRC systems typically involve vehicle on-board units (OBUs) with transceivers and transponders to receive and broadcast relevant information. The OBUs provide situational awareness to drivers, giving them the opportunity to make travel decisions in response to abnormal conditions downstream, intersection approach queues, and incident sites. Unlike V2V communication, V2I involves communication between vehicles and infrastructure equipment. Through this communication system, CVs share information with road side units (RSUs) linked to ground servers and traffic control centers (Yang et al., 2017). V2I communication could convey various real-time traffic information, including the presence of downstream congestion, advisory speed limits on sharp curves, traffic signal status, stop sign warnings, and pedestrian crosswalk warnings. Such real-time information promotes driver awareness, which may improve safety for motorists (Marshall et al., 2016).

One of the largest CV deployments in the nation is in Ann Arbor, Michigan. In this initiative, also known as Safety Pilot Model Deployment (SPMD), more than 3,000 OBUs have been installed in vehicles and are used to collect and send various data, including Signal Phasing and Timing (SPaT) and Basic Safety Messages (BSM) (Xie et al., 2018). BSMs, known globally as messages containing safety information of a vehicle and transmitted periodically to surrounding vehicles, are a subset of the SPMD dataset (Henclewood & Rajiwade, 2015). Several studies have explored the potential of CVs by using this dataset. One study by Liu & Khattak (2016) investigated the potential of instantaneous driving decisions contained in BSMs, such as hard accelerations or braking and quick lane changes, in identifying critical events. The study indicated that BSMs can inform drivers about their driving behaviors or dangers in surrounding roadway environment. The study also pointed out that the information is simple and informative, and helps drivers make informed decisions. Another study by Xie et al. (2018) explored the potential of using CV data to identify high-risk locations in a more proactive manner, without relying on the historical crash data that often takes considerable time to collect. The study concluded that high-risk locations identified by CV data were found to be similar to the ones identified by historical rear-end crash data.

Another CV pilot deployment program was launched in September 2015 to deploy, test and operationalize cutting-edge mobile and roadside technologies and to enable multiple connected vehicle applications. This program was sponsored by the Intelligent Transportation Systems (ITS) Joint Program Office (JPO), under USDOT which awarded cooperative agreements to three agencies; New York City Department of Transportation, Tampa Hillsborough Expressway Authority and Wyoming Department of transportation. Tampa-Hillsborough Expressway Authority (THEA) owns and operates the Selmon Reversible Express Lanes (REL), which is a first-of-its-kind facility to address urban congestion. THEA pilot has deployed a variety of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) applications to relieve congestion, reduce collisions, and prevent wrong-way entry at the REL exit. The THEA CV Pilot uses DSRC to enable transmissions among approximately 1,600 cars, 10 buses, 10 trolleys, 500 pedestrians with smartphone applications, and approximately 40 roadside units along city streets (THEA, 2018). Currently no study has explored the potential of CVs by using this dataset.

CV technologies collect point-to-point traffic data over a road section. Point-to-point traffic data such as travel time better describe traffic conditions over a road section (Hellinga & Knapp, 2000). This advantage seems to solve the detector/sensor spacing which was reported as a major challenge of majority of the AID algorithms. Currently there is no study that has explored the potential of CVs in automated freeway incident detection by using any of CVs pilot dataset. Therefore, this study explores the usability of CVs technology on automated freeway incident detection.

CHAPTER 3 METHODOLOGY

Different traffic parameters such as volume, speed and occupancy respond to traffic incidents differently in time and magnitude (Corby et al., 1997). It is therefore important to select the proper traffic parameters inputs to the proposed algorithm. The proposed algorithm relies on the principle of pattern recognition and the statistical method, by identifying patterns in the data that are considered 'normal' for a stretch of a roadway and use the statistical approach to create a forecasted bound of normal flow. To detect incidents, real-time traffic speed is compared to the predefined thresholds.

Traffic speed has seldom been used in automated incident detection algorithms due the reason that the majority of previous AID algorithms retrieved data from roadway-based sensors such as inductive loop detectors. These detectors collect traffic flow and occupancy only. With advancement of technology, traffic speed can be collected with even more accurate nowadays. Therefore, this study adopts traffic speed as the metric used for the proposed algorithm.

Traffic speed is one of the most commonly used traffic parameter as an indicator of traffic incidents (Lee et al., 2003). When traffic incidents block traffic lanes, vehicles tend to decelerate and others change lanes, and the traffic stream at the upstream of a traffic incident may change from the homogeneous pattern to the stop-and-go pattern. Thus, there is an abrupt change in traffic speed before an incident and just after the incident. Therefore, traffic speed just before an incident and immediately after an incident can be used to detect incident-induced traffic disturbances.

3.1 Input Data to the Proposed Algorithm

Mean Travel Speed (MTS)

MTS is simply the average speed, calculated by Eq. (1). In the previous algorithms, the traffic data were aggregated in the range of 30 seconds to 3 minutes intervals and the threshold values were established for the past 3 to 5 minutes (Parkany et al., 2005). However, the study by Guo et al., (2018) observed lack of stability in traffic flow data for short time intervals, and 15 minutes had been suggested as the time interval to obtain stable traffic flow rates (Smith et al., 2003). Different studies have adopted 15-minutes interval as traffic flow measurement interval. One study by Haule et al., (2018) aggregated the traffic data in 15-minutes intervals to estimate the incident impact duration and the results were promising. Therefore, this study adopted 15-minute interval as the measurement interval. The speed profiles were plotted by considering the mean speed aggregated in 15-minute intervals.

Bounds of Flows

To consider the variations in the recurrent travel speed profiles, the 95% confidence interval was used to define the upper and lower bounds. Upper and lower bounds were calculated by using standard deviation approach. Standard deviation is the measure of the amount of variation of a value from the mean speed. It is simply the square root of the variance as shown on Eq. (2). The standard deviation is commonly used to measure confidence in statistical analyses. The confident limits (margin of error) are computed from the standard error of the mean or alternatively from the product of the standard deviation of the population and the inverse of the square root of the sample size, as illustrated on Eq. (3).

$$\bar{s}_t = \frac{\sum_{i=1}^{i=n} s_i}{n} \tag{1}$$

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^{i=n} (s_i - \bar{s}_t)^2}{n-1}} \tag{2}$$

$$\text{Confident interval } \bar{s}_t \pm z \frac{\sigma_t}{\sqrt{n}} \tag{3}$$

where $\bar{s}_t$ and σ_t = the mean traffic speed and standard deviate at current time interval t, respectively; s_i= the traffic speed at time i; n = the number of time intervals over the sampling period prior to the current time interval t. The Z value at 95% confidence interval (1.96), was used, hence, the upper bound and lower bound were calculated as $\bar{s}_t + 1.96 \frac{\sigma_t}{\sqrt{n}}$ and $\bar{s}_t - 1.96 \frac{\sigma_t}{\sqrt{n}}$, respectively.

3.2 Detection Logic of the Proposed Algorithm

The historical traffic flow pattern and real-time traffic flow condition just prior to a traffic incident and just after a traffic incident have significant influence on the performance of automatic incident detection algorithms. In this proposed algorithm, real-time traffic data are used for preliminary detection while historical traffic flow pattern are used for the persistence test. The persistence test is used to reduce false alarms, due to its ability to distinguish between incident and bottleneck congestion, compression shock-waves, and random fluctuations in the data. If the conditions in both the preliminary test and the persistence test are met, an incident alarm is triggered.

