RISK ANALYSIS USING ARTIFICIAL INTELLIGENCE ALGORITHMS TO PREVENT COLLISIONS ON ROADWAY SEGMENTS

AHMAD MOHAMMADI

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CIVIL ENGINEERING

YORK UNIVERSITY

TORONTO, ONTARIO

September 2022

ABSTRACT

This thesis focused on improving the risk analysis algorithms used in collision avoidance systems (CASs) designed to reduce the risk of three types of collision on roadway segments: animal-to-vehicle collisions, pedestrian-to-vehicle collisions, and pedestrian-to-pedestrian collisions.

Currently available CASs use only one input indicator. This approach is limited as the CASs: apply a simple risk analysis algorithm based on a fixed threshold to identify risky situations; cannot simultaneously capture a variety of important collision contributing factors; and cannot combine multiple contributing factors into a single composite risk indicator.

The goal of this thesis was to use artificial intelligence algorithms to create a composite risk indicator based on a combination of various input indicators.

The thesis goal was achieved through four objectives:

1) Develop a fuzzy rule-based algorithm for a next generation roadside animal detection system;
2) Develop a fuzzy rule-based algorithm for a smart protection system to reduce the number of collisions with police officers on duty on the roadway;
3) Develop a semi-supervised machine learning algorithm for a smart protection system to reduce the number of collisions with police officers on duty on the roadway; and
4) Develop a risk analysis approach to evaluate physical distancing on urban sidewalks.

Improvement of the existing risk analysis algorithm in objective 1 resulted in capturing driver behavior, animal behavior, and the spatial and temporal interaction between animal and vehicle. It also resulted in differentiating risk for following and leading vehicle and generating no-risk when vehicle passed from animal.

Objectives 2 and 3 were part of the same CAS study. Improvement of the existing risk analysis algorithm in both objectives 2 and 3 resulted in capturing pedestrian behavior, driver behavior, the spatial and temporal interaction between pedestrian and vehicle with 94% accuracy when estimating all risk labels, and 88% success when identifying near miss collisions.

Objective 4 successfully reflected the role of density and exposure time in the level of physical distancing. It could help decision-makers to select the most appropriate interventions (e.g., sidewalk expansion) for pedestrians to maintain physical distancing.

ACKNOWLEDGEMENTS

First and foremost, I would like to praise and thank God for granting me power and energy to accomplish this research. I would like to express my deepest appreciation and gratitude to my thesis supervisor, Dr. Peter Park whose invaluable advice, continuous support, patience, and encouragement has been invaluable throughout this research. His immense academic knowledge and plentiful work experience have encouraged me in all the time of my academic research and daily life.

Words can not express how grateful I am to my beloved wife, Parisa who supported me emotionally and thoughtfully during this journey. Parisa has been extremely supportive of me throughout this entire process and has made countless sacrifices to help me get to this point.

I am extremely grateful to my examiners, Dr. Kevin Gingerich, Dr. Ali Asgary, Dr. Chris Lee., and Dr. Andrew Eckford for engaging with my thesis enthusiastically and diligently, providing valuable suggestions. Special thanks to the A.U.G. Signals Ltd, specially, Abir Mukherjee and Xia Liu for providing data used in this research. I would also like to thank the Defense Research and Development Canada and the Natural Sciences and Engineering Research Council for their financial and other support for this study. Thanks to financial support from the Build in Canada Innovation Program, and additional support was provided by Dr. Sarah Boyle of Mount Revelstoke and by Glacier National Park, Parks Canada. I would also like to extend my deepest gratitude to my parents and siblings, deserve special thanks for their continued support and encouragement. I am grateful to Tanvir Chowdhury and Dr. Seungho Yang for working as a team over the summer on a COVID-19 related topic. Furthermore, I would like to say thank to Nesta Morris for providing valuable editing and writing services for this thesis. Lastly, I must also say thank to Sindy Mahal and Gillian Moore for answering all my questions during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AASHTO	American Association of State Highway and Transportation Official
AI	Artificial Intelligence
BC MoTI	British Columbia Ministry of Transportation and Infrastructure
CAS	Collision Avoidance System
CDC	Centers for Disease Control and Prevention
COVID-19	Coronavirus Disease of 2019
ECDC	European Center for Disease Prevention and Control
FBI	Federal Bureau of Investigation
GPS	Global Positioning System
HCM	Highway Capacity Manual
LED	Light-Emitting Diode
LPPD	Level of Pedestrian Physical Distancing
NACTO	National Association of City Transportation Officials
NG RADS	Next Generation Roadside Animal Detection System
NHSTA	National Highway Traffic Safety Administration
PET	Post Encroachment Time
PCL	Pedestrian Comfort Level
PLOS	Pedestrian Level of Service
PS	Pedestrian Space
RADS	Roadside Animal Detection System
RADAR	Radio Detection And Ranging
SSD	Stopping Sight Distance

TAC	Transportation Association of Canada
TCT	Traffic Conflict Technique
TRB	Transportation Research Board
TTC	Time-To-Collision
V2I	Vehicle-To-Infrastructure
WHO	World Health Organization

CHAPTER ONE

1. INTRODUCTION

According to the World Health Organization (WHO, 2018), an average of 1.35 million people worldwide were killed in traffic collisions each year from 2000 to 2016. Unfortunately, traffic collisions are now the leading cause of death in the world for children and young adults aged 5-29 years exceeding HIV/AIDS (the second leading cause), tuberculosis (the third leading cause), and diarrhoeal diseases (the fourth leading cause). In the United States in 2020, traffic collisions were the fourteenth leading cause of death in 2020 (38,824) (CDC, 2020; NHSTA, 2021). In Canada, traffic collisions were the sixteenth leading cause of death in 2020 (with 1,745 death) (Statistics Canada, 2020). In the United States in 2020, 2,282,015 people were injured were traffic collisions, and in Canada, 101,572 were injured.

A collision avoidance system (CAS) is a system designed to reduce the number of traffic collisions. CASs use advanced and emerging technologies such as sensors, computers, wireless communication technologies, and electronic devices that automatically collect and disseminate real-time vehicle data and other important data (Milanes et al., 2012; Wu et al., 2014). In-vehicle CASs have been developed to reduce the risk of vehicle-to-vehicle collisions (Milanes et al., 2012; Houénou et al., 2014; Mattas et al., 2020; Yang, et al., 2021; Choi et al., 2021; Kaul and Altaf, 2022) and to reduce the risk of vehicle-to-pedestrian collisions (Dong et al., 2017; Pathak and Sivraj 2018; Zadobrischi and Negru 2020). Throughout of this thesis, the risk of a collision is defined as the chance (probability) of a collision given the temporal and spatial interaction between road users (i.e., road user closeness in space and time) (Johnsson et al., 2018).

Due to the high number intersections-related collisions, many systems focus mainly on reducing the risk of collisions at intersections. In the United States in 2019, for example, 976,816 (50%) of the

1,949,831 collisions occurred at intersections, and intersections accounted for 25% of all fatal collisions and 49% of all injury collisions (NHTSA, 2019). In Canada in 2019, 47,593 (44%) of the 105,782 collisions occurred at intersections, and intersections accounted for 26% of all fatal collisions and 46% of all injury collisions (Transport Canada, 2019a).

Collisions on road segments, however, usually result in more severe consequences (Lee et al., 2006). For example, in Canada, 51% all fatal collisions occurred on primary or secondary highways, or on local roads where the speed limit at the collision site exceeded 60 km/h (Transport Canada, 2019b). It is clear that developing CASs to reduce the risk of segment collisions needs special attention.

This study selected three types of collision on segments: animal-to-vehicle collisions, pedestrian-to-vehicle collisions particularly collisions between vehicles and people working on a roadway on foot e.g., police officers, first response workers and people involved in road construction, and pedestrian-to-pedestrian collisions.

1.1. Historical Statistics for Segment Collisions

Section 1.1 briefly presents historical statistics for the three types of segment collision selected and then discusses the CASs currently available.

1.1.1. Animal-To-Vehicle Collisions

Animal-to-vehicle collisions are a challenging problem in North America, particularly on rural highway segments. The collisions threaten human lives and cause economic damage and ecological losses (Fahrig and Rytwinski, 2009). The importance of animal-to-vehicle collisions is viewed not only from human safety perspective but also from wildlife conservatives' perspective.

From the human safety perspective, the United States recorded a total of 1,739,687 collisions involving animals during the four years from 2014 to 2017. These collisions led to approximately 200

human deaths, and $2 billion of direct insurance claims each year (National Insurance Crime Bureau, 2018). Canada reported 474 human deaths in collisions involving large animals between 2000 and 2014 (Traffic Injury Research Foundation, 2015).

From the wildlife conservation perspective, Virginia in the United States reported 56,000 deer deaths on the road in 2012 alone (Druta and Alden, 2015), and in 2007 British Columbia in Canada reported 5,913 wildlife deaths of which approximately 76% were deer (BC MoTI, 2010). These are just the reported statistics as many collisions remain unreported (Benten, 2018). British Columbia estimates that the number of wild animal deaths recorded in 2007 represents only about 25% to 35% of the actual number (BC MoTI, 2010). Collisions involving animals have a negative impact on animal populations (Fahrig and Rytwinski, 2009) and direct and indirect consequences for ecosystems (Davenport and Davenport, 2006).

1.1.2. Pedestrian-To-Vehicle Collisions

People whose work involves working on roadway segments as pedestrians are at risk of being hit by a passing vehicle. If hit, there is little or no protection that could absorb the impact of the collision (Yannis et al., 2020). Pedestrian workers usually are distracted by number of activities. For example, work zone workers are dealing with construction/maintenance tasks, and police officers are typically directing traffic, stopping traffic for enforcement, or investigating the scene of a severe collision.

From 2018 to 2019, the United States recorded 252 pedestrian deaths in work zones (126 deaths per year) (NHSTA, 2020a). From 2015 to 2019, the United States recorded 50 police officer deaths from being struck by a vehicle (FBI, 2020).

1.1.3. Pedestrian-To-Pedestrian Collisions

The COVID-19 outbreak spread swiftly around the world, and by June 2022, had infected nearly 540

million people in about 210 countries, with nearly 6.3 million deaths (WHO, 2022). As COVID-19 is known to be spread by physical contact, coughs, sneezes, or even exhaled breath while talking with an infected person, public health agencies strongly recommended avoiding contact and other maintaining the recommended physical distancing between people (WHO, 2020, ECDC, 2020). Many municipalities around the world quickly implemented various interventions (e.g., sidewalk expansion) on urban streets to ensure adequate space for anyone using the street as a pedestrian.

The main goal of such interventions is to ensure that pedestrians do not contact and/or collide with each other and can maintain physical distancing, i.e., to avoid pedestrian-to-pedestrian collisions. This issue is not limited to the current COVID-19, but also relevant any similar health crises that may occur in the future.

1.2. Existing Collision Avoidance Systems

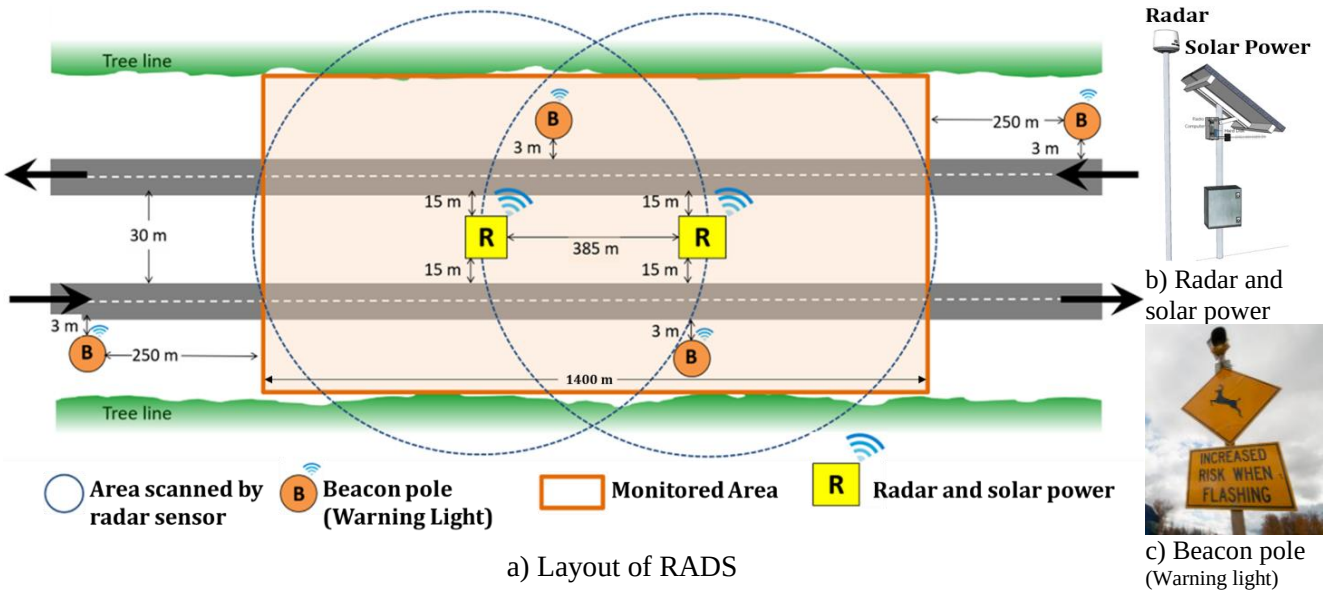
1.2.1. Animal-To-Vehicle Collision Avoidance Systems

Roadside animal detection systems (RADSs) are the most common type of CAS designed to reduce the number of animal-to-vehicle collisions (Huijser et al., 2006). RADSs use a sensor in conjunction with flashing warning lights installed on roadways where the possibility of animals crossing is relatively high. If a RADS detects an animal on or near the roadways, the system activates the flashing warning lights to warn approaching vehicles.

Figure 1-1 is a RADS installed on Highway 416 near Kemptville, Ontario, Canada (Mukherjee et al., 2013). The system is designed to detect the presence of large animals on approximately 1.4 km of highway.

The system has two solar powered radar units (the two Rs in yellow boxes). The warnings are disseminated to drivers via four flashing warning lights (the four Bs in orange circles).

Existing RADSs warn approaching vehicles if the system detects the presence of an animal within a pre-set lateral (i.e., (perpendicular) distance) (e.g., 5 meters) from a roadway (Sharafsaleh et al., 2010; Mukherjee et al., 2013; Grace et al 2017; Druta et al., 2018; Antônio et al., 2019).



a) Layout of RADS

b) Radar and solar power

c) Beacon pole (Warning light)

Figure 1-1: A RADS (Mukherjee et al., 2013)

As the systems use only one input indicator, i.e., the lateral (i.e., perpendicular) distance from the animal to the roadway to activate the warning lights, the RADS simply produces a binary output (on/off warning lights) based on a binary input (presence/absence of an animal). This means that the RADSs simply activate the warning flashers during the time the system detects the presence of an animal within the preset detection area, and then turns off the warning flashers shortly after the animal moves outside of the detection area.

Existing RADSs have major limitations. They do not consider possible differences in factors such as the animal's behavior (e.g., unpredictable jumping, moving slowly, stopping, etc.), the vehicle(s) (e.g., leading/following vehicle(e)), the position of the animal relative to the roadway (on/off the roadway), the

spatial and temporal interactions between the animal and the vehicle (in meters or time), driver behavior (aggressiveness), etc.

1.2.2. Pedestrian-To-Vehicle Collision Avoidance Systems

CASs designed to protect distracted pedestrians from traffic are known as pedestrian protection systems. For example, Wang et al. (2012) developed a pedestrian protection system called "WalkSafe" which aimed to protect a distracted pedestrian by using a smartphone to alert the pedestrian. The system relied on phone cameras to detect vehicles coming towards a pedestrian, and provided a warning four seconds before to possible collision. Tong and Jia (2019) developed augmented reality glasses (i.e., a wearable device) for distracted pedestrians. The device activated a warning when there was a risk of a collision. The screen of the virtual reality glasses displayed the remaining time to the possible collision and showed the expected path of the approaching vehicle.

CASs have also been designed specifically to reduce the risk of collisions with pedestrians who work on roadways. These studies typically used smart cones and/or smart vests to detect the intrusion of a passing vehicle into work zone. For example, Mollenhauer et al. (2019) developed a smart vest that warns work zone workers that they are at risk from a passing vehicle. In the system, smart cones are equipped with sensors that detect a passing vehicle's intrusion into the work zone. Confucian's (2022) "Traffic Accident Warning System" aims to warn police officers working at a collision that a passing vehicle may cause a secondary collision. If a smart cone is hit by a passing vehicle, the cone immediately sends a signal to alert the police officer. Figure 1-2 shows a smart cone developed by smartcone.com. The cone uses a light-emitting diode (LED) to warn the pedestrian.

Existing systems (including smart cones) designed to protect pedestrians working on a road have major limitations. They do not consider differences in pedestrian behavior or circumstances (e.g., distance from pedestrian to roadway), differences in driver behavior (e.g., speed of passing vehicle, aggressiveness

of driver), or spatial and temporal interactions between pedestrian and vehicle, etc.

Figure 1-2: A smart cone developed by smartcone.com

1.2.3. Pedestrian-To-Pedestrian Collision Avoidance Systems

The concept of pedestrian-to-pedestrian collisions brought about by the COVID-19 pandemic has led to the development of specially designed CASs. As these CASs monitor pedestrian density on urban streets, the systems are better described as monitoring systems.

Meivel et al. (2021), for example, developed a monitoring system to detect physical distancing in real-time. The system used a camera in conjunction with a convolutional neural network to detect pedestrians. Figure 1-3 shows an image from the system on the left and a bird's-eye view on the right. Several people are identified. The system automatically generates the bird's-eye view and estimates the distance between the pedestrians. If a pedestrian is at least two meters from another pedestrian, the pedestrian is shown in a yellow rectangle. Pedestrians who are within two meters of another pedestrian

are shown by a red rectangle.

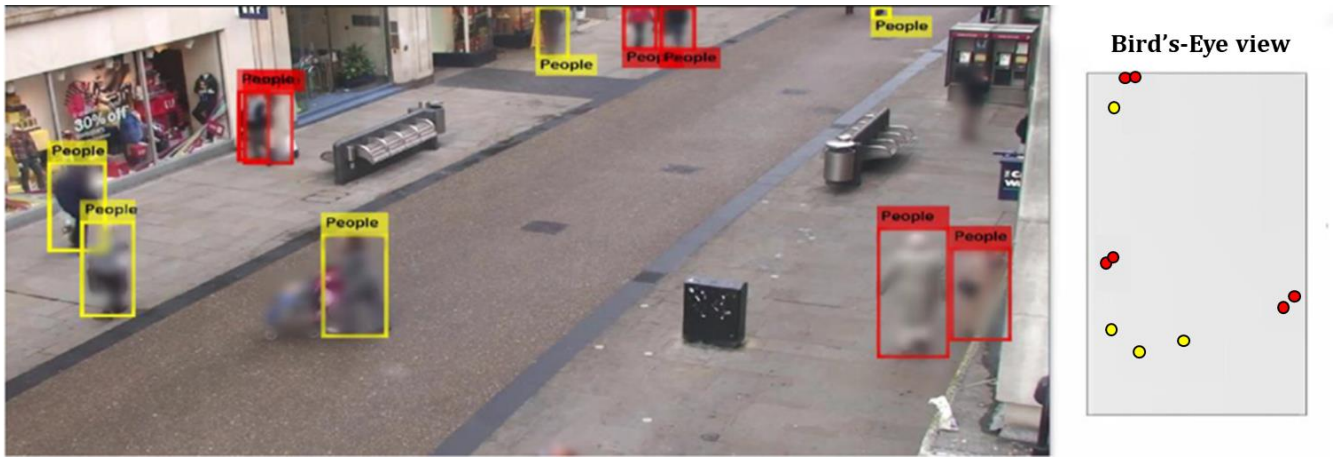


Figure 1-3: Real time monitoring system for physical distancing detection (Meivel et al., 2021)

El Habchi et al. (2022) developed a camera-equipped drone to detect the physical distance between pedestrians. The system uses a deep learning algorithm to detect the real-world position of each pedestrian, and then estimates the distance between the pedestrians to determine whether the distances are considered safe or not.

Loke et al. (2022) developed a prototype device called PADDIE-C19 to monitor the physical distance between pedestrians. The system detects pedestrians and estimates the distance between pedestrians to determine whether the distances are considered safe or not.

Existing physical distancing monitoring systems use only one input indicator (i.e., the 2-meter distance between pedestrians) and a simple risk analysis algorithm to analyze the movement of pedestrians and determine how well physical distancing is being achieved. However, pedestrian physical distancing can be described as a mobility intervention (interventions imposed on mobility of pedestrians by authorities) designed to limit virus infection by keeping people, for example, 2 m apart.