Preliminary test

In the preliminary test, MTS are calculated in 60-second intervals and compared with the bound of flow established in the past 15-minute interval. Traffic incidents are regarded as preliminarily detected if the MTS value calculated in 60-second intervals fall below the threshold values established in the past 15 minutes. These threshold values are different under various traffic flow conditions.

Persistence test

The persistence test is conducted to confirm the preliminarily detected traffic incident and to improve the algorithm performance. If the MTS values calculated in 60-sec interval still fall below the established threshold values using incident-free historical traffic data, an alarm is finally triggered. The threshold values used for the persistence test are established using incident-free historical traffic data, in which the bounds of normal flow are established in 15-minute intervals for each segment of roadway at each time of the day.

Figure 3.1 show a typical speed profile and bound of normal flow. The speed profiles and bound of normal flow are established on every segment of the roadway at each day of the week. The bounds of normal flow are established by the speed data aggregated over 15-minute interval. Bounds of flow (upper and lower bound) are established by using the 95% confident interval of the recurring speed. Real-time speed data aggregated over 60-second interval are compared against the established threshold value. If the aggregated speed data over 60 seconds fall below the established threshold value, an incident alarm is triggered as shown on Figure 3.2 (see the speed drop at about 13:30 hours).

Figure 3.3 summarizes the detection logic of the proposed algorithm. The right part of the chart describes the preliminary detection while the left part illustrates the persistence test. In the preliminary test the threshold values ‘bound of flow’ are established by using real-time traffic data aggregated in the past 15-minute intervals while in the persistence test the threshold values ‘bound of normal flow’ are established by using historical data aggregated over 15-minute intervals.

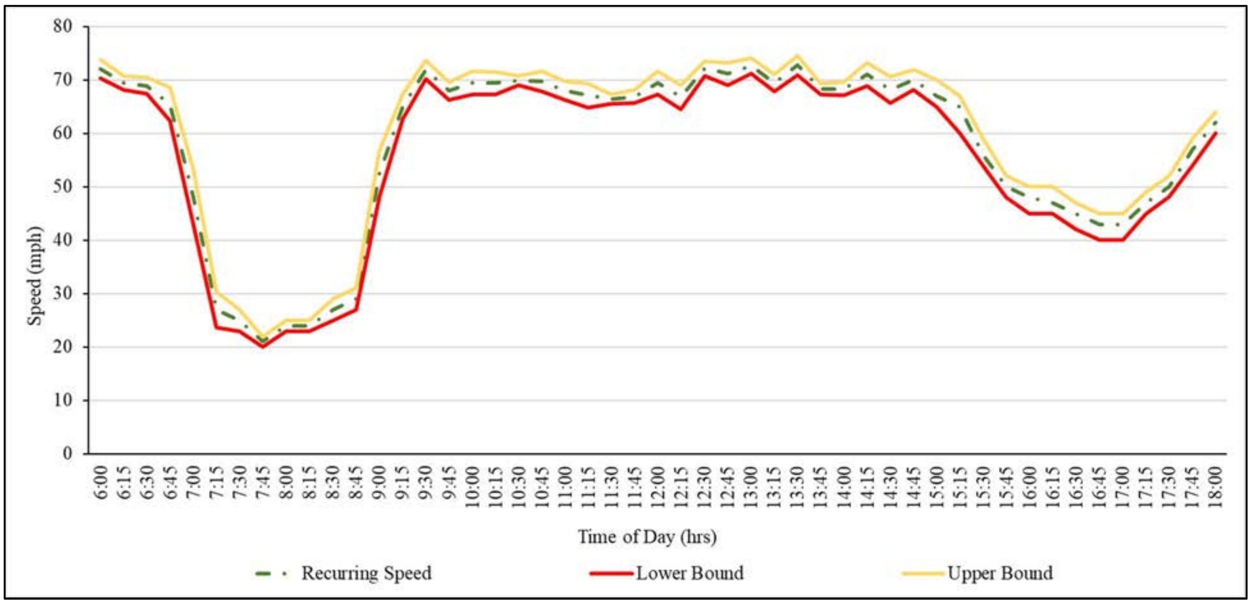

Figure 3.1: Typical speed profile ‘normal bound of flow’