Physical distancing assumes that the risk of transmission declines with the distance between people, but a more sophisticated approach than the current algorithm which is simply based on a single

distance may help health agencies to evaluate and monitor how well physical distancing is being achieved. A sophisticated approach would also consider exposure time and density which play an important role in the possibility of viral transmission. Increased exposure time and density increases the risk of viral transmission (Vuorinen et al., 2020; CDC, 2020; Parshina-Kottas et al., 2020). Therefore, use of only one input indicator (i.e., physical distance between pedestrians) for the risk analysis can be viewed as a major limitation of current systems.

1.3. Research Gaps

As mentioned, existing CASs for each of the three selected types of segment collision (i.e., RADS, pedestrian protection systems, and pedestrian physical distance monitoring systems) use one input indicator and apply a simple risk analysis algorithm to differentiate a safe situation from a potentially risky situation. The simple analysis involves compare the input indicator with a fixed threshold.

Use of one input indicator and apply the simple risk analysis algorithm bring many limitations which we explain them in Section 1.3.1. Research Gaps in Input Indicator Selection and Section 1.3.2. Research Gaps in Existing Risk Analysis Algorithms.

1.3.1. Research Gaps in Input Indicator Selection

The single input indicators used in the risk analysis of existing CASs are known as surrogate safety indicators. RADSs use only the position of the detected animal, pedestrian protection systems use only the remaining time to collision or the intrusion into a work zone by a passing vehicle, and pedestrian physical distancing detection systems use only physical distance.

Surrogate safety indicators are indicators developed by transportation safety engineers to discriminate safe situation from a risky situation (Zheng et al., 2021). Unlike conventional safety studies, such as in-service road safety review projects that rely largely on analyzing historical collision data, CASs

require a real-time estimation of risk (the inverse of safety). As historical collision data are not appropriate, the systems use surrogate safety indicators.

Relying on historical collision data as the main data source for studying safety has many drawbacks: lack of suitable historical collision records, lack of information about the interaction processes that led to the collision, a high degree of under-reporting, and the ethical problem that one has to wait for a collision to occur to be able to investigate possible causes (Laureshyn et al., 2017a). Surrogate safety indicators overcome these drawbacks, and it has been shown that the number of risky situations identified by surrogate safety indicators is statistically correlated to the number of collisions (Amundsen and Larsen, 1977; Migletz et al., 1985; Peesapati, et al., 2013; Tarko, 2018; Charly and Mathew 2019).

Many surrogate safety indicators have been developed. The most commonly applied indicators are reported in Appendix A. Johnsson et al. (2018) and Mahmud et al. (2019) reviewed surrogate safety indicators and divided them into three groups.

The first group of indicators attempts to capture the spatial and temporal interaction between two road users (e.g., a vehicle and a pedestrian) in distance or time. This group includes TTC, post encroachment time (PET) (Cooper 1984), time to crosswalk (Vogel 2003a), time headway (Vogel 2003b), and braking time (Lu et al. 2012).

The second group attempts to capture the behavior of road users; for example, the behavior of drivers as driver behavior is the most important collision contributing factor (Treat 1979; Terzi, 2018). For example, in 2019, 26 % of fatal traffic collisions in the United States involved speeding, and 8% involved a distracted driver (a distracted driver has diverted his attention from the driving task to focus on a different activity or issue) (NHSTA, 2020b; NHSTA, 2022). An aggressive driver increases the chance of a collision by 55% (Song et al., 2021). Driver behavior indicators include, but are not limited to

magnitude of speeding (i.e., driving faster than the posted speed limit) (Terzi, 2018), sudden lane changes (Bergasa, 2014), sudden brake force during deceleration (Chen, 2015a), and variation in speed (Sugimoto et al. 2008).

The third group attempts to capture environmental factors (e.g., weather/road condition) and vehicle characteristics. This group includes quantitative indicators and qualitative indicators. For example, stopping sight distance (SSD) is a quantitative indicator which quantitatively takes into account road condition and vehicle characteristics (e.g., braking technology may differ vehicle by vehicle) (Layton & Dixon, 2012). There are also qualitative indicators (e.g., good, moderate, and bad type of indicators) such as visibility (good, moderate, bad) (Sugimoto et al. 2008), roadway surface condition (dry, wet, snowy, Icy) (Bokare and Maurya, 2017), etc.

Given the large number of indicators that can be developed, a CAS that uses only one indicator to differentiate safe and risky situations clearly has limitations.

The first limitation is that no single indicator can capture simultaneously different aspects of risk associated with spatial and temporal interaction between road users and the behavior road users (e.g., drivers' behavior), environmental factors, and vehicle characteristics (Johnsson et al., 2018). A second limitation is that each indicator can capture only one type of collision (Johnsson et al., 2018). The estimation of the risk of different types of collision requires different indicators and more than one indicator. For example, in the case of rear-end or head-on collision, TTC, time headway, and braking time are useful (Mahmud et al., 2019). In the case of collisions that occur on pedestrian crossings, PET and time to crosswalk are useful (Mahmud et al., 2019). In the case of side-swipe collisions, the lateral distance between two road users is useful. In the case of animal-to-vehicle collisions, a vehicle can be involved in different types of collision, e.g., a head-on collision, an animal crossing the road collision, or a side-swipe collision. No single indicator can capture all these types of collision.

A third limitation is that introducing multiple indicators and conducting a risk analysis separately for each indicator can result in different conclusions (Saisana et al. 2005; Shen et al. 2012; Chen et al. 2015b). For example, Tak et al. (2015), developed a deceleration-based indicator which relied on the SSD of two vehicles. They compared the results obtained using the new indicator with the results obtained using TTC rear-end collisions in which the vehicle braked sharply. In one of their observations, the leading and following vehicles were 50 meters apart, the leading was traveling at 110 km/h and the following vehicle was traveling at 95 km/h. The leading vehicle suddenly braked sharply. In reality, the collision occurred because 50 meters did not give the following vehicle enough distance to avoid the collision. When Tak et al. (2015) used TTC, the situation was evaluated as safe. When they evaluated this situation using the SSD of two vehicles, the situation was evaluated as risky. The study showed that different indicators may produce different results.

The use of a combination of the various risks estimated by multiple input indicators could overcome the above limitations (Johnsson et al., 2018; Arun et al., 2021; Zheng et al., 2021; Jameel, 2019), but the combination of the estimates made by multiple input indicators into a composite risk indicator requires a sophisticated risk analysis algorithm. The strength of a composite risk indicator would be its ability to summarize and condense the complexity of the factors contributing to a collision to a meaningful and manageable amount of information (Godfrey and Todd, 2001).

Section 1.3.1 considers the gaps in the research into risk analysis algorithms designed to create a composite risk indicator.

1.3.1. Research Gaps in Existing Risk Analysis Algorithms

Existing CASs for the three types of segment collision selected apply a simple risk analysis algorithm (i.e., a fixed threshold). For example, existing RADSs compare the input indicator (i.e., the position of detected animals) with a fixed threshold (e.g., 5 meters) to discriminate safe and risky situations and

produce a warning if required. Although this approach is very straightforward, it has several drawbacks. The main drawback is that the approach cannot combine multiple input indicators. When multiple input indicators are relevant to the situation, a RADS would need to conduct a risk analysis by each indicator separately. We already explained that introducing multiple indicators and conducting risk analysis by each indicator separately can result in different conclusions. Another drawback is that this approach cannot capture uncertainty. The selection of the threshold is often arbitrary and results in uncertainty. For example, many studies have used TTC to distinguish safe and risky situations, but the studies selected different thresholds. For example, Vogel (2003a) used a TTC of 2 secs, El-Basyouny and Sayed (2013) used a TTC of 1.5 secs, and Sacchi et al. (2013) used a TTC of 3 secs.

To find a better risk analysis algorithm, we did not limit our search to the three selected types of segment collision. We also searched studies that used risk analysis algorithms to combine multiple input indicators for the analysis of other types of collision (e.g., vehicle-to-vehicle collisions). We found that many studies applied statistical algorithms to combine multiple input indicators to create a composite risk indicator. For example, several studies used a weighted arithmetic average (summations of multiple input indicators) or weighted geometric average (multiplication of multiple input indicators).

The most widely used approach is the traffic conflict technique (TCT) proposed by Hydén (1987) for vehicle-to-vehicle collisions. TCT is the summation of two input indicators (i.e., weighted arithmetic average): time to accident, and conflicting speed. A collision course must be assumed when using TCT. This means that at a certain moment, road users are on a path to collide, and would collide if one or both of them did not take evasive action.

Lu et al. (2012) developed a composite risk indicator called a quantitative index for vehicle-to-vehicle collisions at an intersection. The index is calculated by multiplying incomplete braking time by

TTC. The study used fixed thresholds to classify the index into three risk categories (no conflict, non-serious conflict, and serious conflict).

Wang and Stamatiadis (2014) build a composite risk indicator called an aggregate crash propensity metric for vehicle-to-vehicle collisions at an intersection. They used a weighted average method to combine three surrogate safety indicators (required braking rate, maximum available braking rate, and TTC).

The literature included many statistical algorithms that used a statistical function (logistic/normal distribution) to link independent variables (e.g., TTC) to a dependent variable (collision risk). These algorithms include, but are not limited to, binary logit models (Kadali and Vedagiri, 2016; Uzondu et al., 2018; Xing et al., 2019), ordered logit models (Stipancic et al., 2016; Tageldin, and Sayed, 2019; Li et al., 2019), a multinomial logit model (Dimitriou et al., 2018), and a nonlinear regression model (Gu et al., 2019). Although statistical algorithms are easy to perform and simple to understand, they usually need a large amount of data (include collision/near miss collision data) to calibrate the algorithm, and they cannot capture uncertainty.

In our study, we face limited data availability as no dataset exist for the selected CAS which could provide position data (x- and y-coordinates) and speed of road users in real-time. Furthermore, our limited dataset does not include collision/near miss collision data. This was a major challenge as we had no way to show the outcome of certain combinations of input indicators could be considered to indicate the development of a high or low risk situation. As a result, we could not simply use statistical algorithms (e.g., a regression model) in which the independent variables (e.g., input indicators) and dependent variable (composite risk indicator) are known. In our study, for limited available data, the independent variables were known, but the dependent variable was unknown.

We decided to select artificial intelligence (AI) algorithms to create a composite risk indicator from multiple input indicators. Intelligence is the ability to learn and understand, to solve problems and to make decisions (Negnevitsky et al., 2005). AI algorithms are algorithms that make machines do things that would require intelligence if done by humans (Boden, 1977). The question of whether machines/computers can think or can be intelligent has been an important question since the early days of artificial intelligence in the late 1940s. The answer to the question is not as easy as "yes" or "no." Some people are smarter than others. Some of our decisions are intelligent while some are silly. Some people are good at planning and construction, but bad in history and physics. Humans have the ability to learn and understand, to solve problems and to make decisions, but our abilities are not equal. Therefore, we should expect that if machines/computers can think, some of them might be smarter than others in some ways.

Human intelligence is internal and too complex to be represented as an algorithm. AI tries to simulate human intelligence. There are two popular types of algorithms in AI (Negnevitsky et al., 2005): (1) Rule-based algorithms (2) Learning-based algorithms.

(1) Rule-based algorithms: People are capable of expressing their knowledge in the form of rules for problem solving. For example, Negnevitsky et al. (2005) provided an example. Imagine an alien asks a person how to cross an intersection. The person is able to teach the alien how to cross the intersection by simple IF-THEN rules. The person says:

Rule 1 "IF the traffic light is green THEN the action is go"

Rule 2 "IF the traffic light is red THEN the action is stop"

These statements represented in the IF-THEN form are used to create a rule-based algorithm.

Fuzzy rule-based algorithms are a popular rule-based algorithm used in AI risk analysis (Negnevitsky et al., 2005). Fuzzy rule-based algorithms do not require a large volume of data to calibrate the algorithm and can handle uncertainties (Dimitriou et al., 2008; Liu and Cocea, 2017).

We compared a fuzzy rule-based algorithm with a statistical algorithm to deepen our understanding of these algorithms. In statistical algorithms, input indicators have numerical values, the relationship between the input indicators and the composite risk indicator is defined in terms of a mathematical function, and the composite risk indicator is a numerical value. In a fuzzy rule-based algorithm, the input indicators are converted into fuzzy sets (e.g., safer and riskier), the relationships are defined by if-then rules (e.g., if input indicator A is riskier, then situation becomes risky), and the composite risk indicator is expressed in terms of a fuzzy set (e.g., low, medium, high) (Liu and Cocea, 2017).

Very few studies of CASs for vehicle-to-vehicle collisions have used fuzzy rule-based algorithms to combine multiple input indicator to create a composite risk indicator. Milanés et al. (2012) developed a fuzzy rule-based algorithm to combine two indicators of TTC and time gap to generate a composite risk indicator with three labels (low, medium, high). The study focused on vehicle-to-vehicle rear-end collisions. Nadimi et al. (2016) developed a new composite risk indicator that combined TTC and PET using a fuzzy rule-based algorithm. The study focused on vehicle-vehicle rear-end collisions.

(2) Learning-based algorithms: Learning-based algorithms, known as machine learning algorithms, are also popular algorithms used in AI risk analysis. Whereas rule-based algorithms such as fuzzy rule-based algorithms are designed by an expert, machine learning algorithms learn from experience and historical data. The algorithms automatically analyze and interpret underlying patterns and structures in the data and improve their performance over time. To do this, machine learning algorithms need a large amount of data to calibrate the algorithm to perform well. The algorithms are difficult to understand and difficult to perform, but they are popular because they require less training time to calibrate the algorithm,

they gradually learn over time, and there is the possibility of combining them with other algorithms (e.g., fuzzy rule-based algorithm) to develop hybrid algorithms.

Also, very few studies of CASs for vehicle-to-vehicle collisions have used machine learning algorithms to combine multiple input indicators into a composite risk indicator. For example, Wang et al. (2015) proposed a K-Means method to classify vehicle-vehicle interactions into three risk labels (low, medium, high) based on three input indicators (maximum deceleration, average deceleration, and percentage reduction in vehicle kinetic energy). The authors then used a classification and regression tree to explore the relationships between risk labels, driver/vehicle characteristics, and road environments. Xiong et al. (2019) used a spectral clustering and K-Means method to classify vehicle-to-vehicle interactions into three risk labels (low, medium, high) based on three input indicators (average deceleration, maximum deceleration, and jerk). They then used classified labels to calibrate a fuzzy rule-based algorithm.

However, the above studies were for vehicle-to-vehicle collisions. No study has been conducted into developing a risk analysis using AI algorithms to create a composite risk indicator for a CAS for animal-to-vehicle collisions, pedestrian roadway worker-to-vehicle collisions or pedestrian-to-pedestrian collisions. This thesis can be viewed as the first attempt to apply different AI algorithms to create a composite risk for the three types of collision.

Different statistical algorithms including the binary logit model (Kumar et al., 2019; Xing et al., 2019; Xing et al., 2020) and the ordered logit/probit model (Stipancic et al., 2016; Jiang et al., 2016) have been proposed to estimate the risk associated with collisions. In general, statistical algorithms can be written as $Y = F(X, \beta, \varepsilon)$ where Y is the risk (conflict), X is used for input indicators, β is used for parameters (such as weights for each input indicator), and ε is an error term. Two major conditions need to be met when using any statistical algorithm. The first condition is that each value of Y and X should

be known before conducting any risk analysis. The second condition is a relatively large size dataset that includes data for collisions/near miss collisions. If these two conditions can be met, a statistical algorithm can be used to estimate β through a calibration process.

AI algorithms (fuzzy rule-based algorithm or machine learning) do not require these two conditions. AI algorithms can be used to estimate the risk even when the value of Y is unknown and/or there is no large dataset.

The choice of AI algorithms depends largely on data availability and uncertainty. As we faced limited data availability and the existence of uncertainty, we selected a fuzzy rule-based algorithm to develop the risk analysis algorithm. By collecting more data to include some near miss collisions, we also developed a hybrid algorithm from the fuzzy rule-based algorithm and a machine learning algorithm (a semi-supervised machine learning algorithm) to combine the strengths of both algorithms.

Validating the results of our AI algorithms was a challenge. One approach to validating the results of the proposed risk analysis algorithm was to evaluate the algorithm's accuracy when identifying collisions (Olszewski et al., 2016; Xiong et al., 2019). Another approach was to check that the proposed risk analysis algorithm produced an estimate of increased risk for the time before an actual collision, but did not produce such an estimate before a non-collision (Cunto et al., 2009; Zhao and Lee., 2018). Both above validation approaches need historical collision data or simulated data.

In our study, we used two different approaches to validating our algorithms. Laureshyn et. al. (2017b) and Johnsson et. al. (2018) proposed an approach called "relative product validation" which measures whether the risk analysis algorithm can indicate the expected direction of changes in safety (i.e., improvement or deterioration). In this approach, a 3-D dimensional risk surface is developed to show that the direction of changes in each indicator is correctly reflected in the direction of change in the composite risk indicator. This approach can be used when historical collision data are not available. Our second

approach involved using our proposed risk analysis algorithms to identify near miss collisions and evaluating the accuracy of the identification.

Section 1.4 discusses our observation from an area outside of transportation (air defense systems) which helped us to develop our AI algorithms.

1.4. Observations from Air Defense Systems

The use of surrogate safety indicators is much discussed and developing proper risk analysis using AI algorithms to combine multiple indicators and create a composite risk indicator is recognized as a challenging issue. Recent literature has included comprehensive reviews of surrogate safety indicators and has encouraged further research into suitable methods for combining indicators (Johnsson et al., 2018; Arun et al., 2021; Zheng et al., 2021). Johnsson et al. reviewed 239 articles, reports and dissertations on risk analysis using surrogate safety indicators. They noted that "*Further research focusing on* ***combining the many different indicators*** *could help to strengthen the theoretical base for surrogate safety indicators.*" Arun et al. reviewed 549 studies on risk analysis using surrogate safety indicators and gave special attention to methodological issues. They noted that *"****Mapping several surrogate indicators into a risk indicator*** *is a promising approach that has received relatively little attention."* Zheng et al. (2021) reviewed 168 studies on modeling traffic conflicts using surrogate safety indicators. They noted that "*Road user behavior plays a crucial role in collision occurrence. Given the enormous levels of variation between road users and their reactions to events preceding possible collisions,* ***incorporating human behavior in (risk analysis) models*** *remains a challenge."*

To strengthen our risk analysis using AI algorithms, we searched areas outside of transportation. We found that there are many similarities in the risk analysis of military objects in an air defense system and the risk analysis of traffic objects (i.e., road users include animal, pedestrian, and vehicle) in CASs.

Figure 1-4 shows the similarity between an air defense system and the three types of road user of interest to our study. Each system seeks to protect one or more assets from one or more potential risky objects in real-time. The defense system detects the characteristics of an approaching risky object (e.g., a hostile military aircraft) and uses an algorithm to estimate a composite risk indicator. If a certain risk threshold is met, the system indicates the level of alarm or response required (Johansson and Falkman, 2008).

As shown in Figure 1-4, a typical air defense system uses a sensor (e.g., microwave radar) to detect an approaching risky object. Our proposed CASs take a similar approach.