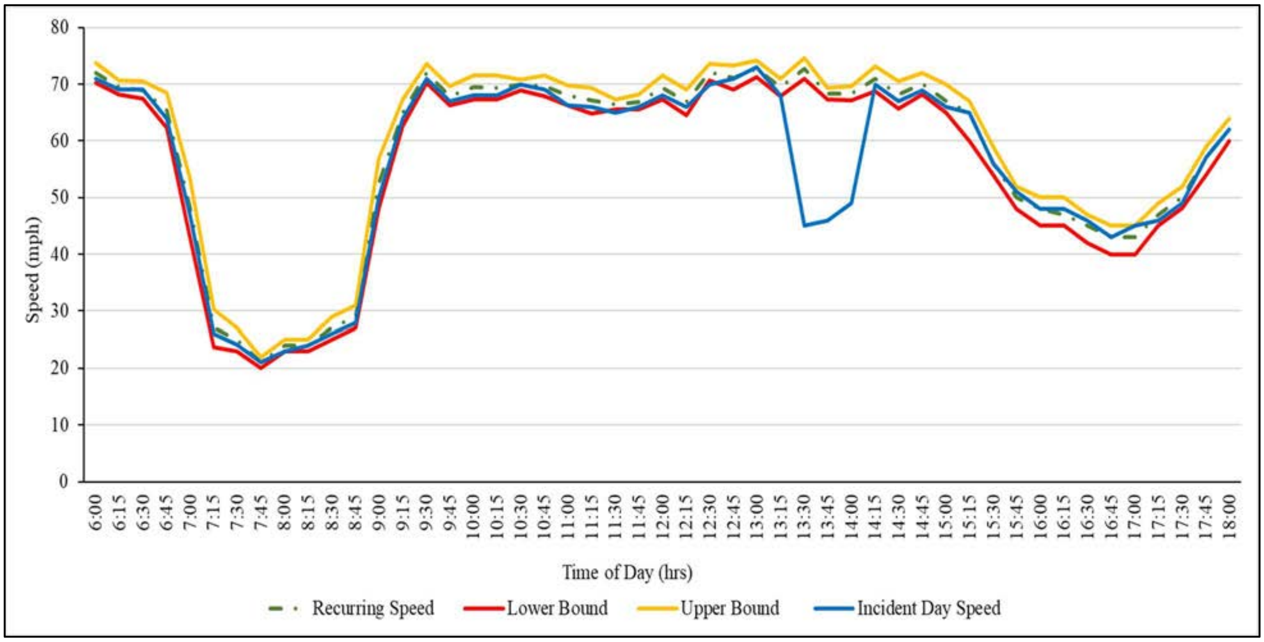

Figure 3.2: Detection of incident from normal speed profile

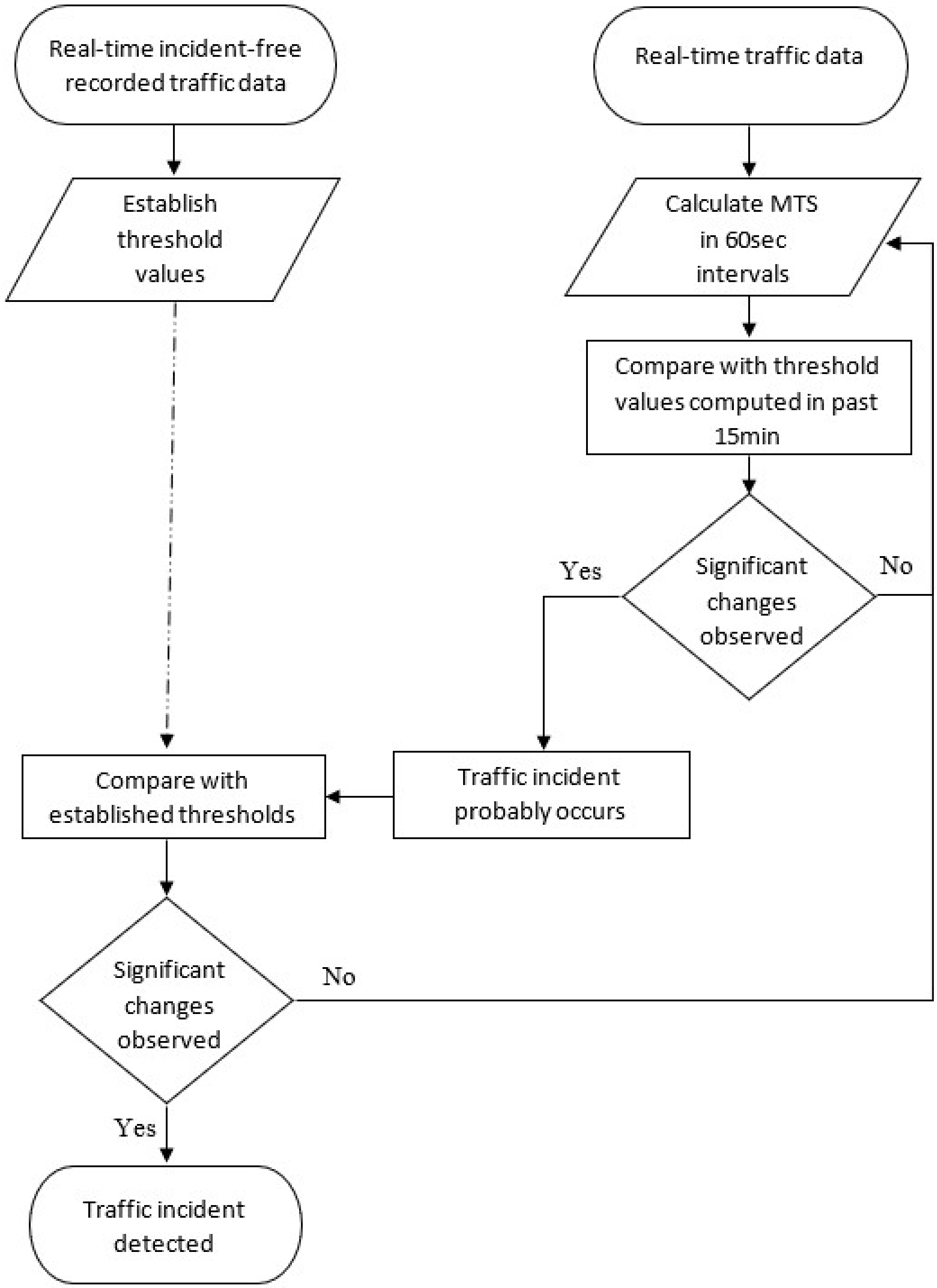

Figure 3.3: Proposed algorithm flow chart

3.3 Measures of performance

Three parameters have been used to measure the performance of incident detection algorithms. These parameters are the detection rate (DR), false alarm rate (FAR), and mean time to detect (MTTD) (Li et al., 2013). This study adopts these three parameters as the measures of the proposed algorithm's performance. The following section describes the three parameters in details.

Detection Rate (DR)

The detection rate is the ratio of the number of correctly detected incidents to the total number of incidents known to have occurred during the observation period (Jiang et al., 2010). This varies according to the definition of an incident, some studies count any stalled vehicle to be an incident, regardless of location, while others only count lane-blocking incidents. Lane-blocking incidents generally have reported higher detection rates than shoulder incidents because shoulder incidents often do not cause sufficient disruption in traffic flow to trigger an alarm (Li et al., 2013).

Mean Time to Detect (MTTD)

Mean time to detect is defined as the time from when the incident occurs until it is detected (Li et al., 2013). This does not include the time taken to verify the incident.

False Alarm Rate (FAR)

The false alarm rate is most often defined as the percentage of incorrect detection signals relative to the total number of algorithm decisions (Martin et al., 2001). This is calculated to determine how many incident alarms were falsely set.

CHAPTER 4 STUDY SITE AND DATA COLLECTION

4.1 Study Site

The Lee Roy Selmon Expressway was chosen for this study (Fig. 4.1) because of the THEA CV data availability. Lee Roy Selmon Expressway, also known as State Road 618, is a 15-mile limited access toll road in Hillsborough county, Florida. It connects the South Tampa neighborhood near MacDill Air Force Base with downtown Tampa and Brandon. For most of its length, the expressway has four lanes (two in each direction) with a small portion of a six-lane segment (three lanes in each direction). The posted speed limit of this road section ranges from 55 to 70 mph with one-direction average daily traffic of about 76,275 vehicles (THEA, 2019).

Figure 4.1: Tampa-Hillsborough Expressway, Tampa, Florida (THEA, 2018)

4.2 Traffic Data

Real-time recorded traffic data were the main input for this study. These data were retrieved from the Regional Integrated Transportation Information System (RITIS) database and THEA CV pilot study for the entire study period from February, 2019 to June, 2019. The following sections provide further details of data sources;

RITIS

RITIS is a data sharing, dissemination, and archiving system that includes real-time data feeds and data analysis tools such as probe, detector, and transit data analytics. From the RITIS database, HERE traffic speed data were collected. HERE GPS technology collect speed data by recording the time when a vehicle passes the known segment, then this information is used to deduce the travel time of the vehicle on that segment and the speed is calculated from the obtained travel time and a known segment length. The following information were retrieved from the HERE data; traffic speed, date and time, segment location and segment length. Real-time traffic speed data were extracted in 1-minute intervals for an entire Selmon Expressway with a total of 33 data collection segments. The 33 data collection segments used have different lengths, with the maximum segment length of 1.82 miles and minimum segment length of 0.01 miles.

THEA CV Data

Tampa-Hillsborough Expressway Authority (THEA) CVs Pilot study generates data from the interaction between vehicles (V2V) and between vehicles and infrastructure (V2I). The dataset consists of BSMs generated by vehicle onboard units (OBUs) and transmitted to road-side units

(RSUs) located throughout the study area. The CV technologies collect point-to-point traffic data over a roadway section.

A Python script was used to download the THEA CV data from USDOT ITS datahub data sandbox stored in AWS S3. THEA CV dataset has about 54 variables, but this study uses the following key data variables: datasets contain data generation (date, time and vehicle id), geographic position (latitude, longitude, and elevation) and motion (speed and acceleration). Data quality check was performed before using the datasets for further analysis. Python was used for data manipulation and ArcGIS software was used for spatial analysis. In the first step of the data cleaning procedure, duplicated records were removed. Unavailable fields and erroneous data points were removed based on the criteria given on metadata, and lastly all observations were formatted and converted into required units. Then the ArcGIS software was used for mapping and sorting out observations. The buffer tool and selection by location tool were used to sort observations at the selected study corridor as shown in Figure 4.2. The section of Selmon Expressway in which CV pilot study was deployed starts from South Franklin Street to Channelside Drive and is a total of 1.12 miles.

Figure 4.2: ArcGIS geoprocessing and selection by location process

The CV one-direction average daily traffic (CV ADT) was calculated based on average trips generated for the entire study period and the value of 4924 vehicles was obtained. The penetration rate of CVs was calculated by comparing the CV directional ADT and the Selmon Expressway directional ADT (76,275 vehicles) recorded near the study corridor (at West Plaza Group Gantry). Based on those figures, an estimated CV penetration rate of 6.46% was obtained.

4.3 Incident Data

Traffic incident data were required for testing and comparison of the proposed algorithm performance. Incident data were retrieved from Signal Four Analytics (crashes only) and Road Rangers reports for the entire study period from February, 2019, to June, 2019. Due to the Road

Ranger operating hours, the incident data were collected only from Monday through Friday, 6:00 am through 6:00 pm. The following sections provide further details of incident data sources;

Signal Four Analytics

Signal Four Analytics is a statewide interactive, web-based geospatial crash analytical tool in the state of Florida. Crash data were extracted from this tool and the ArcGIS software was used for mapping and data reduction. A total of 60 crashes were found to fall within study site for the study period. Crash occurrence time, location, levels of severity namely; fatal, injury, and property damage only (PDO), and other characteristics were extracted from the crash database.

Road Ranger Reports

The Road Ranger Service Patrol in Florida is a freeway service patrol that provides free highway assistance services to motorists. The Road Rangers provide assistance to motorists by quickly clearing traffic incidents. Services can include providing a limited amount of fuel, assisting with tire changing and other types of minor emergency repairs, and providing support at crash sites. This study extracted traffic incident data from the road ranger reports. A total of 292 incidents were reported by Road Rangers for the study period within the study area. Time of arrival to the scene, location, levels of severity, and other characteristics were included in the incident reports.

Table 4.1 gives the summary of the collected incident data. Reporting sources include Road Ranger and Signal Four Analytics (one incident may be reported by both agencies). Incident type was categorized into crashes, vehicle problems (disabled or abandoned vehicles, emergency vehicles, fire and police activity), and traffic hazards (debris, flooding, and spillage). Incident severity was categorized into minor, and moderate/severe. Lane closure included left/right lane blockage, and shoulder/median accidents included all incident that were not lane-blocking. Peak hours included morning peak (0600 to 1000 hours) and evening peak (1500 to 1800 hours).

Table 4.1: Descriptive statistics of incident data

Incident Attributes	Categories	Frequency	Percentage (%)
Reporting Agency	Road Rangers	292	89.30
	Signal Four Analytics	60	18.35
	Both	25	7.65
Incident type	Crash	90	27.52
	Vehicle problems	210	64.22
	Traffic hazards	27	8.26
Severity	Minor	231	70.64
	Moderate/Severe	96	29.36
Lane closure	Lane-blockage	212	64.83
	Shoulder/median	115	35.17
Time of the day	Peak hours	187	57.19
	Off-peak hours	140	42.81

The traffic and accident data were divided into two groups in order to test the proposed AID algorithm. One group consist of the HERE data and all the recorded incidents summarized in Table

4.1 and another group consists of the THEA CV data and incidents that were only found to fall within CV pilot study corridor. The first group comprises of a total of 327 incidents while the second group comprises of 41 incidents. Table 4.2 gives the summary of the incidents that found to fall within the CV pilot study corridor.

Table 4.2: Descriptive statistics of incident data

Incident Attributes	Categories	Frequency	Percentage (%)
Reporting Agency	Road Rangers	33	80.49
	Signal Four Analytics	12	29.27
	Both	04	9.76
Incident type	Crash	19	46.34
	Vehicle problems	17	41.46
	Traffic hazards	05	12.20
Severity	Minor	22	53.66
	Moderate/Severe	19	46.34
Lane closure	Lane-blockage	26	63.41
	Shoulder/median	15	36.59
Time of the day	Peak hours	32	78.05
	Off-peak hours	09	21.95

CHAPTER 5 ANALYSIS AND RESULTS

5.1 Testing of the Proposed Algorithm using HERE Data

HERE speed data were used for testing the proposed algorithm. Real-time speed data was extracted for the entire Selmon Expressway from a total of 33 data collection segments. Speed data were extracted in 1-minute intervals from February, 2019, to June, 2019. The analysis included all incidents that were found to fall within the study corridor for the entire study period. A total of 327 incidents were used in this section.

The first part of the analysis was to obtain incident-free days as indicated in Table 5.1. These were days that did not have any recorded incidents in the selected study period for a duration of 6 am to 6 pm.

Table 5.1: Incident-free days

Day	Date
Monday	Feb 04, April 29, May 13, June 17
Tuesday	March 12, March 19, May 07
Wednesday	Feb 06, March 06, April 10, April 24, May 01
Thursday	March 14, March 28, April 18, May 23, May 30
Friday	Feb 01, May 17, June 07

Establish the Bounds of Normal Flow

The historical traffic data recorded during the incident-free days were used to establish the threshold values. The speed data were aggregated in 15-minute intervals. The 95% confidence interval of the mean speed was calculated to define the upper and lower bounds of the recurrent speed profile. This 95% confidence interval accounted for the variation in speeds on a roadway segment. For each segment, a total of five speed profiles were generated, one for each day of week (Monday through Friday). The speed profiles were established for every segment of the expressway at each time of the day from Monday through Friday. The mean travel speed (MTS) values were calculated in 15-minute intervals, and the upper and lower bounds of the recurrent speed profile was established and adopted as the threshold value.

The python programming language was used to implement the proposed algorithm. The results of the analysis are summarized in Table 5.2.

Table 5.2: Performance of the proposed algorithm using HERE data

Incident Attributes	Categories	Total	Detected	Detection rate (DR) %
Time of the day	Peak hours	216	139	64.35
	Off peak hours	111	93	83.78
Lane closure	Lane blocking	212	212	100.00
	Shoulder	115	20	17.39
Incident type	Crash	90	84	93.33
	Vehicle problem	210	131	62.38
	Hazards	27	17	62.96
Overall performance		327	232	70.95

Figure 5.1 summarizes the algorithm's performance by using HERE data. The chart illustrates the algorithm's DR under different incident attributes.