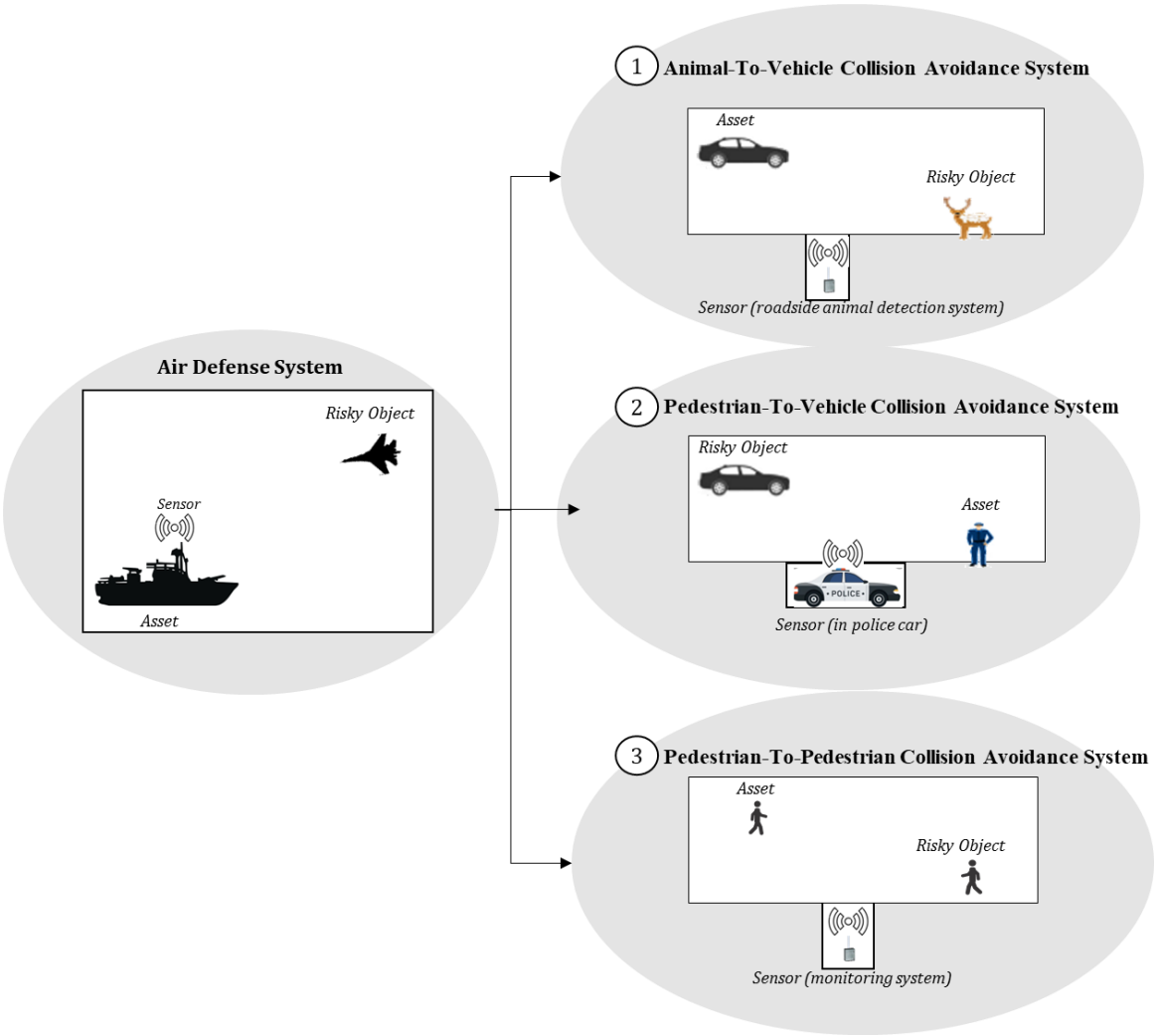


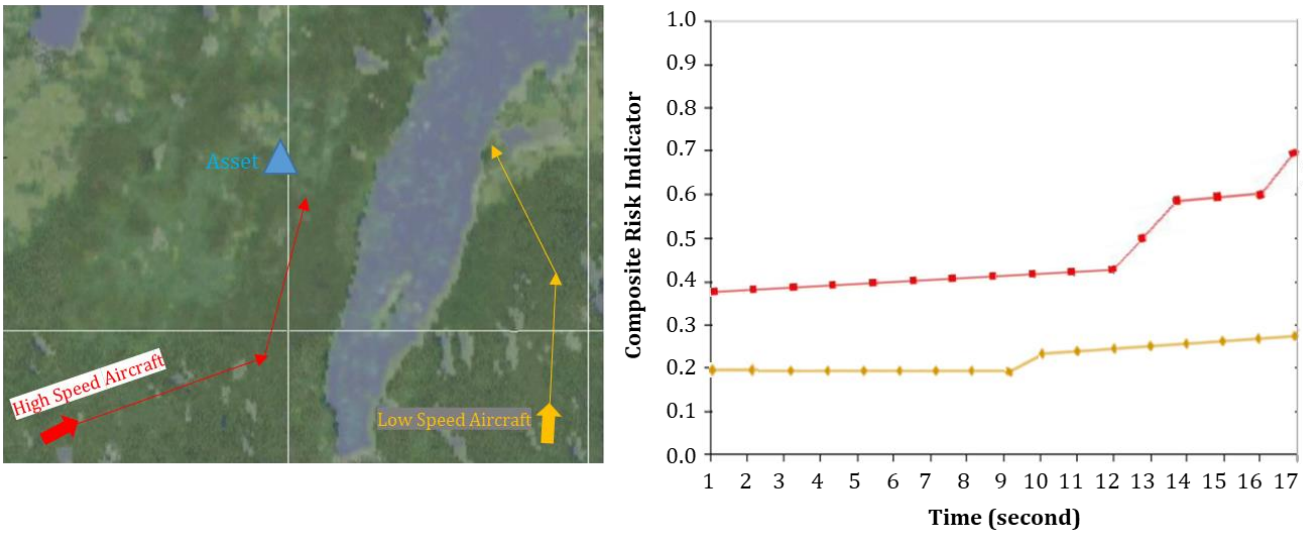
Figure 1-4: Similarities between air defense system and CAS in this study

The animal-to-vehicle CAS detects the characteristics of the animal (risky object) and passing vehicle (asset) and uses an algorithm to estimate the composite risk indicator and determine whether to activate an alarm for passing vehicles.

The vehicle-to-pedestrian CAS detects the characteristics of the passing vehicle (risky object) and pedestrian who work in the roadway (police officer) (asset) and uses an algorithm to estimate the composite risk indicator and determine whether to activate an alarm to alert the pedestrian. The pedestrian-to-pedestrian collision avoidance (monitoring) system detects the characteristics of the pedestrian (risky object) and pedestrian (asset) and uses an algorithm to estimate the composite risk indicator and determine how well physical distancing is being achieved.

The air defense systems combine multiple input indicators to create a composite risk indicator in real-time. Input indicators includes military object behavior (e.g., military aircraft speed) and the spatial and temporal interaction between the military object and the asset (e.g., direct distance and time to hit). The system then applies a risk analysis algorithm using AI (e.g., a fuzzy rule-based algorithm) and produces a composite risk indicator ranging from zero (the lowest risk) to one (the highest risk) (Liang, 2007). Figure 1-5a shows a typical example of the positions and trajectories of two potential risky objects (a high-speed aircraft and a low-speed aircraft) and an asset in an air defense system. Figure 1-5b shows the risk values (i.e., value of the composite risk indicator) generated in real-time. The high-speed aircraft is shown as a red arrow with a red trajectory and a red risk value. The low-speed aircraft is shown as a yellow arrow with a yellow trajectory and a yellow risk value. In Figure 1-5, the risk value gradually increases as the high-speed aircraft approaches the asset, but at second 12, the trajectory of the aircraft suddenly changes, and the risk value suddenly increases from 0.41 at second 12 to 0.58 at second 14 and then to 0.70 at second 17. The risk value for the low-speed aircraft is steady until second 9 when the aircraft changes direction and flies closer to the asset, but as the aircraft is not close to the asset, the risk

value increases only slightly. It is clear form the Figure that the behavior of a military object and the spatial and temporal interaction between the military object and the asset are reflected in the composite risk indicator second by second. By using reviewed studies in the literature of surrogate safety indicators and air defense systems, we try to develop risk analysis using AI algorithms in the three selected CAS in our study. We try to clearly illustrate how our proposed risk analysis algorithms could capture road user behavior and spatial and temporal interaction similarly to air defense system.



a. Positions and trajectories of aircraft and asset

b. Real-time composite risk indicator

Figure 1-5: An example of real-time composite risk indicator in air defense system (Johansson and Falkman, 2008)

1.5. Research Goal and Objectives

The goal of this thesis was to explore the development of composite risk indicators for different CASs by conducting various risk analyses using different AI algorithms. We selected three CASs to demonstrate the whole process of developing risk analysis algorithms from selecting safety indicators to combining multiple indicators into a composite indicator to estimating risk by taking into account road user behavior and the spatial and temporal interactions between road users.

We had four objectives (See Figure 1-6). Objectives 2 and 3 were part of the same CAS study:

1. Develop a fuzzy rule-based algorithm for a next generation roadside animal detection system (NG RADS);
2. Develop a fuzzy rule-based algorithm for a smart protection system to prevent collisions with police officers on duty on the roadway;
3. Develop a semi-supervised machine learning algorithm for a smart protection system to prevent collisions with police officers on duty on the roadway; and
4. Developing a risk analysis approach to evaluate physical distancing on urban sidewalks.

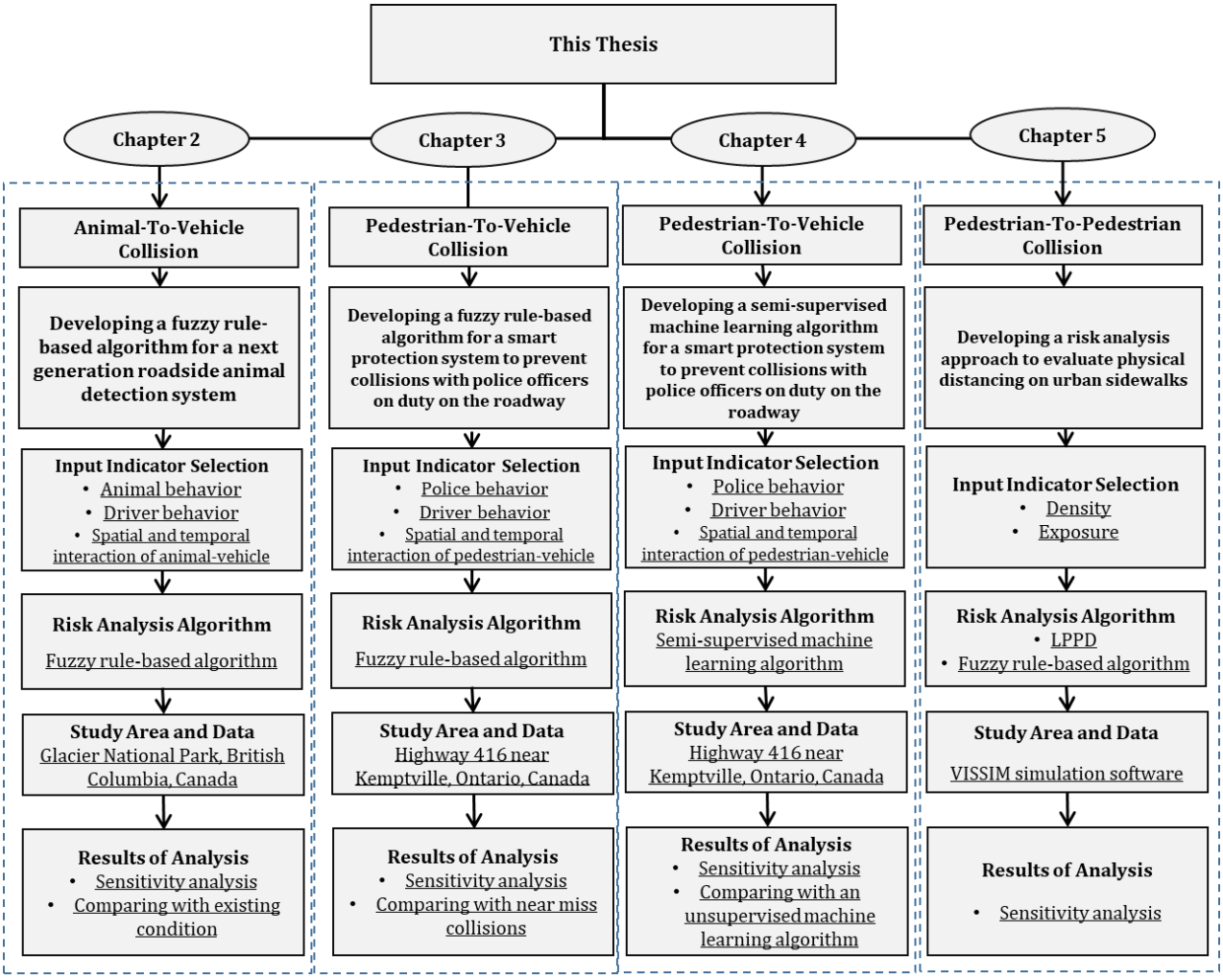


Figure 1-6: Flow chart showing the four objectives of the thesis

Objective 1: Develop a fuzzy rule-based algorithm for a next generation roadside animal detection system (NG RADS). This objective focused on avoiding animal-to-vehicle collisions. The study developed a framework for a NG RADS. The framework used a risk analysis algorithm (fuzzy rule-based algorithm) to create a composite risk indicator. The algorithm combined multiple indicators that can estimate the varying levels of risk posed by animals on the roadway by assessing real-time data on animal behavior, driver behavior and the spatial and temporal interactions between animal and driver.

Objective 2: Develop a fuzzy rule-based algorithm for a smart protection system to prevent collisions with police officers on duty on the roadway. This objective focused on avoiding vehicle-to-pedestrian collisions. The study proposed a smart protection system for pedestrians who work on the roadway (police officer in our case). The system used a risk analysis algorithm (fuzzy rule-based algorithm) to create a composite risk indicator. The algorithm combined multiple indicators to provide police officers on duty on the road with an accurate warning of the risk of being hit by an approaching vehicle by assessing real-time data on officer behavior, driver behavior and the spatial and temporal interactions between the police officers and approaching vehicle.

Objective 3: Develop a semi-supervised machine learning algorithm for a smart protection system to prevent collisions with police officers on duty on the roadway. Like Objective 2, this objective focused on avoiding vehicle-to-pedestrian collisions. The study designed a semi-supervised machine learning algorithm that included multiple processes for learning in a semi-supervised fashion. The study used a large amount of unlabeled data and a small amount of data for near miss collisions, and used driver braking behavior to identify near miss collisions. The study also applied an unsupervised machine learning algorithm to compare with the proposed semi-supervised machine learning algorithm.

Objective 4: Develop a risk analysis approach to evaluate physical distancing on urban sidewalks. This objective focused on avoiding pedestrian-to-pedestrian collisions. The study proposed a

risk analysis to evaluate and monitor sidewalks in terms of the level of physical distancing using a more sophisticated approach than the current binary distance threshold approach based only on a simple physical distance cut off. The study developed two input indicators to capture pedestrian density and pedestrian exposure time. These inputs were used to create a level of pedestrian physical distancing (LPPD) ranging from A to F. The LPPD was designed to help decision makers to select and implement appropriate mobility interventions. The study went on to develop a mathematical process (fuzzy rule-based algorithm) to estimate the relative risk of viral transmission between pedestrians under different pedestrian walking conditions.

As shown in Figure 1-6, we designed a consistent four-step process to achieve the study's goal of exploring the development of composite risk indicators for different CASs. The four-steps were:

(1) input indicator selection. This step involved screening and selecting or developing appropriate indicators representative of the different factors affecting the possibility of a collision occurring. The factors included referred to road user behavior and spatial and temporal interactions between road users;

(2) risk analysis algorithm. This step involved developing appropriate algorithms for combining multiple indicators into a composite risk indicator;

(3) study area and data. This step involved determining the location of study, the type of data (e.g., real-world data/simulation data), and the data specifications; and

(4) results of analysis: This step involved attempts to conduct sensitivity analysis and compare the proposed CAS performance with existing/past systems or studies. Also, we applied validation approaches (mentioned in Section 1.3.1) to validate the results of our studies.

1.6. Thesis Outline

The thesis has six chapters organized as follows:

Chapter 1 introduces and highlights the limitations of risk analysis in CASs for the segment collisions. The chapter includes a brief literature review of existing CASs for animal-to-vehicle collisions, for pedestrian-to-vehicle collisions and for, pedestrian-to-pedestrian collisions, the limitations of using of only one input indicator in the CAS risk analysis, and the research gaps in CAS studies. The chapter also discusses risk algorithms used in existing CAS and considers the similarities between air defense system and the CASs selected for this study. The chapter then presents and discusses the research goal and objectives of this thesis.

Chapter 2 is concerned with animal-to-vehicle collisions. It provides a detailed explanation of the limitations of risk analysis algorithms in existing RADS. It then presents a framework for a proposed next generation roadside animal detection system (NG RADS). The framework uses a sophisticated risk analysis algorithm to create a composite risk indicator. The development of the risk analysis algorithm is discussed in terms of selecting/developing four input indicators to capture animal behavior, driver behavior, and the spatial and temporal interaction between animal and vehicle. The chapter goes on the explain the processes involved in developing the NG RADS's risk analysis algorithm (a fuzzy rule-based algorithm). The chapter discusses how the proposed NG RADS was tested using real-world data from a RADS that covers a 1.4 km section of a two-lane rural highway (Trans-Canada HWY 1) in the Glacier National Park, British Columbia, Canada. The real-time data obtained included the positions (i.e., x- and y-coordinates) of animals and passing vehicles and the second by second speeds of passing vehicles. The chapter continues with an analysis of the results and findings and a discussion of the potential benefits of the NG RADS over existing RADS.

Chapter 3 is concerned with pedestrian-to-vehicle collisions. It provides a literature review of the limitations of risk analysis algorithms in existing pedestrian protection systems designed to prevent pedestrian-to-vehicle (the pedestrian being a police officer in our case) collisions. The chapter proposes a smart protection system that uses a risk analysis algorithm to create a composite risk indicator, and discusses the four input indicators (selected to capture pedestrian behavior, driver behavior and spatial and temporal interaction between pedestrian and vehicle) and the fuzzy rule-based algorithm used to integrate the input indicators into a single composite indicator. The chapter discusses the real-world data obtained from a radar installation on a section of Highway 416 near Kemptville, Ontario, Canada. The real-time (second by second) data obtained include the positions (x- and y-coordinates) of pedestrians working on the roadway and the position and speed of passing vehicles. The chapter discusses the estimation of the composite risk indicator and compares the results of the study with other studies to demonstrate how well the model estimates the risk associated with a police officer's roadway activities.

Chapter 4 is also concerned with pedestrian-to-vehicle collisions. It provides a literature review of the limitations of risk analysis algorithms in existing pedestrian protection systems designed to prevent pedestrian-to-vehicle (the pedestrian being a police officer in our case) collisions. The chapter discusses the large amount of data obtained from a radar installation on a section of Highway 416 near Kemptville, Ontario, Canada. The real-time (second by second) data obtained include the positions (x- and y-coordinates) of pedestrians working on the roadway and the position and speed of passing vehicles. The proposed protection system relies on a risk analysis algorithm. The chapter explains how, to develop the risk analysis algorithm, the analysis uses driver braking behavior to identify near miss collision. The four input indicators (selected to capture pedestrian behavior, driver behavior and spatial and temporal interaction between pedestrian and vehicle) are discussed. The chapter continues by discussing the development of a semi-supervised machine learning algorithm (that uses near miss collisions) and an

unsupervised machine learning algorithm. The chapter concludes by presenting a sensitivity analysis and by comparing the results of the semi-supervised and unsupervised machine learning algorithms.

Chapter 5 is concerned with pedestrian-to- pedestrian collisions. It presents a literature review of the limitations of conducting a risk analysis based on a single indicator (pedestrian physical distancing) for sidewalk CASs. The chapter then discusses two input indicators designed to capture density and exposure in order to develop a quantitative evaluation of how well physical distancing is being achieved on sidewalks. The chapter explains how the evaluation of different levels of physical distancing can help decision makers to select and implement appropriate mobility interventions. The chapter also discusses the development of a fuzzy rule-based algorithm designed to estimate the relative risk of viral transmission between pedestrians under different pedestrian walking conditions. This chapter presents different scenarios on a virtual sidewalk in VISSIM simulation software. The chapter concludes by presenting a sensitivity analysis that is used to asses the results of study.

Chapter 6 summarizes the conclusions for each of the four objectives (i.e., for Chapters 2 to 5). The chapter also discusses the limitations of each study that contributed to the thesis, and makes recommendations for future research.

1.7. Chapter 1 References

Amundsen, F., and Larsen, O., (1977). Traffic conflicts technique-status in Norway. Proceedings of 1st Workshop on Traffic Conflicts, Olso, Norway.

Antônio, W. H., Da Silva, M., Miani, R. S., and Souza, J. R. (2019). A proposal of an animal detection system using machine learning. Applied Artificial Intelligence, 33(13), 1093-1106.

Arun, A., Haque, M. M., Bhaskar, A., Washington, S., and Sayed, T. (2021). A systematic mapping review of surrogate safety assessment using traffic conflict techniques. Accident Analysis & Prevention, 153, 106016.

British Columbia Ministry of Transportation and Infrastructure (BC MoTI), (2010). Wildlife accident monitoring and mitigation in British Columbia, WARS 1988-2007 special annual report. Available at: https://www2.gov.bc.ca/gov/content/transportation/transportation-infrastructure/engineering-standards-guidelines/environmental-management/wildlife-management/wildlife-accident-reporting-system/wars-1988-2007 [Accessed June, 2022]

Bergasa, L. M., Almería, D., Almazán, J., Yebes, J. J., and Arroyo, R. (2014). Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors. In 2014 IEEE Intelligent Vehicles Symposium Proceedings (pp. 240-245). IEEE.

Benten, A. (2018). Evaluation of the efficacy of wildlife warning reflectors to mitigate wildlife-vehicle collisions on roads, Doctoral dissertation, Georg-August-University Gottingen.

Boden, M.A. (1977). Artificial intelligence and natural man. Basic Books, New York

Bokare, P. S., and Maurya, A. K. (2017). Acceleration-deceleration behavior of various vehicle types. World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2017

Centers for Disease Control and Prevention (CDC) (2020). Frequently asked questions and answers. Available at: https://www.cdc.gov/coronavirus/2019-ncov/faq.html [Accessed Oct, 2020].

Centers for Disease Control and Prevention (CDC) (2021), National center for injury prevention and control (NCIPC). Web-based injury statistics query and reporting system (WISQARS). Available at: http://www.cdc.gov/injury/wisqars [Accessed June, 2022]

Charly, A., and Mathew, T. V. (2019). Estimation of traffic conflicts using precise lateral position and width of vehicles for safety assessment. Accident Analysis & Prevention, 132, 105264.

Chen, F., Wang, J., and Deng, Y. (2015a). Road safety risk evaluation by means of improved entropy TOPSIS–RSR. Safety Science, 79, 39-54.

Chen, Z., Yu, J., Zhu, Y., Chen, Y., and Li, M. (2015b). Abnormal driving behaviors detection and identification using smartphone sensors. 12th Annual IEEE International Conference on Sensing, Communication, and Networking, 2015 (pp. 524-532).

Choi, J. G., Kong, C. W., Kim, G., and Lim, S. (2021). Car crash detection using ensemble deep learning and multimodal data from dashboard cameras. Expert Systems with Applications, 183, 115400.

Cooper, P. J. (1984). Experience with traffic conflicts in Canada with emphasis on "post encroachment time" techniques. In International Calibration Study of Traffic Conflict Techniques (pp. 75-96). Springer, Berlin, Heidelberg.

Cunto, F., Duong, D., and Saccomanno, F. F. (2009). Comparison of simulated freeway safety performance with observed crashes. Transportation Research Record, 2103(1), 88-97.

Davenport, J., (2006). The ecology of transportation: managing mobility for the environment (Vol. 385). Dordrecht, the Netherlands: Springer.

Dimitriou, L., Tsekeris, T., and Stathopoulos, A. (2008). Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. Transportation Research Part C: Emerging Technologies, 16(5), 554-573.

Dimitriou, L., Stylianou, K., and Abdel-Aty, M. A. (2018). Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions. Accident Analysis & Prevention, 118, 221-235.

Dong, Y., Li, Y., Liu, W., and Wu, J. (2017). Unconscious behavior detection for pedestrian safety based on gesture features. 18th International Conference on Parallel and Distributed Computing, Applications and Technologies, 2017, (pp. 39-43). IEEE.

Druta, C., and Alden, A. S. (2015). Evaluation of a buried cable roadside animal detection system. Virginia Center for Transportation Innovation and Research. Report No. FHWA/VCTIR 15-R25

Druta, C., Alden, A. S., and Donaldson, B. M. (2018). Evaluation of a buried sensing cable for roadside animal detection. Transportation Research Board 97th Annual Meeting, Washington DC, United States

El-Basyouny, K., and Sayed, T. (2013). Safety performance functions using traffic conflicts. Safety Science, 51(1), 160-164

El Habchi, A., Baibai, K., Moumen, Y., Zerouk, I., Khiati, W., Rahmoune, N., and Bouchentouf, T. (2022). Social distance monitoring using YoloV4 on aerial drone images. In E3S Web of Conferences (Vol. 351, p. 01035). EDP Sciences.

European Centre for Disease Prevention and Control (ECDC), (2020). Considerations Relating to Social Distancing Measures in Response to COVID-19—Second Update. March 23. Available at: https://www.ecdc.europa.eu/en/publications-data/considerations-relating-social-distancing-measures-response-covid-19-second [Accessed June, 2022].

Fahrig, L., and Rytwinski, T. (2009). Effects of roads on animal abundance: an empirical review and synthesis. Ecology and society, 14(1).

Federal Bureau of Investigation (FBI) (2019). Law enforcement officers accidentally killed- type of accident and activity of victim officer at time of incident, 2015–2019. Available at:

https://ucr.fbi.gov/leoka/2019/topic-pages/tables/table-65.xls [Accessed June, 2022].

Godfrey, L., and Todd, C. (2001). Defining thresholds for freshwater sustainability indicators within the context of South African water resource management. Second Waternet Symposium, Cape Town, South Africa, 2001.

Grace, M. K., Smith, D. J., and Noss, R. F. (2017). Reducing the threat of wildlife-vehicle collisions during peak tourism periods using a roadside animal detection system. Accident Analysis & Prevention, 109, 55-61.

Gu, X., Abdel-Aty, M., Xiang, Q., Cai, Q., and Yuan, J. (2019). Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. Accident Analysis & Prevention, 123, 159-169.

Houénou, A., Bonnifait, P., and Cherfaoui, V. (2014). Risk assessment for collision avoidance systems. 17th International IEEE Conference on Intelligent Transportation Systems (pp. 386-391). IEEE.

Huijser, M. P., McGowen, P. T., and Camel, W. (2006). Animal vehicle crash mitigation using advanced technology phase I: review, design, and implementation. Western Transportation Institute. Report No. FHWA-OR-TPF-07-01.

Hydén, C. (1987). The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique. Bulletin Lund Institute of Technology, Department.

Jameel, A. K. (2019). Development of a holistic index for safer roads, Doctoral dissertation, University of Birmingham.

Jiang, X., Zhang, G., Bai, W., & Fan, W. (2016). Safety evaluation of signalized intersections with left-turn waiting area in China. Accident Analysis & Prevention, 95, 461-469.

Johansson, F., and Falkman, G. (2008). A Bayesian network approach to threat evaluation with application to an air defense scenario. 11th International Conference on Information Fusion (pp. 1-7). IEEE.

Johnsson, C., Laureshyn, A., and De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. Transport Reviews, 38(6), 765-785.

Kadali, B. R., and Vedagiri, P. (2016). Proactive pedestrian safety evaluation at unprotected mid-block crosswalk locations under mixed traffic conditions. Safety Science, 89, 94-105.

Kaul, A., and Altaf, I. (2022). Vanet-TSMA: A traffic safety management approach for smart road transportation in vehicular ad hoc networks. International Journal of Communication Systems, 35(9), e5132.

Kumar, A., Paul, M., & Ghosh, I. (2019). Analysis of pedestrian conflict with right-turning vehicles at signalized intersections in India. Journal of Transportation Engineering, Part A: Systems, 145(6), 04019018.

Laureshyn, A., De Ceunynck, T., Karlsson, C., Svensson, A., and Daniels, S. (2017a). In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator. Accident Analysis & Prevention, 98, 46-56.

Laureshyn, A., Johnsson, C., Madsen, T., Varhelyi, A., de Goede, M., Svensson, A and van Haperen, W. (2017b). Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users. The Road Safety & Simulation International Conference, 17-19 October 2017, The Hague, NL.

Layton, R., and Dixon, K. (2012). Stopping sight distance. Kiewit Center for Infrastructure and Transportation, Oregon Department of Transportation.

Lee, C., Hellinga, B., and Saccomanno, F. (2006). Evaluation of variable speed limits to improve traffic

safety. Transportation research part C: emerging technologies, 14(3), 213-228.

Li, M., Chen, L., Chang, F., and Huang, H. (2019). Safety evaluation of discontinuous lane design of intersection approach in China. Journal of Transportation Safety & Security, 11(4), 398-413.

Liang, Y. (2007). An approximate reasoning model for situation and threat assessment. 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) (Vol. 4, pp. 246-250). IEEE.

Lin, C. H., Chen, Y. T., Chen, J. J., Shih, W. C., and Chen, W. T. (2016). psafety: A collision prevention system for pedestrians using smartphone. 84th Vehicular Technology Conference (VTC-Fall) 2016. (pp. 1-5). IEEE.

Liu, H., and Cocea, M. (2017). Fuzzy rule-based systems for interpretable sentiment analysis. 9th International Conference on Advanced Computational Intelligence (pp. 129-136). IEEE.

Lu, G., Liu, M., Wang, Y., and Yu, G. (2012). Quantifying the severity of traffic conflict by assuming moving elements as rectangles at intersection. Procedia-Social and Behavioral Sciences, 43, 255-264.

Loke, C. H., Adam, M. S., Nordin, R., Abdullah, N. F., and Abu-Samah, A. (2021). Physical Distancing Device with Edge Computing for COVID-19 (PADDIE-C19). Sensors, 22(1), 279.

Mahmud, S. S., Ferreira, L., Hoque, M. S., and Tavassoli, A. (2019). Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment. IATSS research, 43(1), 27-36.

Mattas, K., Makridis, M., Botzoris, G., Kriston, A., Minarini, F., Papadopoulos, B. and Ciuffo, B. (2020). Fuzzy surrogate safety metrics for real-time assessment of rear-end collision risk. A study based on empirical observations. Accident Analysis & Prevention, 148, 105794.

Meivel, S., Devi, K. I., Maheswari, S. U., and Menaka, J. V. (2021). Real time data analysis of face mask detection and social distance measurement using Matlab. Materials Today: Proceedings. DOI: https://doi.org/10.1016/j.matpr.2020.12.1042

Migletz, D., Glauz, W., and Bauer, K., (1985). Relationship between traffic conflicts and accidents. Report No. FHWA/RD-84/042, Federal Highway Administration, Virginia.

Milanés, V., Pérez, J., Godoy, J., and Onieva, E. (2012). A fuzzy aid rear-end collision warning/avoidance system. Expert Systems with Applications, 39(10), 9097-9107.

Mollenhauer, M. A., White, E., and Roofigari-Esfahan, N. (2019). Design and evaluation of a connected work zone hazard detection and communication system for connected and automated vehicles. Virginia Tech Transportation Institute, Report No.03-050

Mukherjee, A., Stolpner, S., Liu, X., Vrenozaj, U., Fei, C., and Sinha, A. (2013). Large animal detection and continuous traffic monitoring on highways. Sensors, 2013 IEEE (pp. 1-3).

Nadimi, N., Behbahani, H., and Shahbazi, H. (2016). Calibration and validation of a new time-based surrogate safety measure using fuzzy inference system. Journal of traffic and transportation engineering, 3(1), 51-58.

National Insurance Crime Bureau (2018), Animal-related insurance claims top 1.7 million in four years, Available at: https://www.nicb.org/news/news-releases/animal-related-insurance-claims-top-17-million-four-years [Accessed June, 2022]

National Highway Traffic Safety Administration (NHSTA) (2019), Fatality and injury reporting system tool (FIRST), motor vehicle crash data querying and reporting, U.S. Department of Transportation, Available at: https://cdan.dot.gov/query [Accessed June, 2022]

National Highway Traffic Safety Administration (NHSTA) (2020a), Fatality Analysis Reporting System (FARS), U.S. Department of Transportation, Available at: https://www-fars.nhtsa.dot.gov/People/PeopleAllVictims.aspx [Accessed June, 2022]

National Highway Traffic Safety Administration (NHSTA) (2020b), Traffic Safety Facts, A Compilation of Motor Vehicle Crash Data, U.S. Department of Transportation, Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813141 [Accessed June, 2022]

National Highway Traffic Safety Administration (NHSTA) (2021), Traffic Safety Facts, Crash Stats, U.S. Department of Transportation, Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813283 [Accessed June, 2022]

National Highway Traffic Safety Administration (NHSTA) (2022), Traffic Safety Facts, Distracted Driving 2020, U.S. Department of Transportation, Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813309 [Accessed June, 2022]

Negnevitsky, M. (2005). Artificial intelligence: a guide to intelligent systems. Pearson education. Available at: http://www.academia.dk/BiologiskAntropologi/Epidemiologi/DataMining/Artificial_Intelligence-A_Guide_to_Intelligent_Systems.pdf [Accessed June, 2022]

Olszewski, P., Osinska, B., Szagała, P., Włodarek, P., Niesen, S., Kidholm Osmann Madsen, T. and Rabjerg Meltofte, K. (2016). Review of current study methods for VRU safety. Part 1–Main report. Available at: https://lup.lub.lu.se/search/publication/403c2502-ae84-426b-b909-f8a21c6ed7ec [Accessed June, 2022]

Parshina-Kottas, Y., Saget, B., Patanjali, K., Fleisher, O., and Gianordoli, G. (2020). This 3-D simulation shows why social distancing is so important. The New York Times. Available at:

https://www.nytimes.com/interactive/2020/04/14/science/coronavirus-transmission-cough-6-feet-ar-ul.html [Accessed June, 2022]

Pathak, R., and Sivraj, P. (2018). Selection of algorithms for pedestrian detection during day and night. Computational Vision and Bio Inspired Computing. Part of the Lecture Notes in Computational Vision and Biomechanics book series (volume 28 pp. 120-133). Springer, Cham.

Peesapati, L. N., Hunter, M. P., and Rodgers, M. O. (2013). Evaluation of post encroachment time as surrogate for opposing left-turn crashes. Transportation Research Record, 2386(1), 42-51.

Sacchi, E., Sayed, T., and Deleur, P. (2013). A comparison of collision-based and conflict-based safety evaluations: The case of right-turn smart channels. Accident Analysis & Prevention, 59, 260-266

Saisana, M., Saltelli, A., and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168(2), 307-323

Sharafsaleh, M., Huijser, M., Kuhn, T., Spring, J., and Felder, J. (2010). Evaluation of an animal warning system effectiveness (No. UCB-ITS-PRR-2010-22).

Shen, Y., Hermans, E., Brijs, T., Wets, G., and Vanhoof, K. (2012). Road safety risk evaluation and target setting using data envelopment analysis and its extensions. Accident Analysis & Prevention, 48, 430-441.

Song, X., Yin, Y., Cao, H., Zhao, S., Li, M., and Yi, B. (2021). The mediating effect of driver characteristics on risky driving behaviors moderated by gender, and the classification model of driver's driving risk. Accident Analysis & Prevention, 153, 106038.

Sugimoto, C., Nakamura, Y., and Hashimoto, T. (2008). Prototype of pedestrian-to-vehicle communication system for the prevention of pedestrian accidents using both 3G wireless and

WLAN communication. 3rd International Symposium in Wireless Pervasive Computing, 2008. ISWPC 2008. pp. 764-767. IEEE.

Statistics Canada (2020). Vital Statistics - Death database 2020. Available at: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310039201 [Accessed June, 2022]

Stipancic, J., Zangenehpour, S., Miranda-Moreno, L., Saunier, N., and Granié, M. A. (2016). Investigating the gender differences on bicycle-vehicle conflicts at urban intersections using an ordered logit methodology. Accident Analysis & Prevention, 97, 19-27.

Tageldin, A., and Sayed, T. (2019). Models to evaluate the severity of pedestrian-vehicle conflicts in five cities. Transportmetrica A: Transport Science, 15(2), 354-375.

Tak, S., Kim, S., and Yeo, H. (2015). Development of a deceleration-based surrogate safety measure for rear-end collision risk. IEEE Transactions on Intelligent Transportation Systems, 16(5), 2435-2445.

Tarko, A.P. (2012). Use of crash surrogates and exceedance statistics to estimate road safety. Accident Analysis and Prevention, 45: 230–240.

Tarko, A. P. (2018). Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution–A theoretical and numerical exploration. Accident Analysis & Prevention, 113, 63-73.

Terzi, R., Sagiroglu, S., and Demirezen, M. U. (2018). Big data perspective for driver/driving behavior. IEEE Intelligent Transportation Systems Magazine.

Traffic Injury Research Foundation (2015). Wildlife-vehicle collisions: 2000-2014. The National Fatality Database, Available at: https://tirf.ca/TIRFCAD17X [Accessed June, 2022]

Transport Canada (2019a). Canadian Motor Vehicle Traffic Collision Statistics: Available at: https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2019 [Accessed June, 2022]

Transport Canada (2019b). Collisions by Roadway Configuration Statistics: Available at: https://wwwapps2.tc.gc.ca/Saf-Sec-Sur/7/NCDB-BNDC/p.aspx?c=100-0-0&l=en [Accessed June, 2022]

Tong, Y., Jia, B., Wang, Y., and Yang, S. (2019). Detecting pedestrian situation awareness in real-time: algorithm development using heart rate variability and phone position. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting. (Vol. 62, No. 1, pp. 1579-1583). Sage CA: Los Angeles, CA: SAGE Publications.

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hume, R. D., Mayer, R. E., and Castellan, N. J. (1979). Tri-level study of the causes of traffic accidents: final report. Executive summary. Indiana University, Bloomington, Institute for Research in Public Safety.

Uzondu, C., Jamson, S., and Lai, F. (2018). Exploratory study involving observation of traffic behavior and conflicts in Nigeria using the Traffic Conflict Technique. Safety Science, 110, 273-284.

Vogel, K. (2003a). Modelling driver behavior-A control theory-based approach. Doctoral thesis Dissertation, University of Linkoping, Sweden.

Vogel, K. (2003b). A comparison of headway and time to collision as safety indicators. Accident Analysis and Prevention, 35: 427–433.

Vuorinen, V., Aarnio, M., Alava, M., Alopaeus, V., Atanasova, N., Auvinen, M., and Hayward, N. (2020). Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors. Safety Science, 130, 104866. DOI:

https://doi.org/10.1016/j.ssci.2020.104866

Wang, T., Cardone, G., Corradi, A., Torresani, L., and Campbell, A. T. (2012). Walksafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads. 12th Workshop on Mobile Computing Systems & Applications. (pp. 1-6).

Wang, C., and Stamatiadis, N. (2014). Derivation of a new surrogate measure of crash severity. Transportation Research Record, 2432(1), 37-45.

Wang, J., Zheng, Y., Li, X., Yu, C., Kodaka, K., and Li, K. (2015). Driving risk assessment using near-crash database through data mining of tree-based model. Accident Analysis & Prevention, 84, 54-64.

World Health Organization (WHO) (2018), Global Status Report on Road Safety 2018. Available at: https://apps.who.int/iris/rest/bitstreams/1164010/retrieve [Accessed June, 2022]

World Health Organization (WHO) (2020). Coronavirus disease (COVID-19) advice for the public. June 4. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [Accessed June, 2022].

World Health Organization (WHO) (2022). Coronavirus disease (COVID-2019) situation report of June 29. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [Accessed June, 2022].

Wu, C., Peng, L., Huang, Z., Zhong, M., and Chu, D. (2014). A method of vehicle motion prediction and collision risk assessment with a simulated vehicular cyber physical system. Transportation Research Part C: Emerging Technologies, 47, 179-191.

Xing, L., He, J., Abdel-Aty, M., Cai, Q., Li, Y., and Zheng, O. (2019). Examining traffic conflicts of up stream toll plaza area using vehicles' trajectory data. Accident Analysis & Prevention, 125, 174-187.

Xing, L., He, J., Abdel-Aty, M., Wu, Y., & Yuan, J. (2020). Time-varying analysis of traffic conflicts at the upstream approach of toll plaza. Accident Analysis & Prevention, 141, 105539.

Xiong, X., Wang, M., Cai, Y., Chen, L., Farah, H., and Hagenzieker, M. (2019). A forward collision avoidance algorithm based on driver braking behavior. Accident Analysis & Prevention, 129, 30-43.

Yang, Z., Yu, Q., Zhang, W., and Shen, H. (2021). A comparison of experienced and novice drivers' rear-end collision avoidance maneuvers under urgent decelerating events. Transportation Research Part F: Traffic Psychology and Behavior, 76, 353-368.

Yannis, G., Nikolaou, D., Laiou, A., Sturmer, Y. A., Buttler, I., and Jankowska-Karpa, D. (2020). Vulnerable road users: Cross-cultural perspectives on performance and attitudes. IATSS research, 44(3), 220-229.

Zadobrischi, E., and Negru, M. (2020). Pedestrian detection based on TensorFlow YOLOv3 embedded in a portable system adaptable to vehicles. In 2020 International Conference on Development and Application Systems (pp. 21-26). IEEE.

Zhao, P., and Lee, C. (2018). Assessing rear-end collision risk of cars and heavy vehicles on freeways using a surrogate safety measure. Accident Analysis & Prevention, 113, 149-158.

Zheng, L., Sayed, T., and Mannering, F. (2021). Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. Analytic Methods in Accident Research, 29, 100142.

CHAPTER TWO

2. DEVELOP A FUZZY RULE-BASED ALGORITHM FOR A NEXT GENERATION ROADSIDE ANIMAL DETECTION SYSTEM

Abstract: Some North American jurisdictions have installed a roadside animal detection system (RADS), a collision avoidances system designed to warn drivers of the possible presence of large animals on a section of rural highway. This study firstly provides a framework for the system to develop a next generation (NG) RADS. The framework has three phases: I) detection, II) risk analysis, and III) communication. This study focuses on the second phase, i.e., developing a risk analysis that can estimate the varying levels of risk posed by animals on the roadway using real-time data on animal behavior, driver behavior and other factors. To estimate the level of risk, the study developed a fuzzy rule-based algorithm that combines four input indicators (e.g., direct distance between animal and vehicle, and stopping sight distance required). The methodology was investigated using real-world traffic and animal data collected from a conventional RADS in British Columbia, Canada. The results showed that the proposed risk analysis in NG RADS has significant advantages over the conventional RADS. In particular, the NG RADS uses varying levels of warning according to the estimated level of the risk rather than the constant level of warning generated by a conventional RADS. The proposed NG RADS can also be designed to use vehicle-to-infrastructure communication technology to establish direct wireless communication with vehicles at risk. In a high-risk situation, this communication could, for instance, automatically control a vehicle's speed to avoid a collision with a large animal.