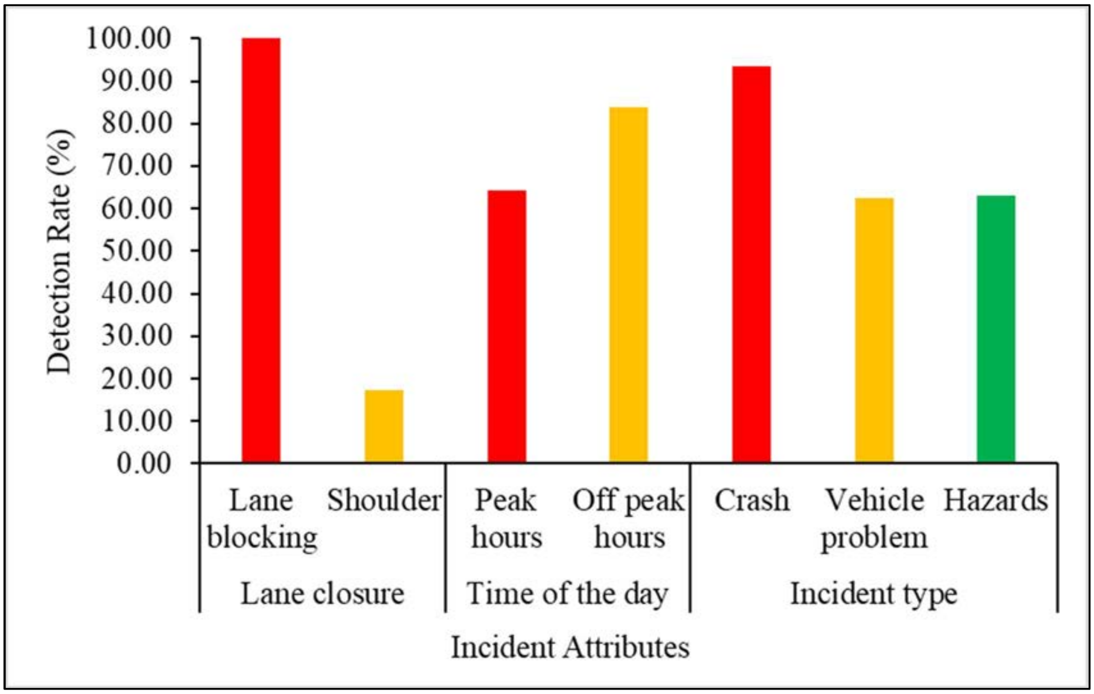

Figure 5.1: Algorithm's performance using HERE data

Results of the Algorithm Performance using HERE Data

The overall performance of the proposed algorithm shows the DR of 70.95% and FAR of 9.02% using traffic data obtained from the HERE database. Moreover, the proposed algorithm detects 64.35% of all incidents that occurred during peak hours and 83.78% of all incidents that occurred during off peak hours. Also, the algorithm detected 93.33% of all crashes, 62.38% of all vehicle problems and 62.96% of all hazards. Again, the algorithm detected 100% of all lane-blocking incidents and 17.39% of all shoulder incidents. This finding is consistent with results reported by previous studies that observed high DR when considering lane-blocking incidents only. Lane-blocking incidents have generally reported higher detection rates than shoulder incidents because shoulder incidents often do not cause sufficient traffic flow disruption to trigger an alarm (Li et al., 2013).

This study faced challenges in analyzing MTTD, as the reported time is sometimes not the same as the real incident occurrence time. Road Ranger reports did not include incident occurrence time, while Signal Four Analytics data have incident occurring time. Nevertheless, the analysis shows that the reported occurrence time may not be the actual incident occurrence time. Surprisingly, with respect to crashes extracted from Signal Four Analytics Database, more than 60% had recorded occurrence times that were rounded to the nearest 5 minutes. This imply that the personnel reporting these incidents may not report actual time. Due to this observation, the MTTD was considered unreliable.

5.2 Testing of the Proposed Algorithm using THEA CV Data

The proposed algorithm was also tested by using THEA CV generated data. This section included only the incidents that were found to fall within the CV pilot study corridor for the entire study period. A total of 41 incidents were used in this section.

The same incident-free days used for the previous section were also adopted in this section. To simplify the analysis, the minimum segment length of 0.01 mile used for HERE data was also adopted for CV data. Therefore, the study corridor was divided into 112 equal sections with the length of about 0.01 mile. The speed profiles were established for every section at each time of the day from Monday through Friday. The mean travel speed (MTS) values were calculated in 15-minute intervals, and the upper and lower bounds of the recurrent speed profile were established and adopted as the threshold value.

The python programming language was used to implement the proposed algorithm. The results of the analysis are summarized in Table 5.3.

Table 5.3: Performance of the proposed algorithm using THEA CV data

Incident Attributes	Categories	Total	Detected	Detection rate (DR) %
Time of the day	Peak hours	32	11	34.38
	Off peak hours	9	2	22.22
Lane closure	Lane blocking	26	10	38.46
	Shoulder	15	3	20.00
Incident type	Crash	19	8	42.11
	Vehicle problem	17	3	17.65
	Hazards	5	2	40.00
Overall performance		41	13	31.71

Figure 5.2 summarizes the algorithm's performance by using CV data. The graph illustrates the algorithm's DR under different incident attributes.

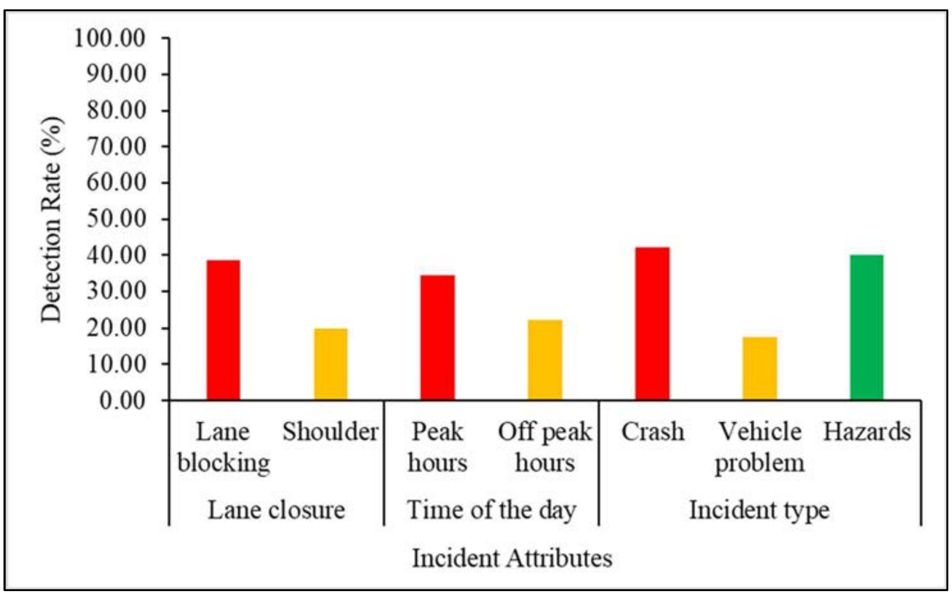

Figure 5.2: Algorithm's performance using CV data

Results of the Algorithm Performance using CV Data

The proposed algorithm shows DR of 31.71% and FAR of 0.00% using traffic data obtained from the HERE database. Moreover, the proposed algorithm detects 34.38% of all incidents that occurred during peak hours and 22.22% of all incidents that occurred during off peak hours. Also, it detected 42.11% of all crashes, 17.65% of all vehicle problems and 40.00% of all hazards. Again, the algorithm detected 38.46% of all lane-blocking incidents and 20.00% of all shoulder incidents. However, the market penetration rate was very low, study findings show high DR when considering lane-blocking incidents only, which is consistent to the findings of the previous studies (Li et al., 2013).

5.3 Algorithm Performance Comparison Between HERE Data and CV Data

Comparison Between DR and FAR

The algorithm performance in terms of DR and FAR indicated that CV data resulted into 31.71% DR and zero FAR while HERE data gave 70.95% DR and 9.02% FAR. Figure 5.3 illustrates the algorithm's performance difference between two data sources. Although the CV penetration rate was low (6.46%), the generated data provide promising results in terms of automated freeway incident detection. This means that once the CV market penetration increases, CVs will generate data that will provide reliable results for automated freeway detection. The fact that CV data resulted into zero FAR proves the accuracy of CV data compared to HERE data. Hence, there is high potential of using CVs in incident management.