***Keywords**: Animal collision, Vehicle-to-Infrastructure, Roadside animal detection system, Fuzzy algorithm, Road safety*

2.1. Introduction

Animal-to-vehicle collisions are a challenging safety concern that threatens human lives and leads to economic and ecological losses. The statistics in North America highlight the seriousness of animal-to-vehicle collisions and show that effective methods for reducing the number of collisions involving animals are crucial to improving road safety. For example, the United States recorded 1,739,687 animal-to-vehicle collisions during the four years from 2014 to 2017 (National Insurance Crime Bureau, 2018). Collisions involving large animals (e.g., deer), accounted for 584,165 of the collisions and $2 billion of direct insurance claims each year. In 2017, 194 deaths were associated with collisions involving animals in the United States (NHSTA, 2018). In 2019 in Michigan, 55,531 (17.6%) of the state's 314,377 reported collisions involved large animals. These collisions resulted in 12 fatalities and 1,265 injuries (Michigan State Police, 2020).

In Canada, Vanlaar et al. (2019) reported 474 fatalities due to collisions involving large animals between 2000 and 2014 (15 years). In Ontario, the three most recent years (2016-2018) of collision data include 10,351 collisions involving animals per year, 15% of Ontario's collisions. Animal-to-vehicle collisions were the third largest type of collision in the province (after inattentive driving collisions (13,663 collisions per year (20%)) and speed-related collisions (11,296 collisions per year (16%)) (Ontario Provincial Police, 2018). The Traffic Injury Research Foundation (2017) reported that 28% of Canadian drivers experienced one or more near-miss animal collisions in 2016.

2.1.1. Traditional Animal-To-Vehicle Collisions Mitigation Methods

More than 40 mitigation methods (roadway lighting, roadside clearing, etc.) have been proposed to reduce the number of animal-to-vehicle collisions (Huijser et al., 2007). Each approach has its advantages and disadvantages. The most widely used mitigation methods are static animal warning signs, animal fencing, underpasses, and overpasses (e.g., eco-bridges).

Static animal warning signs are the most affordable method, but their effectiveness appears limited (Bond and Jones, 2013). Rogers and Premo (2004), for instance, reported that the installation of static animal warning signs did not significantly reduce the number of animal-to-vehicle collisions. Animal fencing is highly effective, reducing the number of animal-to-vehicle collisions by around 87%. Initial installation costs are around $112,500 (2006 USD) for a 1.5 km section of highway, but animal fencing is considered ecologically unfriendly and environmentally undesirable as it blocks crossing opportunities for animals and may destroy continuity of animal habitat. Crossing structures such as underpasses and overpasses are expected to reduce animal-to-vehicle collisions by around 86% when combined with fencing. Ecologically, underpasses and overpasses can provide continuity of animal habitat, but where traffic is heavy, effectiveness may be reduced by traffic noise deterring animals from using the structures especially underpasses. In addition, underpasses and overpasses are the most expensive mitigation methods. For example, a 60 m wide overpass crossing a two-lane rural highway in Banff National Park, Alberta, Canada cost $3.76M in 2007 (Huijser 2008).

2.1.2. Roadside Animal Detection System (RADS)

A conventional Roadside Animal Detection System (RADS) is an alternative mitigation approach. A RADS is a relatively modern ITS technology designed to detect the presence of large animals near or on the road and to warn approaching drivers of their presence by activating dynamic flashing warning lights (Huijser et al., 2006). When the warning lights of a RADS flash, drivers usually decrease speed which increases the time available to react to animals on the roadway (Huijser et al., 2006; Grace et al., 2017). In principle, animal fencing and crossing structures attempt to change animal behavior to improve road safety whereas a RADS attempts to increase drivers' awareness of the presence of animals and change human drivers' behavior by providing a timely warning message. Providing a warning that changes human behavior appears to be an easier goal than changing animal behavior.

A typical RADS has two phases. In the first phase, the system uses a sensor (e.g., perimeter surveillance radar) to detect the presence of one or more animals near or on the roadway. In the second phase, the system uses an alarm (e.g., flashing warning lights) to warn approaching vehicles of the animal presence. The collision reduction effect of a RADS is reported as 87% for collisions involving large animals such as moose and deer (Huijser et al., 2008).

A RADS has important advantages. The cost is around $130,000 (2014 CAD) for a 1.5 km section of highway. This is similar to the cost of animal fencing and very much less expensive than crossing structures. As a RADS does not restrict animal movements, the system is far more ecologically friendly than animal fencing or crossing structures (Grace et al., 2017). An additional advantage is that the system can be used to collect and monitor real-time 24/7 traffic flow data (e.g., traffic volume, vehicle speed and vehicle classification) (Mukherjee et al., 2013). The data collected may provide valuable inputs for road safety or other transportation projects.

Several studies have investigated conventional RADS installed on rural highways in Europe and North America, Huijser et al. (2006) reviewed 34 RADS sites (22 in Europe and 12 in North America). All the sites reviewed were conventional RADS that produced a binary output (on/off warning lights) based on a binary input (presence/absence of an animal), i.e., all the installations simply activated the flashing warning lights during the time that the system detected the presence of an animal within the detection area, and simply turned off the warning lights shortly after the animal moved outside of the detection area. Other studies reporting on such systems include Grace et al. (2017), Mukherjee et al. (2013), Sharafsaleh et al. (2010), Druta et al. (2018), Antônio et al. (2019) and Druta et al. (2020).

A conventional RADS system does not consider possible differences in factors such as the animal's behavior (e.g., unpredictable jumping, moving slowly, stopping, etc.), the vehicle(s) (e.g., leading/following vehicle(e)), the position of the animal relative to the roadway (on/off the roadway), the

spatial and temporal interactions between the animal and the vehicle (in meters or time), driver behavior (aggressiveness), etc.

Situations where animals are encountered may differ considerably. For example, we can imagine that 1) an animal may be walking or running parallel to the roadway a few meters away from the edge of a roadway, or 2) the animal may be walking or running on the roadway or crossing the roadway. It can be reasonably assumed that the latter situation is the riskier situation for approaching vehicles and that it needs more immediate attention for drivers than does the first situation. We can imagine another example in which two or more vehicles are traveling closely behind each other. In this case, it can also be reasonably assumed that the leading vehicle would be in a riskier situation than the following vehicle(s) when animals are on the roadway assuming all other conditions (e.g., vehicle type and speed) are the same for each vehicle. The current RADS does not tailor the warning to the different levels of risk facing the vehicles involved. This limitation is due mainly to the fact that a current RADS' warning relies on a simple binary input/output regardless of different and changing animal and/or traffic situations.

2.1.3. Study Goal and Objectives

This study aims to create a framework for a Next Generation Roadside Animal Detection System (NG RADS). The framework relies on developing a risk analysis algorithm to create a composite risk indicator from a combination of multiple input indicators. The study has three objectives:

1) Identify appropriate input risk indicators associated with animal' diverse activities, and traffic situations;
2) Develop an advanced algorithm for a NG RADS to estimate the composite risk indicator of a collision with an animal and disseminate levels of warning for each approaching vehicle on a real-time basis; and
3) Demonstrate the possible usefulness of the proposed algorithm using a real-world study.

Section 2.2 of this chapter addresses the framework for the NG RADS. Section 2.3 presents the proposed framework and its application in the case study. Section 2.4 discusses the results of the case study. Section 2.5 discusses the study's findings and recommendations for future work.

2.2. Proposed Framework for a Next Generation Roadside Animal Detection System

Figure 2-1 shows the three-phase framework of our proposed NG RADS. The framework is adapted from the framework for a military air defense system (Johansson and Falkman, 2008). The three phases are I) detection, II) risk analysis, and III) communication.

I. Phase I (detection): The system uses radar sensors to detect the characteristics of the animal (e.g., position, direction of movement, and speed) and the vehicle on a real-time (second by second) basis. The characteristics are used as seed information for Phase II.

II. Phase II (risk analysis): In the input indicator selection, the system uses the seed information from Phase I to develop various input indicators to capture animal behavior, driver behavior, and the spatial and temporal interaction between the animal and the vehicle. These indicators include, but not limited to the magnitude of speeding (the difference between the driver's speed and the posted speed limit), and the direct distance between the animal and the vehicle. The system then estimates various levels of risk using a fuzzy rule-based algorithm that combines all the information into a composite risk indicator.

III. Phase III (communication): The system can use advanced Vehicle-to-Infrastructure (V2I) communication technologies to broadcast different levels of warning according to the estimated risk. Different levels of warning can be directed to different vehicles on a real-time basis to alert drivers to the need to take appropriate action to avoid a collision. One possibility is that we can use different colors representing different levels of warning. Another possibility is to broadcast

audible and/or visual warnings via the built-in audio and/or visual display system of the approaching vehicles.

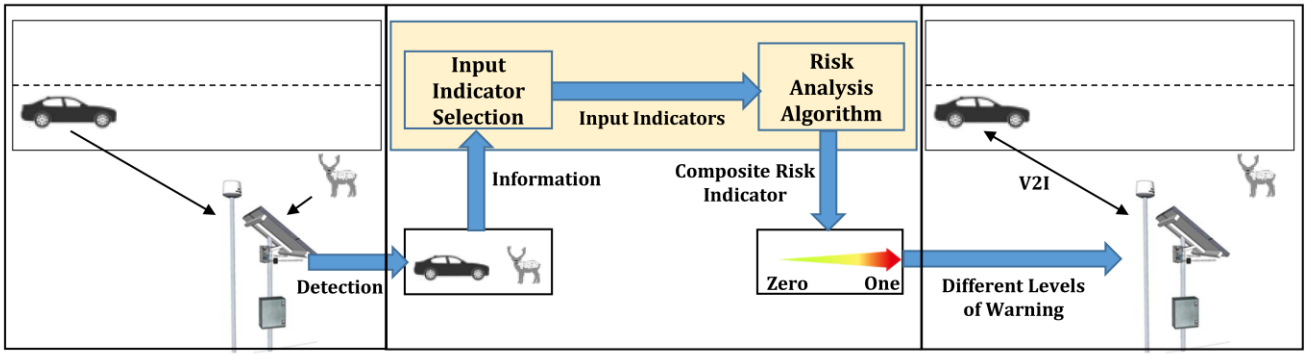


Figure 2-1: Proposed framework of NG RADS

This study focuses on addressing Phase II (risk analysis). As a conventional RADS system can already detect seed information such as the animal and vehicle positions and velocities second by second, the technology needed for Phase I (detection) already exists. Phase III (communication) requires advanced V2I communication technologies (e.g., wireless, bi-directional communication using short- and medium-range frequencies) to transfer data and deliver the alarm of a potential risk. With the rapid and on-going development of V2I technology and wireless communication technology (e.g., 5G), communication between a NG RADS system and vehicles approaching within a considerable distance (e.g., 2 km with two sets of RADS installed) is an achievable goal in the foreseeable future (Rahman and Tepe, 2014).

Section 2.2.1 (input indicator selection) and Section 2.2.2 (risk analysis algorithm) discuss the Phase II issues in detail.

2.2.1. Input Indicator Selection

The purpose of the input indicator selection is to capture animal behavior, driver behavior, and the spatial and temporal interaction between animal and vehicle through the selection and measurement of appropriate indicators.

A review of research into collisions that involved an animal found that the indicators used to trigger the warning messages of conventional RADS systems (Grace et al., 2017; Mukherjee et al., 2013; Sharafsaleh et al., 2010; Druta et al., 2018; Antonio et al., 2019; Druta et al., 2020) were limited to the presence/absence of an animal within a detection area. As safety indicators are usually designed to analyze vehicle-to-vehicle collisions (Mahmud, et al., 2017) rather than animal-to-vehicle collisions, additional indicators were required. This study proposed four indicators for the input indicator selection: 1) Indicator A (lateral distance from animal to roadway), 2) Indicator B (magnitude of speeding), 3) Indicator C (stopping sight distance; SSD), and 4) Indicator D (direct distance). The indicators are shown in Figure 2-2.

1) Indicator A (lateral distance from animal to roadway). Animals may attempt to cross a roadway at any time and at any location (Lima et al., 2015). In principle, a conventional RADS simply uses Indictor A as its sole input indicator, i.e., if a RADS detects the presence of an animal within a pre-set lateral (i.e., perpendicular) distance (e.g., 5 meters) from a roadway, the system turns on the flashers installed on the roadside (see Figure 2.1). Figure 2-2a shows Indicator A. As the lateral distance from an animal to the roadway decreases, the situation's risk increases. Notice, however, that this distance cannot capture all the possible risks to vehicles using the road. If, for example, an animal is moving parallel to the roadside, Indicator A will hardly change although such a situation clearly includes the possibility of a greatly increased level of risk if the distance between vehicles and animals decreases. A conventional RADS does not consider this possibility although animals in such a situation may clearly become nervous, and suddenly change course. A conventional RADS cannot differentiate between drastically different risky situations such as "animals are walking on the roadway and moving towards a vehicle" and "animals are walking off the roadway and moving away from vehicles."

2) Indicator B (magnitude of speeding). Driver behavior plays an important role in the collisions.

It is reasonable to assume that aggressive driving may increase the probability of a collision with an animal. This study used magnitude of speeding to assess a driver's aggressiveness. The study assumed that drivers traveling under the posted speed limit tend to be conservative drivers (with a reduced chance of any kind of collision) and that drivers who exceed the posted speed limit tend to be aggressive drivers (with an increased chance of any kind of collision) (Castignani, et al., 2015; Eftekhari and Ghatee, 2018).

Magnitude of speeding refers to the difference between the driver's speed and the posted speed limit. Figure. 2-2b and Equation (2-1) explain Indictor B. As the magnitude of speeding increases, the situation's risk increases.

$$MS_i = TS_i - PSL \tag{2-1}$$

where:

MS_i = magnitude of speeding for vehicle *i* (km/h);

TS_i = travel speed for vehicle *i* (km/h); and

PSL = posted speed limit on the highway segment considered (km/h).

A higher driving speed requires a longer stopping sight distance to allow for braking and avoiding a collision. In the United States, speeding accounted for 29% of traffic fatalities, 13% of injury crashes, and 10% of property-damage-only collisions in 2020 (NHSTA, 2020). We used magnitude of speeding (speed higher than posted speed limit) to capture aggressive driver behaviour.

Several studies have clearly shown the impact of changes in the posted speed limit on the number of animal-vehicle collisions (Zachary and Danks 2010; Meisingset et al., 2014). For example, Zachary and Danks (2010) showed that an 8-km/hr increase in the posted speed limit increased the chance of a moose-to-vehicle collision by 35%, and Meisingset et al. (2014) showed that an increase in the speed limit from 50 to 70 km/hr increased the relative risk of an animal-vehicle collision by a factor of 6.8, and that an increase from 50 to 80 km/hr increased the relative risk by a factor of 14.4.

Indicators A and B do not assess the spatial and temporal interaction between the animals and the vehicles. These interactions are assessed using Indicators C (SSD) and D (direct distance).

3) Indicator C (stopping sight distance; SSD). SSD is an estimate of the minimum distance at which a driver notices an object (in this study, an animal) on the roadway and can make a complete stop to avoid the collision with the object (AASHTO, 2018; TAC, 2019). See Fig. 2-2c. SSD is an important highway design parameter that considers a variety of safety elements including perception-reaction time, speed, and friction coefficient as shown in Equation (2-2). In this study, Indicator C represents a binary condition. If the distance between the vehicle and the nearest point of conflict is less than the SSD, the situation is considered risky. If the distance is longer than the SSD, the situation is considered safe.

$$SSD = Vt_{pr} + \frac{V^2}{2\mu(f \pm g)} \quad (2\text{-}2)$$

In Equation (2):

SSD = stopping sight distance (m);

V = vehicle speed (m/s);

t_{pr} = perception-reaction time (s);

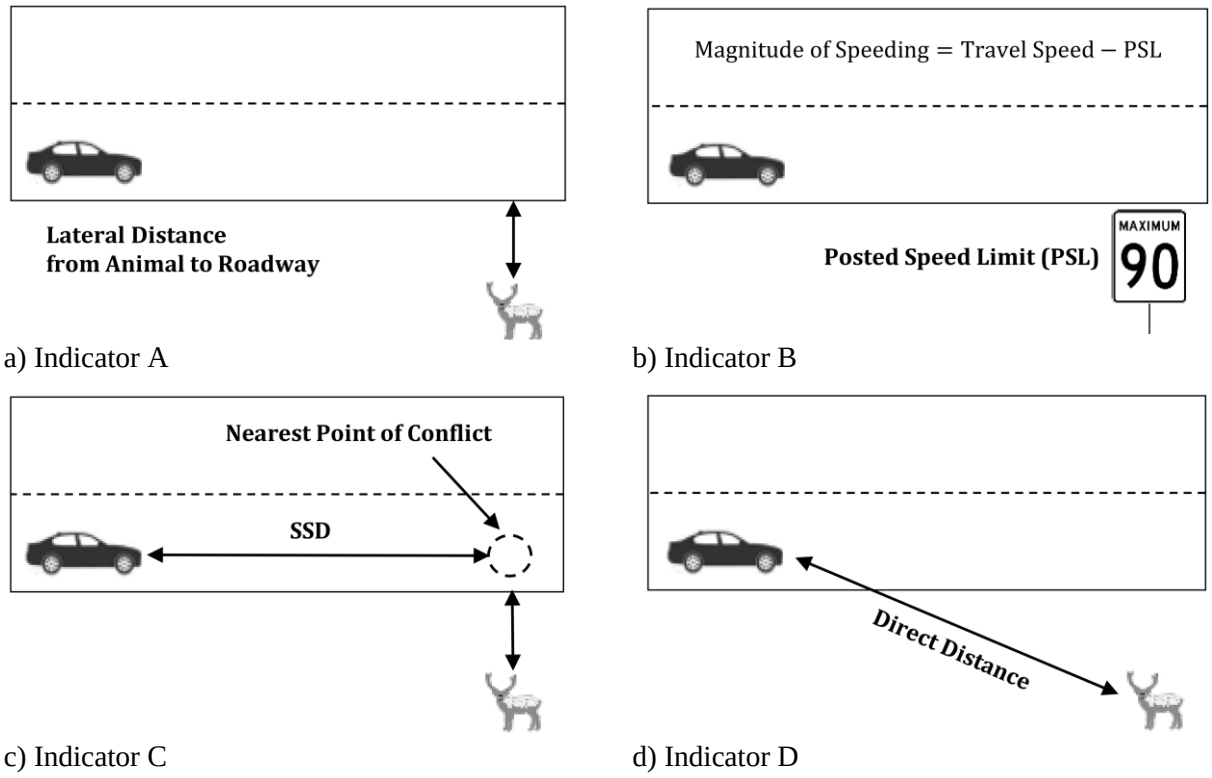
μ = 9.8 m/s^2;

f = friction coefficient; and

g = roadway grade.

4) Indicator D (direct distance): In certain situations, Indicators A, B and C cannot capture the risk and will not trigger a warning for approaching vehicles. For example, if an animal is located outside the pre-set distance from a roadway, an approaching vehicle is traveling at less than the posted speed limit, and the SSD estimated for the vehicle appears to be adequate, the RADS may not trigger a warning. However, animal behavior is largely unpredictable, and an animal's speed and direction of movement may

suddenly change. For example, animals such as moose and deer sometimes unexpectedly jump directly towards a moving vehicle as if attacking the vehicle. Frid and Dill (2002) explained this behavior using the "fight-or-flight theory" originally introduced by Cannon (1916), more than a century ago. Indicator D (see Figure 2-2d) was used to assess the level of risk related to unpredictable behavior by taking into account the actual distance between an animal and a vehicle to capture the varying level of risk associated with an animal suddenly changing course and other movements that reduce the distance between an animal and a vehicle. As direct distance decreases, the situation's risk increases.



a) Indicator A b) Indicator B

c) Indicator C d) Indicator D

Figure 2-2: The four indicators used for input indicator selection

Although the study did not include a separate indicator for the running speed of a moving animal, three of the four proposed indicators (Indicators A, C and D) indirectly reflect the speed of a moving animal as the estimated values are updated every second. As an example, Indicator A (lateral distance from animal to roadway) will be decreasing faster for an animal each second that is moving towards the

roadway faster than an animal that is moving slowly towards the roadway.

Also, although this study did not use the traveling speed of vehicles as a separate indicator, Indicators C (SSD) and D (direct distance) indirectly capture the impact of travel speed of vehicle. Equation (2-2) shows that the SSD required increases with higher vehicle travel speed. Higher vehicles speeds will produce a higher level of risk than do lower vehicle speeds. As the proposed system analyzes the situation second by second, the direct distance captures the changing effect of vehicles' traveling speed.

2.2.2. Risk Analysis Algorithm

Artificial intelligence algorithms used in risk analysis require combination of the various input indicators to estimate varying levels of risk. Algorithms that combine multiple input indicators include statistical algorithms, machine learning algorithms, and fuzzy rule-based algorithms (Ding et al., 2019). The choice of an appropriate algorithm depends largely on the nature of the problem and issues such as the complexity of the interrelationships, the causes of uncertainty, the availability of data, the type of data (qualitative or quantitative), etc. (Balmat et al., 2011). Teodorović (1994) noted that a collision avoidance system (CAS) in intelligent transportation systems often needs to deal with traffic flows in situations where data are very uncertain and/or scarce. RADS installations are usually on highway segments with little traffic where limited and missing observed data are major constraints particularly when creating a rural CAS (as in this study).

Most algorithms require a relatively large volume of input data to calibrate the system, and some input parameters may involve a range of uncertainty (e.g., the driver perception-reaction time used in Equation (2-2) varies and is affected by a driver's health age, quality of vision, etc.). Calibrating a RADS requires consideration of the number of actual or near miss collisions between vehicles and animals, but as these collisions are rare, it is not possible to observe and record the large amount of data required for

system calibration. In addition, collecting detailed driver information in the field can be costly and/or impossible due to the privacy by-laws of some jurisdictions.

Fuzzy rule-based algorithms are designed to capture and analyze uncertainty without requiring a large amount of data. They were originally proposed by Zadeh (1965). A fuzzy rule-based algorithm uses a membership function to determine a membership degree between 0 and 1 for each input indicator. The membership function can be changed into a "fuzzy set" designed to capture the uncertainty of the input and output variables. These fuzzy sets can be defined as linguistic variables (safer or riskier in our case) and can evaluate binary situation such as yes/no, true/false, etc. Notice that fuzzy rule-based algorithms have been widely used for many different transportation safety studies (Shrinivasan, et al., 2007; Shrinivasan, et al., 2018). A fuzzy rule-based algorithm has three main stages: 1) fuzzification, 2) rule evaluation, and 3) deffuzication (Mendel, 1995). In the fuzzification stage, we used fuzzy linguistic variables to convert the four indicators defined in Section 2.2.1 into input fuzzy sets and we determined the shape of the membership functions for each input indicator. In the fuzzy rules stage, we defined linguistic IF-THEN rule constructions. For example, a fuzzy rule can be defined as: if distance from animal to roadway (indicator A) is "safer," and magnitude of speeding (Indictor B) is "safer," and SSD (Indicator C) is "safer," and direct distance (Indictor D) is "safer," then output fuzzy set is "very low."

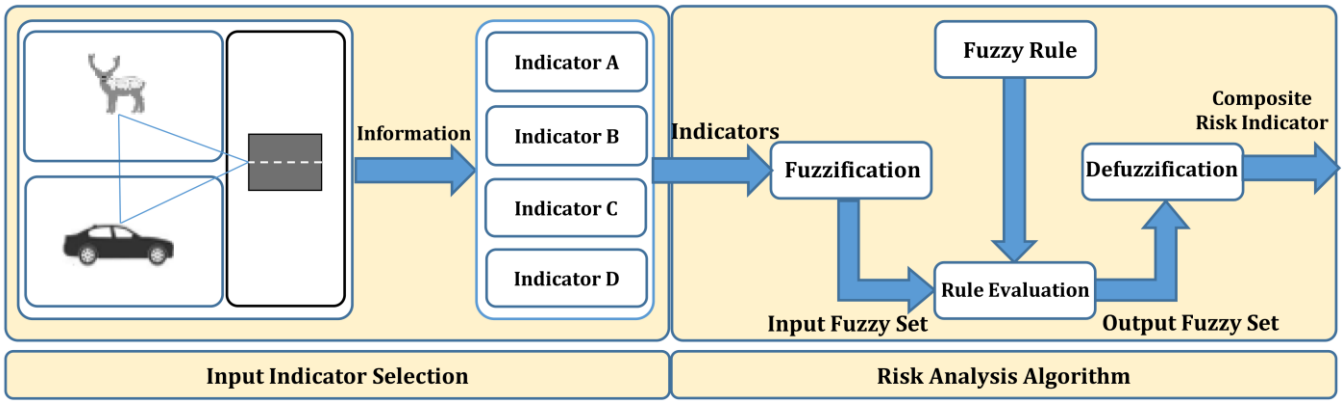


Figure 2-3: Input indicator selection and risk analysis algorithm

The defuzzification stage converts the fuzzy output set to risk. Figure 2-3 shows the concepts involved in the input indicator selection and the risk analysis algorithm.

2.3. Application of the Proposed Algorithm to a Real-World Case

2.3.1. Study Data

In this study, an "event" was defined as a situation in which the movement of one or more animals and the movement of one or more vehicles were detected during the same time window for at least 10 seconds. To demonstrate the application of the proposed fuzzy rule-based algorithm to a real-world situation, the study analyzed two months of RADS data (June, 2018 to July, 2018) from a RADS that covers a 1.4 km section of a two-lane rural highway (Trans-Canada HWY 1) in the Glacier National Park, British Columbia, Canada. The posted speed limit of the highway section is 90 km/h. During the data collection period, the RADS detected 35 events. For all the events, animals were detected close to the roadway (lateral distance from animal to roadway was less than 15 meters).

2.3.2. Calibration of the Fuzzy Rule-Based Algorithm

A five-step calibration process was required to localize the fuzzy rule-based algorithm to our study location: 1) determination of input fuzzy set, 2) determination of membership function for input fuzzy set, 3) determination of boundary values for membership function, 4) determination of output fuzzy set and fuzzy rule, and 5) clarification of stopping criterion in fuzzy rule-based algorithm.

2.3.2.1. Determination of Input Fuzzy Set: This study assigned the most basic number (two) of input fuzzy sets for each indicator (see Table 2-1).

Table 2-1: Input fuzzy sets for indicators

Input Indicator	Input Fuzzy Set
A. Lateral distance from animal to roadway	Safer, Riskier
B. Magnitude of speeding	Safer, Riskier
C. SSD	Safer, Riskier
D. Direct distance	Safer, Riskier

2.3.2.2. Determination of Membership Function for Input Fuzzy Set: An appropriate shape for the membership function is then determined for each input fuzzy set. We used a "trapezoid" membership function as this is the most commonly used functional form in most practical applications (Barua et al, 2014).

2.3.2.3. Determination of Boundary Values for Membership Function: The boundary values of a membership function play an important role in a fuzzy rule-based algorithm. In this study, the boundaries were determined by using information published in the relevant literature and by analyzing field data collected from the RADS installed in the study area.

The British Columbia clear zone regulations were consulted to establish the boundary values for Indicator A. Table 2-2 shows the clear zone width currently applied in British Colombia for rural highways where the average daily traffic is between 1,500 and 6,000. Table 2-2 shows how the clear zone width increases as the design speed increases (BC MoTI, 2019).

Table 2-2: Clear zone width by design speed in British Colombia for rural highways with average daily traffic of 1,500 to 6,000

Design Speed (km/h)	Clear Zone Width (m)
<70	3.5 - 4.5
80	5.0 - 5.5
90	6.0 - 6.5
100	8.0 - 9.0
>110	8.5 – 10.0

Fitzpatrick et al. (2003) analyzed a significant amount of field data to investigate the relationship between design speed, posted speed, and operating speed. The study found that "*while a relationship between operating speed and posted speed limit can be defined, a relationship of design speed to either operating speed or posted speed cannot be defined with the same level of confidence.*" The study also found that design speed appears to have minimal impact on operating speeds unless a tight horizontal radius or vertical curve is present. The study further noted that the posted speed limit sometimes exceeds

the design speed which may result in liability concerns. Unfortunately, we did not have clear information about the design speed for our study section. As a result, this study used the minimum and maximum clear zone widths presented in Table 2-2 and set the boundary values for the input fuzzy set of Indicator A as 3.5 m and 10 m respectively.

The boundary values for Indicator B were estimated from the vehicle travel speeds collected by the RADS at the study site. Figure 2-4 is a boxplot showing the distribution of vehicle travel speeds. As the posted speed limit of the study section is 90 km/h, the boundary values for the fuzzy sets for Indicator B were 0 km/h (90 km/h – posted speed limit (90 km/h)) and 30 km/h (120 km/h – posted speed limit (90 km/h)), i.e., the magnitude of speeding varied from 0 km/h to 30 km/h.

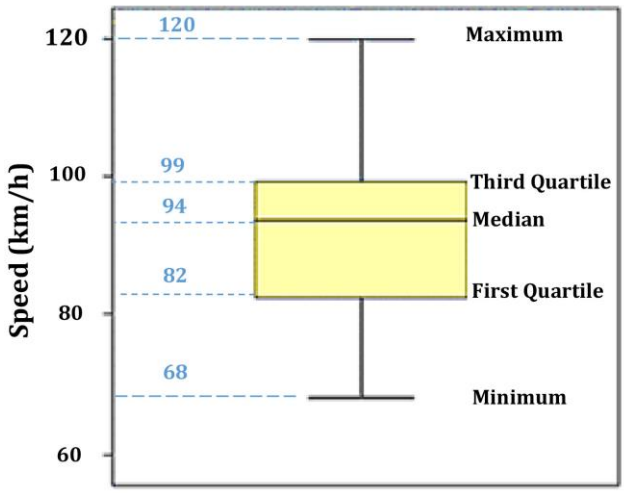
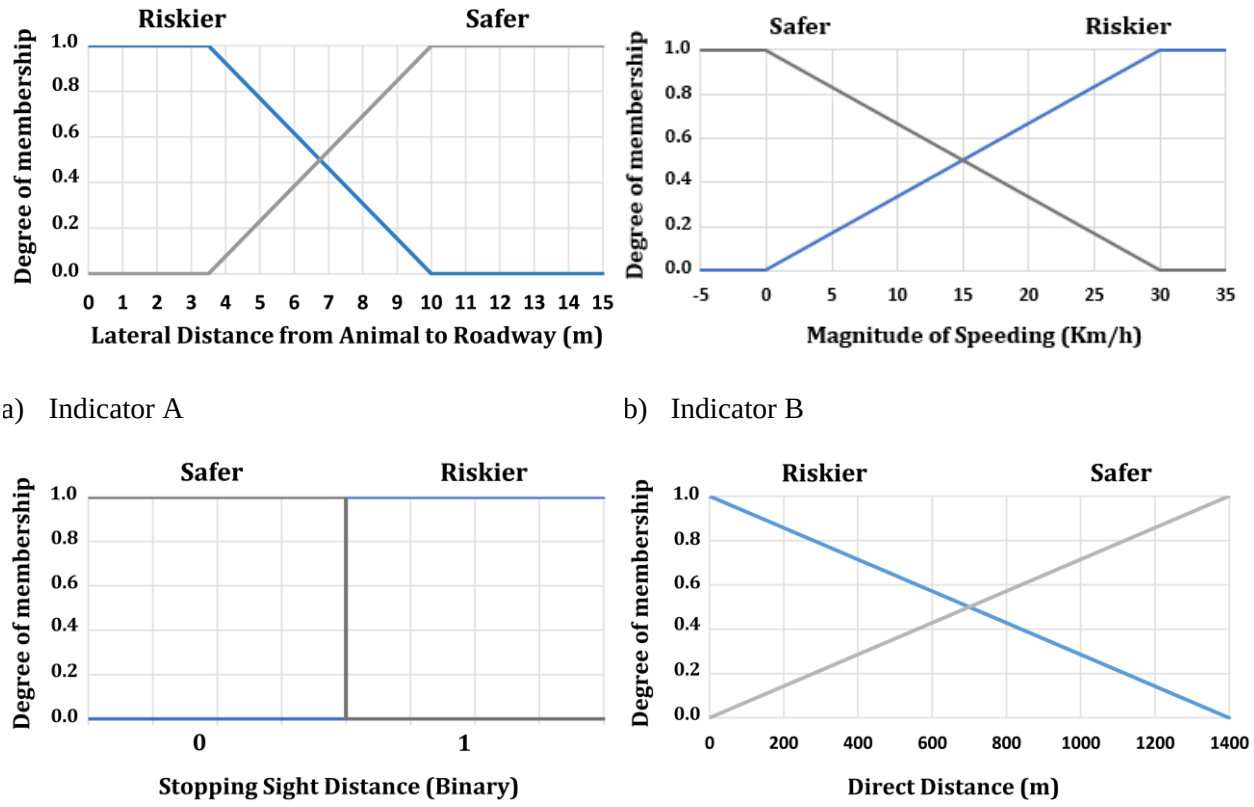


Figure 2-4: Distribution of vehicle travel speeds

The boundary value for Indictor C was based on the SSD presented in the most recent North American geometric guidelines (AASHTO, 2018; TAC, 2019). This study estimated the distance from an approaching vehicle to the nearest point of conflict on the roadway and compared with the SSD. If the distance obtained was positive (higher than or equal to zero), the situation was considered safe. If the distance obtained was negative (lower than zero), the situation was considered risky. Therefore, SSD has a binary value of zero represents safer and one represents riskier.

To determine the boundary condition for Indicator D, this study used the maximum distance detectable by the RADS. This distance, approximately 1.4 km, was the distance used as the maximum boundary value representing the maximum direct distance between an animal and a vehicle. As the theoretical minimum direct distance between an animal and a vehicle is zero (the distance at which a collision occurs), the minimum boundary value for the input fuzzy set for Indicator D was zero.

Figure 2-5 shows the boundary values and functional form of the input fuzzy sets used for the study's four indicators.



a) Indicator A

b) Indicator B

c) Indicator C

d) Indicator D

Figure 2-5: The boundary values and membership function of input fuzzy sets

2.3.2.4. Determination of Output Fuzzy Set and Fuzzy Rule: The determination of fuzzy rules provides an important link between the four indicators and the composite risk indicator and allows us to apply IF-THEN rule constructions. As there are four indicators and each indicator has two input fuzzy sets (safer

and riskier), the total number of combinations becomes 16 (= 2^4). Each combination presents a distinct fuzzy rule.

To determine the 16 fuzzy rules, it was necessary to define the output fuzzy sets associated with the composite risk indicator. The procedure suggested in the literature was used to generate pre-defined output fuzzy sets (Phillis and Davis, 2009)

We assigned an integer value of 0 (safer) or 1 (riskier) to the two input fuzzy sets for each indicator. Each fuzzy rule maps each input indicator (A, B, C, D) to the output fuzzy set for the composite risk indicator (φ,) where $(A, B, C, D) \in \{0,1\}$, and $\varphi = A + B + C + D$.

The minimum φ value of 0 (= $0 + 0 + 0 + 0$) occurs when all four input fuzzy sets are assigned as "safer," and the maximum φ value of 4 (= $1 + 1 + 1 + 1$) occurs when all four input fuzzy sets are assigned as "riskier." In this study, there were five output fuzzy sets as φ could have one of five values $\{0, 1, 2, 3, 4\}$. The five output fuzzy sets were interpreted as ranging from a "very low {0}" risk to a "very high {4}" risk. The five output fuzzy sets were mapped evenly as shown in Figure. 2-6 to produce the composite risk indicator that varies from 0 to 1.

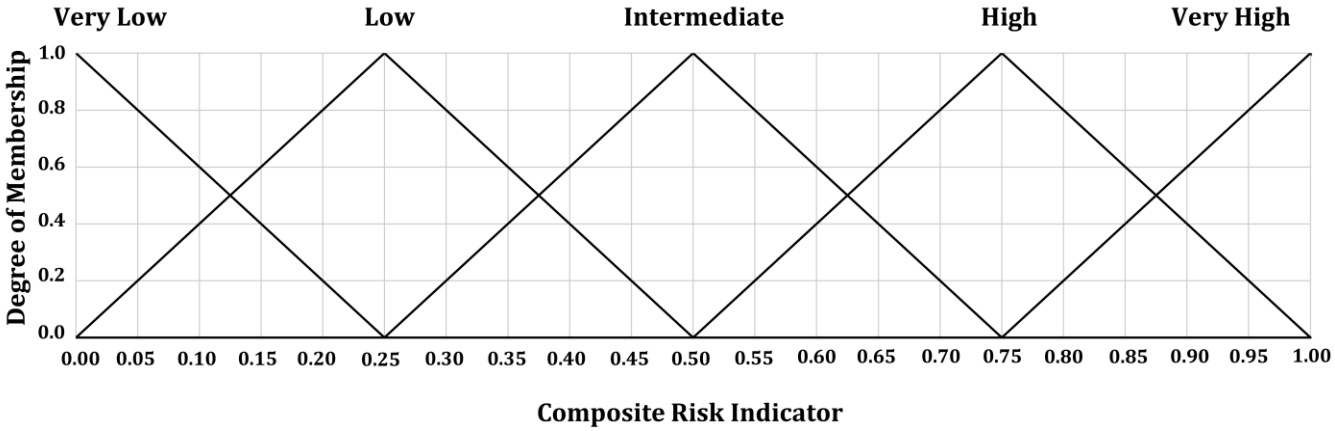


Figure 2-6: Membership function for output fuzzy sets

Here, we explain the risk and the relationship between four input indicators and the composite risk indicator in detail. The risk of a collision is defined as the chance (probability) of a collision given the temporal and spatial interaction between road users (i.e., road user closeness in space and time) (Johnsson et al., 2018).

To capture the risk, we selected four input indicators each of which had two fuzzy sets ("riskier" and "safer"). Figure 2-5 shows each fuzzy set for input indicators together with the boundary values. Furthermore, the composite risk indicator (output) has five fuzzy sets: "very-low", "low", "intermediate", "high", and "very-high". Figure 2-6 shows each fuzzy set for the composite risk indicator together with the boundary values. Table 2.3 shows the 16 rules we defined to connect the four input indicators and their fuzzy sets with the five levels of the composite risk indicator. For example:

- If the fuzzy sets of the four input indicators are all "safer," then the fuzzy set of the composite risk indicator is "very low" (see rule# 1);

Table 2-3 The fuzzy rule-based algorithm's 16 rules

Rule ID	Indicator_A	Indicator_B	Indicator_C	Indicator_D	Composite Risk Indicator
1	Safer	Safer	Safer	Safer	Very Low
2	Riskier	Safer	Safer	Safer	Low
3	Safer	Riskier	Safer	Safer	Low
4	Safer	Safer	Riskier	Safer	Low
5	Safer	Safer	Safer	Riskier	Low
6	Riskier	Riskier	Safer	Safer	Intermediate
7	Riskier	Safer	Riskier	Safer	Intermediate
8	Riskier	Safer	Safer	Riskier	Intermediate
9	Safer	Riskier	Riskier	Safer	Intermediate
10	Safer	Riskier	Safer	Riskier	Intermediate
11	Safer	Safer	Riskier	Riskier	Intermediate
12	Riskier	Riskier	Riskier	Safer	High
13	Riskier	Safer	Riskier	Riskier	High
14	Riskier	Safer	Riskier	Riskier	High
15	Safer	Riskier	Riskier	Riskier	High
16	Riskier	Riskier	Riskier	Riskier	Very High

- If the fuzzy sets of the four input indicators are all "riskier," then the fuzzy set of the composite risk indicator is "very high" (see rule# 16); and
- If the fuzzy sets of two indicators are "safer" and the fuzzy sets of two other indicators are "riskier," then the fuzzy set of the composite risk indicator is "intermediate" (see rules #6-11).

Figure 2-7 shows two examples of a risk value estimated by our fuzzy rule-based algorithm. The first example is the vertical blue line which intersects the "Intermediate" fuzzy set a 0.8 degree of membership (top horizontal blue line) and intersects the "Low" fuzzy set with a 0.2 degree of membership (bottom horizontal blue line) resulting in a risk value of 0.45. The second example is the vertical red line which intersects the "Intermediate" fuzzy with a 1.0 degree of membership (top horizontal red line) and intersects the "Low" fuzzy set with a 0.0 degree of membership (bottom horizontal red line) resulting in a risk value of 0.50.