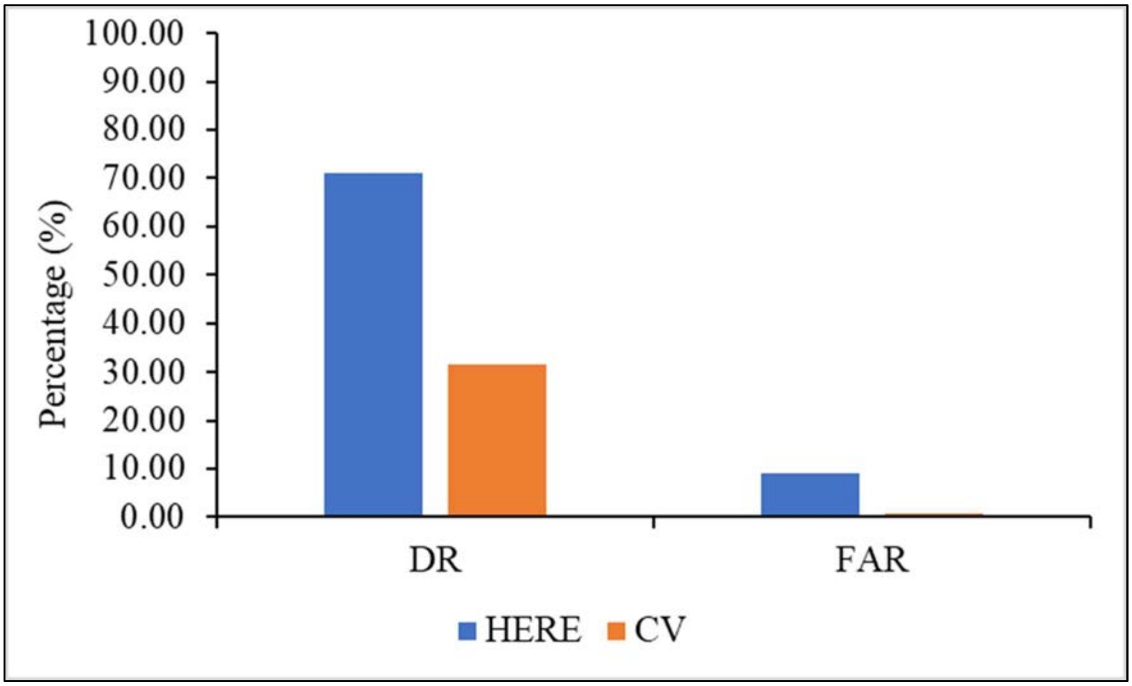

Figure 5.3: Algorithm's performance comparison between HERE data and CV data

Correlation Analysis of the Algorithm's Detection Times

Statistical comparison between algorithm's detection time using HERE data and CV data was conducted. The statistical analysis was conducted to observe the correlation between the two detection times. A correlation coefficient depicts the strength of the link between two quantitative variables. Since data consist of continuous variables, the Pearson correlation coefficient was adopted and the test was conducted at 95% significance level. The test was conducted to determine whether there is a statistical relationship between the algorithm's detection time using HERE data and CV data. Correlation significance tests were also performed and the Pearson's correlation coefficient between algorithm's detection time using HERE data and CV data was calculated.

Results of Correlation Analysis

The Pearson's correlation coefficient between the two detection times was found to be 0.99 and the p-values of the Pearson's correlation coefficient was found to be less than 0.001. Figure 5.4 shows a scatter plot of the correlation between the two detection times, as shown on the graph, there is high correlation between the algorithm's detection time using HERE data and CV data. Therefore, the incidents detected by the CV data were similar to the one detected by the HERE data, despite of the lower CV market penetration rate.

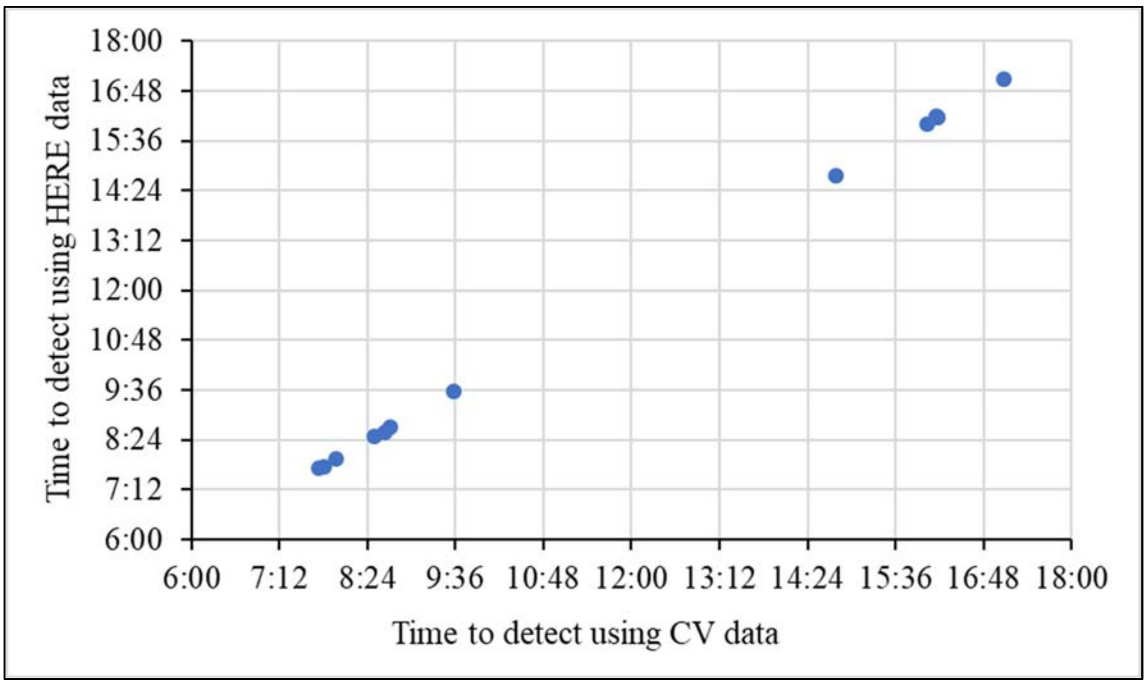

Figure 5.4: Pearson's correlation scatter plot

Algorithm's Detection Time Difference

The algorithm performance in term of difference between the two detection times was also analyzed. The analysis was performed to determine how faster algorithm performs when using one data source compared to another. To achieve this, algorithm's detection time difference was calculated by taking algorithm's detection time using HERE data minus CV data. The results were recorded as summarized in Figure 5.5. The positive time difference implies algorithm detect an incident earlier when using CV data than HERE data and the negative difference implies the vice versa. Zero-time difference implies algorithm's detection time is the same for both CV data and HERE data. Based on recorded data shown in Figure 5.5 and summarized in Table 5.4, the proposed algorithm can detect an incident 2 minutes earlier when using CV data compared to HERE data, also the proposed algorithm can give 2 minutes delay in detection when using the same data compared to HERE data. This means that the proposed algorithm can either detect an incident faster or delays when using CV data as compared to HERE data and vice versa is true. Moreover, algorithm shows no difference in detection most of time, this implies that the proposed algorithm can also give same result when using either CV or HERE data.

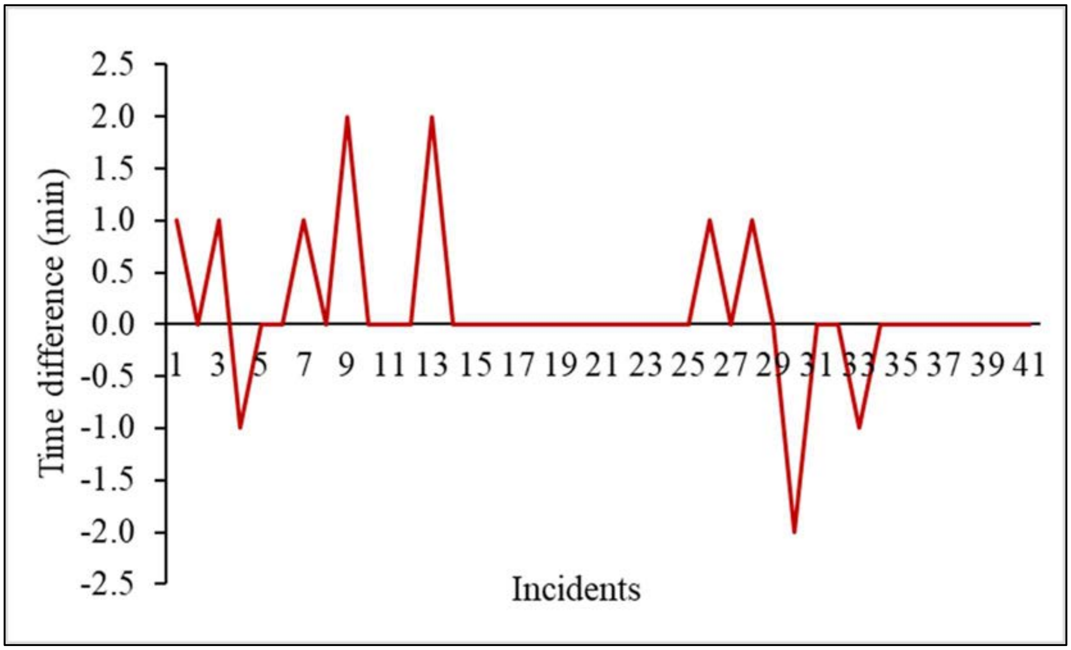

Figure 5.5: Algorithm's detection time difference between CV data and HERE data

Table 5.4: Summary of Statistics

Variable	Observations	Obs. with missing	Obs. without	Minimum	Maximum	Mean	Std. deviation
CV	41	0	41	0.000	2.000	0.220	0.525
HERE	41	0	41	0.000	2.000	0.098	0.374

The hypothesis test was also conducted with null hypothesis (Ho): there is no difference between the two detection times and alternative hypothesis (Ha): there is difference between the two detection times.

Null hypothesis: H_0: μ_difference = 0

Alternative hypothesis: H_1: μ_difference $\neq$ 0

Where: μ_difference = difference between algorithm's detection time using HERE and CV data.

Since the test involves the comparison between the two detection times, analysis of variance (ANOVA) was adopted. ANOVA is a collection of statistical models and their associated estimation procedures used to analyze the differences among group means in a sample. It provides a statistical test of whether two or more population means are equal. The results of analysis are summarized in Table 5.5.

Table 5.5: Analysis of Variance

Source	DF	Sum of squares	Mean squares	F-Value	P-Value
Model	1	1.737	1.737	7.296	**0.010**
Error	39	9.287	0.238		
Corrected	40	11.024			

Results of the Detection Time Difference

Since the p-value (0.01) is small than significance level (0.05) then the null hypothesis is rejected, and hence there is difference between the two detection times. The results show that the proposed algorithm can detect 17.07% of all incidents earlier when using CV data compared to HERE data. This is due to the reason that CV data are point-point hence more accurate than the HERE data which are collected as average segment data. Although the proposed algorithm shows earlier detection when using CV data, it also shows delay in detection. About 7.32% of all incidents were detected earlier by HERE data when compared to CV data. The reason might be the small market penetration rate of CVs.

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

This study presents a newly developed automated freeway incident detection (AID) algorithm based on the available data collected in Tampa, Florida. The proposed algorithm was tested by using THEA CV data and HERE data in conjunction with incidents data obtained form Signal Four Analytics and Road Rangers reports. The study was based on Selmon expressway due to the CV data availability.

Based on the algorithm's performance, by using THEA CV data the proposed algorithm detected 31.71% of the traffic incidents and no false alarm was reported. Again, by using HERE data the proposed algorithm detected 70.95% of the traffic incidents. The false alarm rate (FAR) was about 9.02%, which equal to one false alarm per month. Although the DR by using CV data seemed to be very small compared to the one obtained from HERE data, the results are promising due to small CV market penetration rate (6.46%). The fact that CV data resulted into zero false alarm rate, this prove that the connected vehicles data are more accurate as they are point-to-point records compared to HERE data.

An automated freeway incident detection algorithm only can be effective when flows are heavy enough that traffic is substantially interrupted by an incident. Incidents that occur during low traffic volume and shoulders incidents do not usually cause enough disturbance to the flow to be recognized. This is acknowledged by several studies (Li et al., 2013; Parkany et al., 2005) and consequently leave out all shoulder incidents from the reported detection rates. Considering lane-blocking incidents only, the proposed algorithm gave DR of 100% by using HERE data and

38.46% using THEA CV data. Hence proposed algorithm shows high performance under lane-blocking incidents as they cause high traffic disturbance compared to shoulder incidents.

The Pearson's correlation analysis show that the incidents detected by the CV data were similar to the one detected by the HERE data, despite of the low CV market penetration rate. Moreover, the performance comparison of the proposed algorithm's detection times using ANOVA shows that there is a difference in the algorithm's detection time when using CV data and HERE data. About 17.07% of all incidents were detected earlier when using CV data compared to HERE data, while 7.32% were detected earlier when using HERE data compared to CV data. Furthermore, in 75.61% of all incidents the proposed algorithm shows the same detection time when using the two data sources, CV and HERE data.

Although the CV market penetration rate was low, the algorithm performance as tested using real-time CV data provides a promising result. The algorithm shows low DR due to missing data caused by the low CV market penetration rate. Data availability was challenging as there were missing data in a lot of days and also in most of times. The proposed algorithm gave no FAR, perhaps because of the accuracy of the CV data, when available. Therefore, the proposed algorithm has good potential for application of automated freeway incident detection by using CV data.

Since the market penetration rate is low and CV technology will take time until high penetration rate for full detection rate, the proposed algorithm can be used with other backed up technologies like WAZE, CCTV cameras, highway patrol services and other ITS technologies.

Findings from this thesis will not only add to the body of knowledge related to CV technology, but also will be useful to transportation agencies as the CV technology becomes increasingly more utilized in transportation management strategies. Traffic engineers and other traffic related agencies may also apply the research findings in planning and traffic operations efforts that involve CV technology.

This study has several limitations that are worth mentioning. First, the CV market penetration rate was low, hence data collection and availability were challenging. The second limitation relates to a short study period (only 5 months), which resulted in having only a small number of incidents for analysis. Lastly, actual incident occurrence time was a significant limitation. Road Ranger reports that were obtained from THEA did not include this attribute (showed only the time Road Ranger arrived on scene) hence the study faced a challenge in calculating algorithm's mean time to detect incident. Lastly, further research is suggested to validate the proposed algorithm when the CV market penetration rate is increased, and more incident data are collected.