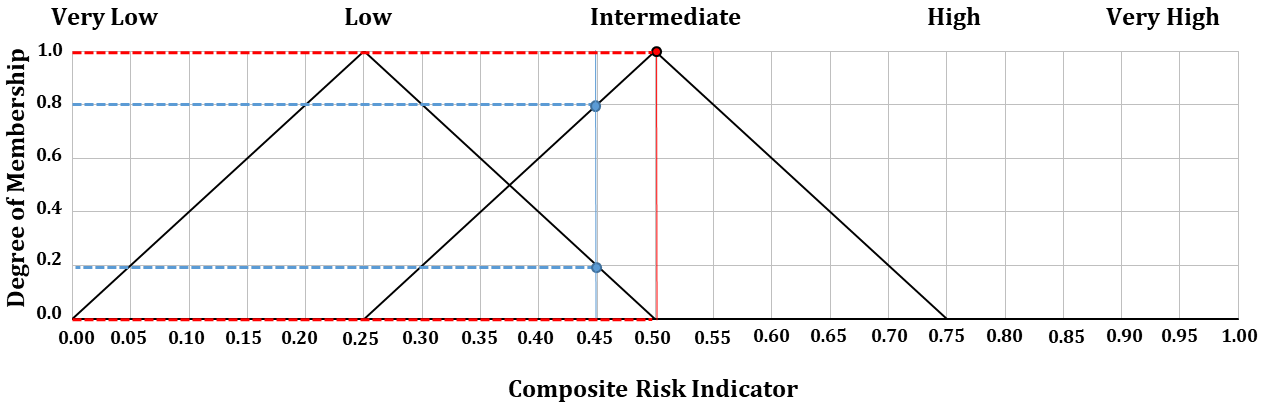


Figure 2-7. An example of two risk values generated by fuzzy rule-based algorithm

2.3.2.5. Clarification of stopping criterion in fuzzy rule-based algorithm: We defined the fuzzy rule-based algorithm's stopping criterion as being when the algorithm generates a zero risk. The stopping criterion refers to the moment when a vehicle passes the nearest point of conflict with an animal (see Figure 2-2) and the animal cannot follow the vehicle even at its maximum running speed. As the collisions involved with animals were mostly with bear (max. running speed = 48 km/h), moose (max. running speed = 56

km/h), or deer (max. running speed = 64 km/h) (The Wildlife Collision Prevention Program, 2019), we used 64 km/h as the maximum speed of an animal. As all of the input data are provided and analyzed second by second, the estimated risk can suddenly become zero in the next second. Note that if the input data periods were less than a second (e.g., a tenth of a second), the estimated value of risk would decrease more gradually.

2.4. Results of Analysis of Case Study

The analysis used the Matlab Simulink package. This study randomly selected one of the 35 real-world events detected by the study area RADS to demonstrate the proposed approach. The event selected showed the movement of one animal and two vehicles over 13 consecutive seconds. See Figure 2-8. The leading vehicle was vehicle #1 and the following vehicle was vehicle #2. During the 13 seconds, vehicle #1 increased its speed from 90.5 km/h to 100.5 km/h, and vehicle #2 decreased its speed from 106 km/h to 87 km/h.

Figure 2-8: The movement of one animal and two vehicles over 13 consecutive seconds of the event.

Figure. 2-9 shows the changes in the four indicators (labeled a) to d)) and the changes in the estimated composite risk indicator (labeled e)) during the 13 seconds of the event.

1) Indicator A (lateral distance from animal to roadway). Figure. 2-9a shows that an animal 10.9 m from the roadway moved closer to the roadway and was 1.9 m away at 13 seconds. Indicator A can be expected to contribute to increasing the value of composite risk indicator for both vehicles as time passes.

2) Indicator B (magnitude of speeding). Figure 2-9b shows that the magnitude of speeding increased from 0.5 km/h at the start of the event to 10.5 km/h for vehicle #1 and decreased from 16.4 km/h to 0 km/h for vehicle #2. Indicator B can be expected to contribute to increasing the value of composite risk indicator for vehicle #1 and decreasing the value of composite risk indicator for vehicle #2 as time passes.

3) Indicator C (stopping sight distance). For vehicle #1, Figure 2-9c shows that the SSD increased from 0 to 1m at 5 seconds, i.e., at 5 seconds, vehicle #1 could not stop in time if the animal jumped onto the nearest point of conflict. For vehicle #2, the same situation arose at 7 seconds. Furthermore, for vehicle #1, SSD does not have value after second 8 when vehicle passes from animal. For vehicle #2, the same situation arose at 12 seconds. Indicator C can be expected to contribute to increasing the value of composite risk indicator at 5 seconds for vehicle #1 and at 7 seconds for vehicle #2.

4) Indicator D (direct distance). Figure. 2-9d shows that for vehicle #1, the direct distance decreased until 9 seconds and then increased. For vehicle #2, the direct distance decreased until 11 seconds and then increased. For the first 9 seconds for vehicle #1 and for the first 11 seconds for vehicle #2, Indicator D can be expected to contribute to increasing the value of composite risk indicator. The risk can then be expected to decrease to zero (no risk) as the vehicles increase their distance after passing the animal.

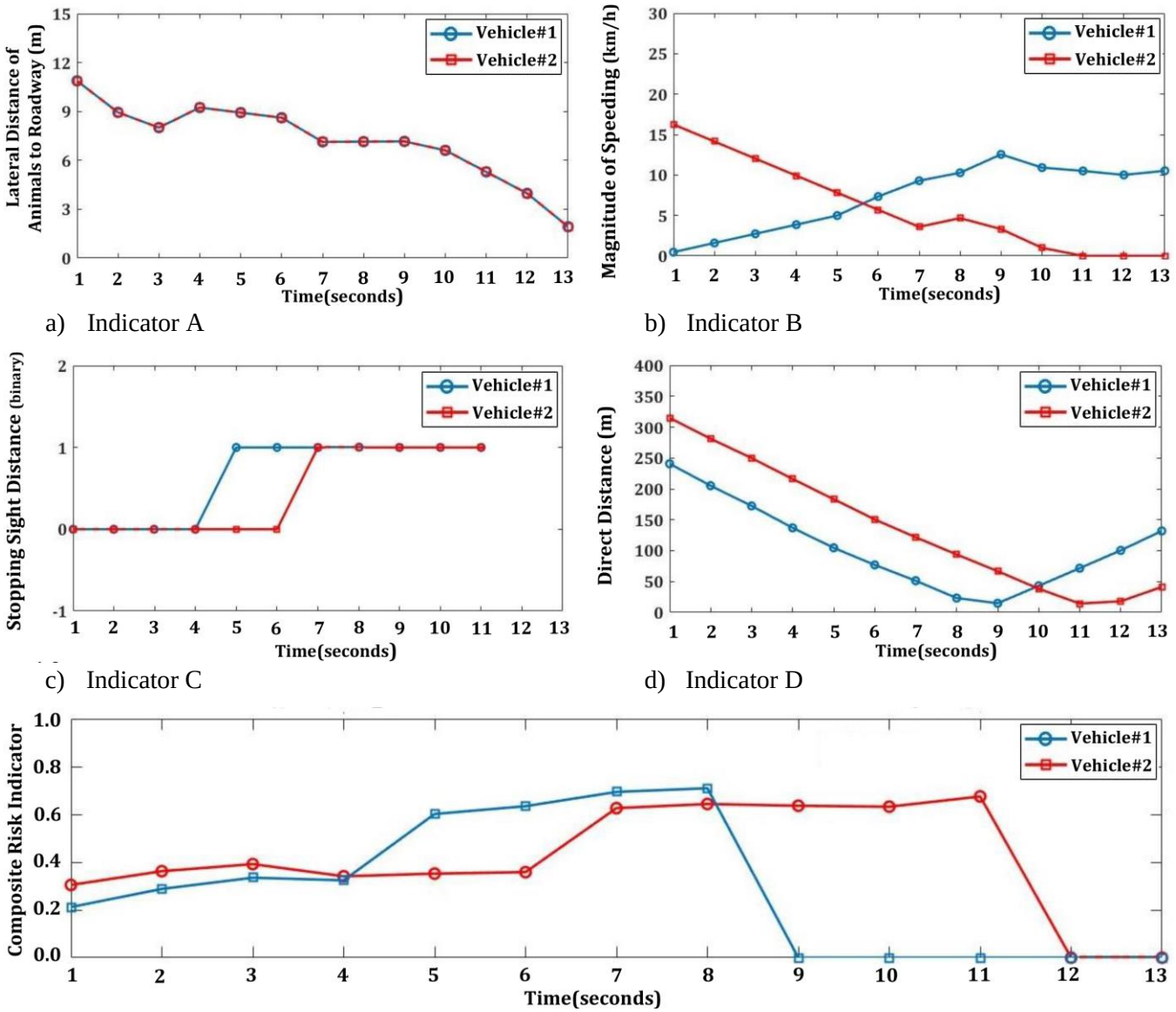


Figure 2-9: Changes in the four indicators and the estimated composite risk indicator during the 13 seconds of the event

5) Figure 2-9e shows the estimated risk for the two vehicles over the 13 seconds. Vehicle #1's composite risk values increased for the first 8 seconds largely because the vehicle was increasing speed. There was a relatively large increase in the value of composite risk indicator (from 0.34 to 0.53) from 4 seconds to 5 seconds. This increase occurred largely because Indicator C changed from 0 to 1 indicating

that vehicle #1 could not stop in time if the animal jumped onto the roadway. Fortunately, as the animal was still 7.5 m away from the roadway at vehicle #1's peak risk of 0.63 at 8 seconds, the value of composite risk indicator remained well under the maximum risk of 1. The risk for vehicle #2 were similar to those of vehicle #1 but peaked at 11 seconds rather than at 8 seconds because vehicle #2's was decreasing travel speed and because it was the following vehicle.

6) The value of composite risk indicator for vehicle #2 was higher than the value of composite risk indicator for vehicle #1 during the first four seconds. This might be unexpected as we normally expect a higher risk for the leading vehicle (vehicle #1) because it is closer to the animal than is the following vehicle (vehicle #2). In this case, however, vehicle #2 had a much higher magnitude of speeding than did vehicle #2 during the first four seconds (see Figure. 2-9b) resulting in a higher estimated risk for vehicle #2. We consider these results to be appropriate as they imply that the proposed NG RADS takes a conservative approach by providing an opportunity for the following vehicle(s) to receive the same or even higher levels of warning as the leading vehicle. This approach will help to increase the reaction or evasion time available to the following vehicle to avoid a collision with a large animal ahead even when a leading vehicle ahead blocks sight of the animal.

In summary, the proposed model used the multiple inputs to produce varying risk as intended. We think that varying risk provide a clear advantage over a conventional RADS that relies on a binary input and a binary output.

2.4.1. Sensitivity Analysis

Sensitivity analysis is a common approach used to show the adequacy of a new product or model, particularly when observed data are scarce as in this study (Laureshyn et al., 2017). In the context of road safety, Laureshyn et al. (2017) explained that sensitivity analysis is a procedure designed to investigate whether a measure can indicate the direction of change (improvement or deterioration) in safety, but not

the absolute extent of the change. For example, a sensitivity analysis could show whether the proposed system generates outputs (risk in this study) in the direction expected from changes in the input values (the four indicators in this study).

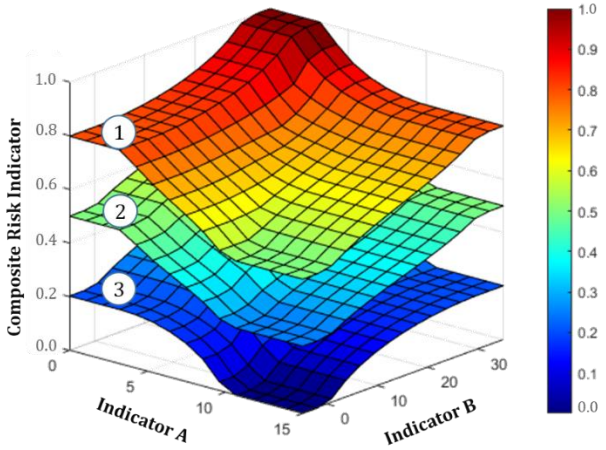
Figure. 2-10 shows the results of the sensitivity analysis, i.e., how the composite risk values changed as the input indicators changed. The changes are shown as 3-D surfaces. In Figure 2-10a, the values of Indicators A (distance from animal to roadway from 0 m to 15 m) and B (magnitude of speeding from 0 km to 30 km/h) change. In Figure 2-10b, the values of indicators C (SSD from 0 to 1) and D (direct distance from 0 m to 1,400 m) change. Three examples of each pair of indicators are provided as shown in the tables under the surfaces. In general, the direction of change in the value of composite risk indicator is what we expect.

For instance, the three surface layers in Figure 2-10a present the three different composite risk indicator surfaces associated with the different input values for Indicators C and D shown in the tabulation below the surfaces. Regardless of changes in Indicators C and D, the value of composite risk indicator increases when Indicator A decreases and Indicator B increases, but the range of the value of composite risk indicator depends on Indicators C and D. The value of composite risk indicator changes from 0 (the minimum value) to 1 (the maximum value). These results are as expected.

Figure 2-10b shows three surface layers that present the three different composite risk indicator surfaces associated with the different input values for Indicators A and B shown in the tabulation below the surfaces. When indicator D decreases, the value of composite risk indicator increases regardless of the values for Indicators D and C. Indicator C is binary and shows a sudden increase in the value of composite risk indicator when its value C changes from zero to one. This result is as expected.

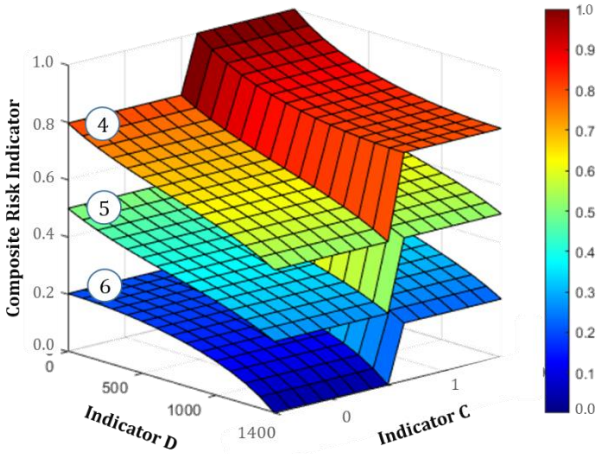
In general, the two lowest surfaces (surfaces 3 and 6) show combinations of indicator values that deliver a relatively low level of risk (highest risk value is 0.5), and the two highest surfaces (surfaces 1

and 4) represent combinations of indicator values that deliver a relatively high level of risk (minimum risk value is 0.5) that may require a strong and urgent level of warning for drivers.



	Indicator C	Indicator D	Composite Risk Indicator
Surface 1	1	0	(0.50,1.00)
Surface 2	0	0	(0.25,0.75)
Surface 3	0	1400	(0.00,0.50)

a) Estimated composite risk indicator when Indicators A and B change



	Indicator A	Indicator B	Composite Risk Indicator
Surface 4	0	30	(0.50,1.00)
Surface 5	0	0	(0.25,0.75)
Surface 6	15	0	(0.00,0.50)

b) Estimated composite risk indicator when Indicators C and D change

Figure 2-10: Sensitivity analysis results

2.5. Discussion and Recommendations

2.5.1. Potential Benefits of NG RADS over Conventional RADS

As noted, this study can be viewed as a feasibility (proof-of-concept) study for developing a framework (algorithm) for an advanced NG RADS that can be integrated with V2I technology. The proposed NG RADS can produce a varying output (different levels of risk) and a varying output (different levels of warning).

Figure 2-11 shows the potential benefits of our proposed system compared with a conventional RADS. In the Figure, imagine a situation when a vehicle is traveling a section of highway during the time from time *i* to time *n*. The vehicle passes an animal near the roadway at time *k* (the nearest point of conflict).

As shown in Figure 2-11**,** a conventional RADS (the horizontal brown line) generates a fixed and constant level of warning (the brown line) over the entire duration of warning. The warning begins as soon as the system detects the presence of an animal within a pre-set area and the same level of warning continues for a certain period of time which may even include a period after the disappearance of an animal. The conventional RADS does not consider the intensity or duration of potential interactions between an animal and a vehicle.

The proposed NG RADS, however, can disseminate different levels of warning depending on the composite risk indicator generated for a particular circumstance. Figure 2-11 provides examples of two alternative approaches to the warnings disseminated by the NG RADS, #01 (the purple curve) and #02 (the orange curve).

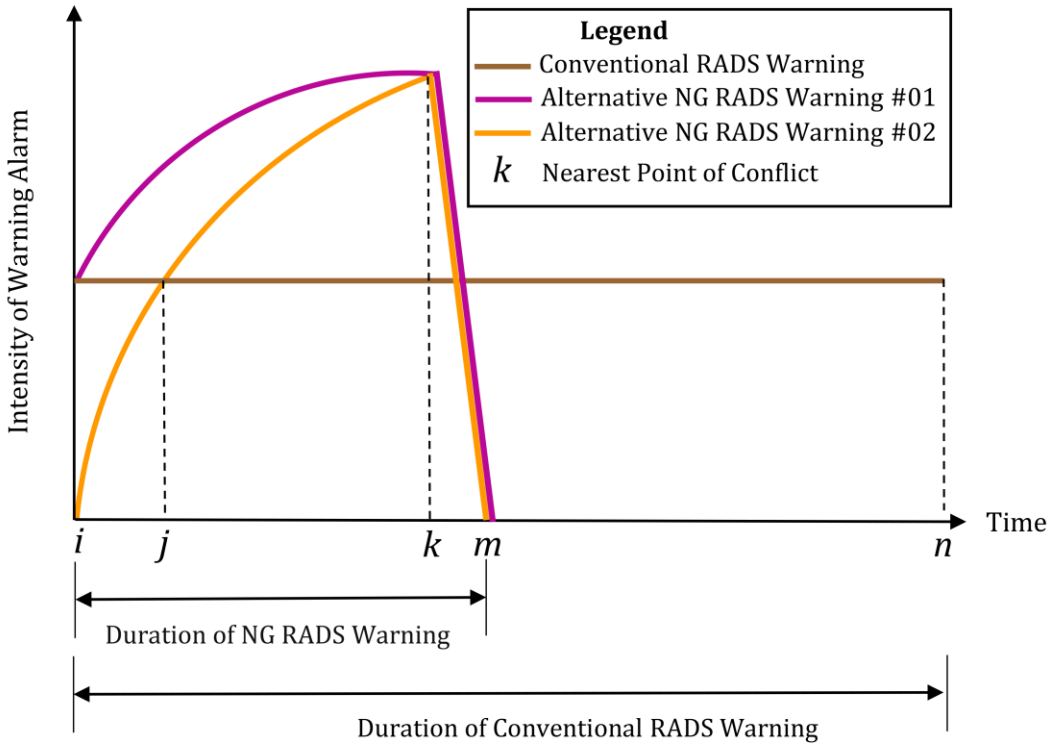


Figure 2-11: Benefits of NG RADS versus the conventional RADS

In warning #01, the NG RADS detects the presence of an animal at time *i* and activates a medium

level of warning that then increases in intensity until the vehicle passes the nearest point of conflict with the animal (at time *k*) after which the vehicle is no longer under risk and the warning level quickly drops to zero (at time m). This approach has two important benefits over the conventional RADS: 1) relevantly timed warnings, i.e., a driver will receive a warning message only while necessary and will not be distracted and/or stressed by nuisance warning messages after the risk has passed; and 2) a stronger level of warning for emergency situations that require a driver's immediate attention, e.g., animals on roadway rather than roadside.

In warning #02, the NG RADS also detects the presence of an animal at time *i* and activates a warning that takes the risk level at time *I* into account. In the example, the risk is low between time *i* and time *j* so the level of warning is not intense. If the value of composite risk indicator increases, the intensity of the warning increases as shown from *j* to *k*. Approach #02 is the favoured approach because it begins by disseminating relatively less intense warning messages and only disseminates stronger warning messages after time *j* if necessary and therefore minimizes distracting and/or stressing drivers with nuisance warnings.

Although the proposed NG RADS can estimate the varying levels of risk second by second, the system should not automatically disseminate a stream of varying levels of warning message to drivers. Providing the most appropriate level of warning the most appropriate time in response to the estimated level of risk is expected to be a challenging issue that should be fully investigated in Phase III (communication). Too many different risk levels may be confusing and difficult for drivers to interpret and react to appropriately especially when time is limited. The system may decide to limit the number of warning levels to three (low, medium, and high).

Although the advanced V2I communication technology envisaged for investigation in Phase III (communication) is not a focus of this study, the auto industry is rapidly developing connected

autonomous vehicles (CAVs) which could deliver even greater benefits from the NG RADS. Future NG RADS could automatically and continuously trigger changes in the traveling speed of CAVs and could even completely stop one or more vehicles if the risk level is high.

It is also possible to imagine different ways for the NG RADS to disseminate varying levels of warning for human drivers. Firstly, it may be useful to consider additional warning lights and/or sound, for example, warning lights that can display multiple colors or LED lights that flash frequently with/without an extra sound warning similar to the warning sound at a grade-crossing. If the level of risk increases, light and/or sound can become more intense. Secondly, some near-future vehicles (still with human drivers) may be able to communicate directly with the NG RADS in which case many means of disseminating different levels of warning for human drivers could become available (e.g., vibrating steering wheels, warning messages on the vehicle's built-in display system, and/or alarm sounds broadcast through the vehicle's built-in audio system). Thirdly, multiple NG RADS could cover many kilometers of highway thereby detecting the presence and movement of large animals along much longer sections of highway than can the in-vehicle detection systems (e.g., radar or LIDAR) currently available on individual autonomous vehicles.