REFERENCES

Amin-Naseri, Chakraborty, Sharma, Gilbert, & Hong. (2017). Evaluating the reliablity, coverage, and added value of crowdsourced traffic incident reports from Waze. *Transportation Research Part C: Emerging Technologies*.

Baykal-Gürsoy, M., Xiao, W., & Ozbay, K. (2009). Modeling traffic flow interrupted by incidents. *European Journal of Operational Research*, *195*, 127–138. https://doi.org/10.1016/j.ejor.2008.01.024

Bernstein, D., & Kanaan, A. Y. (1993). Automatic vehicle identification: technologies and functionalities. *IVHS Journal Print*, *1*(2), 191–204. https://doi.org/10.1080/10248079308903792

Corby, M. J., & Saccomanno, F. F. (1997). Analysis of freeway accident detection. *Transportation Research Record*, *1603*(1603), 80–89. https://doi.org/10.3141/1603-11

Hellinga, B., & Knapp, G. (2000). Automatic vehicle identification technology-based freeway incident detection. *Transportation Research Record*, (1727), 142–153. https://doi.org/10.3141/1727-18

Henclewood, D., & Rajiwade, S. (2015). *Data capture and management programm: Datasets task, Safety Pilot Model Deployment (SPMD)-Sample Data Environment*. 1–126.

Iqbal, M. S., Khazraeian, S., & Hadi, M. (2018). A Methodology to Assess the Quality of Travel Time Estimation and Incident Detection Based on Connected Vehicle Data. *Transportation Research Record*, *2672*(42), 203–212. https://doi.org/10.1177/0361198118773199

Jiang, G., Niu, S., Li, Q., & Chang, A. (2010). Automated incident detection algorithms for urban expressway. *Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010*, *3*, 70–74. https://doi.org/10.1109/ICACC.2010.5486740

KARLAFTIS, M. G., LATOSKI, S. P., RICHARDS, N. J., & SINHA, K. C. (1999). ITS Impacts on Safety and Traffic Management: An Investigation of Secondary Crash Causes. *ITS Journal - Intelligent Transportation Systems Journal*, *5*(1), 39–52. https://doi.org/10.1080/10248079908903756

Lee, C., Hellinga, B., & Saccomanno, F. (2003). Proactive freeway crash prevention using real-time traffic control. *Canadian Journal of Civil Engineering*, *30*(6), 1034–1041. https://doi.org/10.1139/l03-040

Li, X., Lam, W. H. K., & Tam, M. L. (2013). New Automatic Incident Detection Algorithm Based on Traffic Data Collected for Journey Time Estimation. *Journal of Transportation Engineering*, *139*(8), 840–847. https://doi.org/10.1061/(asce)te.1943-5436.0000566

Liu, J., & Khattak, A. J. (2016). Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles. *Transportation Research Part C: Emerging Technologies*, *68*, 83–100. https://doi.org/10.1016/j.trc.2016.03.009

Mahmassani, H., Haas, C., Zhou, S., & Peterman, J. (1998). Evaluation of incident detection methodologies. *Fhwa/Tx-00/1795-1*, 1. Retrieved from https://pdfs.semanticscholar.org/a406/fda93354cce4c28e78a82436ebb87faabec4.pdf%0Ahttps://ctr.utexas.edu/wp-content/uploads/pubs/1795_1.pdf%0Ahttp://medcontent.metapress.com/index/A65RM03P4

874243N.pdf%5Cnhttp://www-math.cc.utexas.edu/research/ctr/pdf_repor

Marshall, D., Author, C., Wu, X., Brown, T., & Ng Boyle, L. (2016). *Connected vehicle auditory alerts: One size doesn't fit all scenarios*.

Martin, P. T., Perrin, J., Hansen, B., Kump, R., & Moore, D. (2001). Incident Detection Algorithm Evaluation (MPC-01-122). *Neural Networks*, (March). Retrieved from https://www.ugpti.org/resources/reports/downloads/mpc01-122.pdf

NHTSA. (2016). USDOT Releases 2016 Fatal Traffic Crash Data. Retrieved November 26, 2019, from https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., & Trego, T. (2010). Traffic Incident Management Handbook. In *Science Applications International Corporation American Transportation Research Institute*. https://doi.org/10.1093/her/cys022

Owens, N. D., Armstrong, A. H., Mitchell, C., Brewster, R., & Administration, U. S. F. H. (2009). *Federal Highway Administration Focus States Initiative: Traffic Incident Management Performance Measures Final Report*. (December).

Parkany, E., & Xie, C. (2005). A Complete Review of Incident Detection Algorithms & Their Deployment: What Works and What Doesn't. *The New England Transportation Consortium*, (00), 120.

Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2015). *2015 URBAN MOBILITY SCORECARD*.

Smith, B. L., & Ulmer, J. M. (2003). Freeway Traffic Flow Rate Measurement : Investigation into Impact of Measurement Time Interval. *Journal of Transportation Engineering*, *129*(June), 223–229. https://doi.org/https://doi.org/10.1061/(ASCE)0733-947X(2003)129:3(223)

Stephanedes, Y. J., Chassiakos, A. P., & Michalopoulos, P. G. (1992a). Comparative Performance Evaluation of Incident Detection Algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, (1360), 50–57. Retrieved from http://onlinepubs.trb.org/Onlinepubs/trr/1992/1360/1360-009.pdf

Stephanedes, Y. J., Chassiakos, A. P., & Michalopoulos, P. G. (1992b). Comparative Performance Evaluation of Incident Detection Algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, (1360), 50–57. Retrieved from https://www.researchgate.net/publication/284930603_Comparative_performance_evaluation_of_incident_detection_algorithms

THEA. (2018). THEA Connected Vehicle Pilot. Retrieved November 20, 2019, from https://www.tampacvpilot.com/

THEA. (2019). *2018 ANNUAL REPORT*. Retrieved from https://www.tampa-xway.com/wp-content/uploads/2019/04/TEX-497-Annual-Report-V8.pdf

Waze. (2006). Free Driving Directions, Traffic Reports & GPS Navigation App by Waze. Retrieved March 30, 2019, from https://www.waze.com/about

Xie, K., Di Yang, Ozbay, K., & Yang, H. (2018). Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2018.07.002

Yang, H., Wang, Z., & Xie, K. (2017). Impact of connected vehicles on mitigating secondary crash risk. *International Journal of Transportation Science and Technology*, *6*, 196–207. https://doi.org/10.1016/j.ijtst.2017.07.007

VITAE

HENDRY IMANI

Education

University of North Florida**,** Jacksonville FL

- Master of Science in Civil Engineering, December 2019

University of Dar es Salaam, Dar es Salaam, Tanzania

- Bachelor of Science in Civil Engineering, November 2018

Related Work Experience

University of North Florida, Jacksonville, Florida. January 2019-December 2019

Graduate Research Assistant

1. Collecting, processing, interpreting, analyzing, and compiling traffic data for research projects.
2. Microscopic modeling of safety applications of connected vehicles.

Teaching Assistant

1. Civil Engineering Geomatics, Spring 2019
2. Advanced Research Methods for Engineers, Fall 2019.

Campus Involvement and Volunteer Experience

1. Institute of Transportation Engineers, UNF Chapter Media
2. College tours at University of North Florida School of Engineering
3. Sorting and packing food items at Feeding Northeast Florida (FNEF), Jacksonville FL