Although this study has highlighted the potential benefits of the NG RADS over a conventional RADS, decision makers need to know the effectiveness of the proposed system, i.e., the expected percentage reduction in the number of collisions (see US Highway Safety Manual, 2010). Unfortunately, the effectiveness of the NG RADS can only be estimated with a scientific before and after study of the change in the number and severity of collisions recorded at a site and such a study can only be conducted after an appropriate period of operation in the field. Uncertainty is a challenge when deciding to install any new system (including the proposed NG RADS).

2.5.2. Limitations of Current Study and Recommendations for Future Work

It is important to note the study's limitations and opportunities for future improvement. In terms of limitations, Indicator C (see Equation (2-2)), for example, includes the input parameters perception-reaction time and friction coefficient. This study used a fixed value (taken from the AASHTO, 2018) for each one, but both are hard to observe in real-time in the field and in reality, both vary considerably. For example, perception-reaction times vary from one driver to another, and surface conditions associated with a roadway's friction coefficient change over time depending, on changing weather conditions, etc. Future work could include a sensitivity analysis to explore the impact of varying input parameter values for perception-reaction times and friction coefficients. Future work will also need to search for an appropriate wireless technology that provides realistic values for input parameters. The technology would, for example, need to connect to the closest weather station to collect data such as real-time rainfall to update the friction coefficient as required.

Future work needs to address many issues associated with Phase III (communication). For instance, future work needs to develop a process that can properly determine a range of thresholds and the appropriate level of warning including the threshold for triggering an intense or emergency level of warning. If the value for a threshold is too low, too many events may trigger a warning message that is stronger than needed causing unnecessary distraction/stress for drivers. If, on the other hand, the value for a threshold is too high, the system may not alert drivers to situations that require a driver to react.

Future work also needs to screen and select the most effective V2I technologies for disseminating varying levels of warning message. As outlined above, we can use a visual, audible or tactile type of communication or a combination of these. Future work needs to consider environmental factors (e.g., adverse weather conditions) that may affect the behavior and choices of an animal or driver. Such a study would probably require extensive and possibly costly data collection.

As the scope of this study did not include searching for the best mathematical approach to risk analysis, future work may need to consider improving the fuzzy rule-based algorithm applied in this study. It might be useful, for example, to investigate a potentially more appropriate type of membership function (e.g., a Gaussian function rather than the linear function used in this study). Future study could also test various other algorithms (e.g., machine learning, Bayesian inference, etc.) to compare the accuracy and efficiency of the approaches in generating more accurate composite risk indicator. Such tests could be costly as they may require collecting substantial amounts of detailed field data over a considerable period of time to calibrate the algorithms for various different situations.

Lastly, future work needs to design the most affordable NG RADS. As many collisions involving large animals occur in remote areas, connecting the NG RADS to the nearest electrical grid may require substantial construction and/or installation costs. Where feasible, solar power may be worth considering, but this approach also has costs such as the cost of the solar panels, installation and maintenance (e.g., regular solar power battery replacement).

Although careful further analysis is required, we trust that this study has provided good seed information for the eventual development of a NG RADS to increase safety on rural roads.

2.6. Chapter 2 References

American Association of State Highway Transportation Officials (AASHTO). (2010). Highway Safety Manual, Washington, DC. Federal Highway Administration.

American Association of State Highway and Transportation Officials (AASHTO). (2018). “A Policy on Geometric Design of Highway and Streets”. Washington, DC. Federal Highway Administration.

Antonio, W. H., Da Silva, M., Miani, R. S., and Souza, J. R. (2019). A proposal of an animal detection system using machine learning. Applied Artificial Intelligence, 33 (13), 1093-1106.

Balmat, J. F., Lafont, F., Maifret, R., and Pessel, N. (2011). A decision-making system to maritime risk assessment. Ocean Engineering, 38(1), 171-176.

Barua, A., Mudunuri, L.S., and Kosheleva, O. X. (1995). Why trapezoidal and triangular membership functions work so well. J. Uncertain Syst. 8(3), 164–168.

Bond, A. R., and Jones, D. N. (2013). Wildlife warning signs: public assessment of components, placement and designs to optimise driver response. Animals. 3(4), 1142-1161.

British Columbia Ministry of Transportation and Infrastructure (BC MoTI) (2019), Memorandum, BC Supplement to TAC Geometric Design Guide. 3rd ed., British Columbia Ministry of Transportation and Infrastructure, British Columbia

Cannon, W.B. (1916). Bodily changes in pain, hunger, fear, and rage: an account of recent research into the function of emotional excitement. D. Appleton, New York

Castignani, G., Derrmann, T., Frank, R., and Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. IEEE Intelligent Transportation Systems Magazine, 7(1),91-102.

Christiansen, P. E. R. (2002). Locomotion in terrestrial mammals: the influence of body mass, limb length and bone proportions on speed. Zoological Journal of the Linnean Society, 136(4), 685-714.

Danks, Z. D., & Porter, W. F. (2010). Temporal, spatial, and landscape habitat characteristics of moose—vehicle collisions in western Maine. The Journal of Wildlife Management, 74(6), 1229-1241.

Ding, W., Jing, X., Yan, Z., and Yang, L.T. (2019). A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion. Inf. Fusion 51, 129–144.

Druta, C., Alden, A.S., and Donaldson, B.M. (2018). Evaluation of a buried sensing cable for roadside animal detection, No. 18–00437, Transportation Research Board 97th Annual Meeting, Washington DC, United States

Druta, C., and Alden, A. S. (2020). Preventing animal-vehicle crashes using a smart detection technology and warning system. Transportation Research Record, 2674(10), 680-689.

Eftekhari, H. R., and Ghatee, M. (2018). Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition. Transportation Research Part F: Traffic Psychology and Behavior, 58, 782-796.

Fitzpatrick, K., Carlson, P., Brewer, M. A., Wooldridge, M. D., and Miaou, S. P. (2003). NCHRP report 504: Design speed, operating speed, and posted speed practices. Transportation Research Board of the National Academies, Washington, DC.

Frid, A., and Dill, L. (2002). Human-caused disturbance stimuli as a form of predation risk. Conservation Ecology, 6(1), 11.

Grace, M. K., Smith, D. J., and Noss, R. F. (2017). Reducing the threat of wildlife-vehicle collisions during peak tourism periods using a roadside animal detection system. Accident Analysis & Prevention, 109, 55-61.

Huijser, M. P., McGowen, P. T., and Camel, W. (2006). Animal vehicle crash mitigation using advanced technology phase I: review, design, and implementation (No. FHWA-OR-TPF-07-01). Available at: https://rosap.ntl.bts.gov/view/dot/21863/dot_21863_DS1.pdf [Accessed June, 2022]

Huijser, M. P., McGowen, P. T., Fuller, J., Hardy, A., and Kociolek, A. (2007). Wildlife-vehicle collision reduction study: Report to congress. Available at: https://wafwa.org/wp-content/uploads/2021/04/2007-Report-to-Congress.pdf [Accessed June, 2022]

Huijser, M.P., McGowen, P., Clevenger, A.P., and R. Ament, (2008). Wildlife–vehicle collision reduction study: best practices manual, FHWA-HEP-09-022, U.S. Federal Highway Administration.

Johansson, F., and Falkman, G. A. (2008). Bayesian network approach to threat evaluation with application to an air defense scenario. 11th International Conference on Information Fusion, (pp. 1-7). IEEE.

Johnsson, C., Laureshyn, A., and De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. Transport Reviews, 38(6), 765-785.

Klein, L.A., and Klein, L.A. (2004). Sensor and data fusion: A tool for information assessment and decision making. Vol. 324. SPIE, Bellingham.

Laureshyn, A., Johnsson, C., Madsen, T., Varhelyi, A., de Goede, M., Svensson, A., and van Haperen, W. (2017). "Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users". In the Road Safety & Simulation International Conference, The Hague, NL.

Lima, S. L., Blackwell, B. F., DeVault, T. L., and Fernández-Juricic, E. (2015). Animal reactions to oncoming vehicles: a conceptual review. Biological Reviews, 90(1), 60-76.

Mahmud, S. S., Ferreira, L., Hoque, M. S., and Tavassoli, A. (2017). Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs. IATSS research, 41(4),153-163.

Meisingset, E. L., Loe, L. E., Brekkum, Ø., & Mysterud, A. (2014). Targeting mitigation efforts: The role

of speed limit and road edge clearance for deer–vehicle collisions. The Journal of Wildlife Management, 78(4), 679-688.

Mendel, J. M. (1995). Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE, 83(3), 345-377.

Michigan State Police, (2020). Crash Database. Available at: https://www.michigan.gov/documents/msp/2019Year_End_File_for_Web_691956_7.pdf [Accessed June, 2022]

Mukherjee, A., Stolpner, S., Liu, X., Vrenozaj, U., Fei, C., and Sinha, A. (2013). Large animal detection and continuous traffic monitoring on highways. Sensors, Baltimore, MD, pp. 1–3.

National Insurance Crime Bureau (2018), Animal-related insurance claims top 1.7 million in four years, Available at: https://www.nicb.org/news/news-releases/animal-related-insurance-claims-top-17-million-four-years [Accessed June, 2022]

National Highway Traffic Safety Administration (NHSTA) (2018), Traffic safety facts 2017, Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812806 [Accessed June, 2022]

National Highway Traffic Safety Administration (NHSTA) (2020), Traffic safety facts 2020, Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813320 [Accessed Oct, 2022]

Ontario provincial police, (2018), Annual report, Available at: https://www.opp.ca/tms/entrydata.php?fnc=3&_id=5d30cd4b241f6e3c9e566f63 [Accessed June, 2022]

Phillis, Y. A., and Davis, B. J. (2009). Assessment of corporate sustainability via fuzzy logic. Journal of Intelligent and Robotic Systems, *55*(1), 3-20.

Rahman, K. A., and Tepe, K. E. (2014). Towards a cross-layer based MAC for smooth V2V and V2I communications for safety applications in DSRC/WAVE based systems. In 2014 IEEE Intelligent Vehicles Symposium Proceedings. Dearborn, MI, pp. 969–973.

Rogers, E., and Premo, D.B. (2004). An ecological landscape study of deer-vehicle collisions in Kent county. White Water Associates. Transportation Research Record, Amasa, MI

Sharafsaleh, M., Huijser, M., Kuhn, T., Spring, J., and Felder, J. (2010). Evaluation of an animal warning system effectiveness, No. UCB-ITS-PRR2010-22, California PATH Program, Institute of Transportation Studies, Available at: https://merritt.cdlib.org/d/ark%3A%2F13030%2Fm5jm2c9w/2/producer%2FPRR-2010-22.pdf [Accessed June, 2022].

Srinivasan, D., Sanyal, S., and Sharma, V. (2007). Freeway incident detection using hybrid fuzzy neural network. IET Intelligent Transport Systems, 1(4), 249-259.

Shrinivasan, L., and Raol, J. R. (2018). Interval type-2 fuzzy-logic-based decision fusion system for air-lane monitoring. IET Intelligent Transport Systems, 12(8), 860-867.

Teodorović, D. (1994). Fuzzy sets theory applications in traffic and transportation. European Journal of Operational Research. 74(3), 379-390.

The Wildlife Collision Prevention Program (2019). Wildlife Vehicle Collision (WVC) Facts. Available at: https://www.wildlifecollisions.ca/collision/collision-facts.htm. [Accessed Sep. 2022].

Traffic Injury Research Foundation (2017). Road safety monitor 2016: Driver behavior and wildlife on the road in Canada. Available at: http://tirf.ca/wp-content/uploads/2018/09/RSM-2016-Driver-Behaviour-and-Wildlife-on-the-Road-in-Canada-5.pdf [Accessed June, 2022]

Transportation Association Canada (TAC), Geometric Design Guide (2019). Available at: https://www2.gov.bc.ca/assets/gov/driving-and-transportation/transportation-infrastructure/engineering-standards-and-guidelines/highway-design-and-survey/tac/tac-2019-supplement/bctac2019-chapter-0000-front.pdf [Accessed June, 2022]

Vanlaar, W. G., Barrett, H., Hing, M. M., Brown, S. W., and Robertson, R. D. (2019). Canadian

wildlife-vehicle collisions: an examination of knowledge and behavior for collision prevention. Journal of Safety Research. 68, 181-186.

Zadeh, L. A. (1965). Fuzzy sets. Information and control. 8(3), 338-353.

CHAPTER THREE

3. DEVELOP A FUZZY RULE-BASED ALGORITHM FOR A SMART PROTECTION SYSTEM TO PREVENT COLLISIONS WITH POLICE OFFICERS ON DUTY ON THE ROADWAY

Abstract: Police officers on duty on the road for traffic stops, vehicle collisions, traffic direction, etc. are exposed to the risk of being hit or even killed by a passing vehicle. Very few studies have tried to develop a collision avoidance system (CAS) that can warn pedestrians or police officers on duty on the road to take proactive evasive action. This study proposes a smart protection system for police officers on duty on the road. The development of the system envisaged involves three essential phases: I) detection, II) risk analysis, and III) warning and communication. This study has focused on the risk analysis phase. We developed a fuzzy rule-based algorithm that combines four input indicators (lateral distance from police officer to traveled lane, magnitude of speeding, stopping sight distance, and direct distance) into a single composite risk indicator. The study used data from a real-world situation on Highway 416 in Ontario, Canada to demonstrate the application of the proposed algorithm. The results clearly demonstrate that the proposed algorithm could identify near miss collisions with 78% success rate and could be used to give timely warning of a possible collision risk to police officers at work on a road.

Keywords: *Smart protection system, Fuzzy rule-based algorithm, Collision avoidance system, Police officers, Risk analysis algorithm, Pedestrian-to-vehicle collisions*

3.1. Introduction

Police officers have a number of duties such as directing traffic and pedestrians, stopping traffic for speeding, assisting/investigating severe collisions, etc. that require officers to be on the road and exposed to the risk of being hit by a passing vehicle. Between 2003 and 2013, 138 police officers were struck and killed on US highways (NHTSA, 2013). Data for 2015 to 2019 showed that 50 police officer deaths resulted from being struck by a vehicle. During this five-year period, traffic stops accounted for 24% of the deaths and assisting/investigating vehicle collisions for 22% (FBI, 2020).

Although a lot of research has considered the road safety of the general public (Moudon et al., 2008; Hassein et al., 2018; Ka et al., 2019), but not much research has been paid attention about reducing the number of collisions involving police officers on duty. This study proposes a new mathematical algorithm for developing a real-time smart protection system to protect police officers who are on duty on the road. The smart protection system refers to a variety of wireless connected devices such as sensors, detectors, and computers that can be used to provide field officers with real-time situational awareness and early warning of a possible collision.

The adoption of smart technologies in road safety is creating new ways to improve situational awareness for people using the roads. Examples are built-in vehicle devices, such as an automatic pedestrian detection system, that use advanced sensing technologies (LiDAR or radar) (Tong et al., 2018). Another example is a prototype system called “WalkSafe” which aims to protect distracted pedestrians by alerting pedestrians who are using a smartphone. The system relies on phone cameras for detecting vehicles coming towards a pedestrian (Wang et al., 2012). Foerster et al. (2014) proposed a similar pedestrian protection system called “SpareEye”. This system also uses smartphone cameras. The system alerts the pedestrian to vehicles along the pedestrian’s walking path and alarms the pedestrian. Lin et al. (2017) developed a smart system called “pSafety” which uses a smartphone’s GPS data to estimate the

position of a pedestrian and to provide warning of a potential collision with an approaching vehicle 3 to 4 seconds before the collision might occur. Tong and Jia (2019) developed augmented reality glasses (i.e., a wearable device) which activate a warning when there is a risk of a collision. The screen of the virtual reality glasses displays the remaining time to a collision and shows the expected path of the approaching vehicle.

Some studies have considered the safety of construction zone workers by investigating the possibility that wearable devices known as smart vests could reduce collision risk with approaching vehicles. The researchers embedded sensors into traffic cones to detect and analyze approaching risk and the sensors could then alarm construction zone workers by light emitting diode lights or by vibration in the case of a high-risk situation. The study technology warned the construction zone workers of a collision risk as soon as a single cone was hit by an approaching vehicle (Dayal et al., 2013; Roofigari-Esfahan et al., 2021).

In a smart protection system, a high-risk situation is generally estimated using surrogate safety indicators (Mullakkal-Babu et al., 2017). Surrogate safety indicators are performance measures used to assess the level of risk/safety in real-time. Many surrogate road safety indicators have been introduced (time to collision, post encroachment time, time headway, etc.). Mahmud et al. (2017) and Johnsson et al. (2018) provided comprehensive reviews of many surrogate safety indicators. The indicators can capture a portion of the risk associated with certain types of collision. For instance, time to collision and time headway are useful indicators if capturing the risk associated with rear-end collisions whereas post encroachment time better captures the risk associated with angle collisions at intersections. Currently, no single indicator can successfully capture all types of collisions on roadways.

The use of multiple indicators to assess collision risk requires a complex and sophisticated algorithm that can generate a composite risk indicator by combining different input indicators. An

integrated system that can produce a single composite indicator is obviously attractive for decision makers needing to set target goals and priorities. Arun et al. (2021) regarded the integration of outcomes from several surrogate indicators into a single composite indicator that estimates the risk associated with collisions as a promising approach, but their study focused on vehicle-to-vehicle collisions. No research has considered a composite indicator that could alert police officers to the risk of being hit by a vehicle.

Teodorović (1994) noted that a collision avoidance system (CAS) in intelligent transportation systems often need to deal with traffic flow situations where data are very uncertain or scarce to collect. As a result, the choice of an appropriate method for a study that aims to develop a new product often depends largely on the nature of the available data and missing observed data are often a major constraint. Many advanced algorithms (e.g., deep learning) require a fairly large volume of input data to calibrate the system and some data are inevitably uncertain. In this study, it is necessary to consider actual and/or near miss collisions between vehicles and police officers, but as these collisions occur only rarely, it is not possible to observe and record a large amount of relevant data for system calibration.

This study used a fuzzy rule-based algorithm to produce a single composite risk indicator. Zadeh (1965) proposed fuzzy rule-based algorithms as a tool for capturing and analyzing uncertainty without requiring a large amount of input data. A fuzzy rule-based algorithm uses a membership function to determine a membership degree between 0 and 1 for each input indicator. The membership function can be changed into a “fuzzy set” designed to capture the uncertainty of the input and output indicators. These fuzzy sets can be used to evaluate binary situation such as yes/no, true/false, etc. and defined as linguistic variables (e.g., safer or riskier). Fuzzy rule-based algorithms have been widely used for many different transportation safety studies (Xiao et al., 2006; Lee et al., 2006; Li et al., 2012; Liu et al., 2017).

3.1.2. Study Goal and Objectives

The goal of this study is to propose a smart protection system for pedestrian who works on roadway (police officer in our case) which relies on designing a risk analysis algorithm (fuzzy rule-based algorithm) to create a composite risk indicator. The study objectives are:

1) Identify appropriate input indicators of different risks associated with different risks associated with police officer' diverse activities, and traffic situations;
2) Apply a risk analysis algorithm that can combine outputs from multiple indicators, estimate varying levels of risk, generate a single composite risk indicator, and provide an opportunity for timely warning to the police officer; and
3) Validate the use of the algorithm in a case study involving field police officers on duty on the road.

3.2. Proposed Smart Protection System

Figure 3-1 shows the three phases of the proposed smart system: I) detection; II) risk analysis, and III) warning and communication.

I. Phase I (detection): A 360° scanning radar sensor can be installed, for instance, on top of a police vehicle. This sensor can detect an object within the line of sight up to a maximum radius of 700 meters (Mukherjee et al., 2013). The outputs of the detection phase are the positions (i.e., x and y coordinates), trajectories and speeds of passing vehicles and police officers in real-time (e.g., second by second).

II. Phase II (risk analysis): The system uses the seed information from Phase I to develop multiple input indicators to capture police officer behavior, driver behavior, and the spatial and temporal interaction between the police officer and passing vehicle. The input indicators could include the

lateral distance from officers to traveled lane, the magnitude of speeding, the stopping sight distance, and the direct distance. The input indicators are similar to input indicators used in Chapter 2; however, the calibration (i.e., boundaries of membership functions) of risk analysis algorithm is different. A risk analysis algorithm using a sophisticated risk analysis algorithm is then developed to combine input indicator and to estimate a composite risk indicator. The composite risk indicator ranges from zero (lowest risk) to one (highest risk). The risk indicator is the outcome of the risk analysis phase and is used as an input to generate the appropriate level of warning in Phase III.

III. Phase III (warning and communication): If the risk generated in Phase II exceeds various pre-set thresholds, the system can generate the appropriate level of audio and/or vibration warning to the police officer. The warning can be disseminated to the police officer's equipment (e.g., mobile radio) or wearable device (e.g., wrist band) and the police officer can take evasive action.

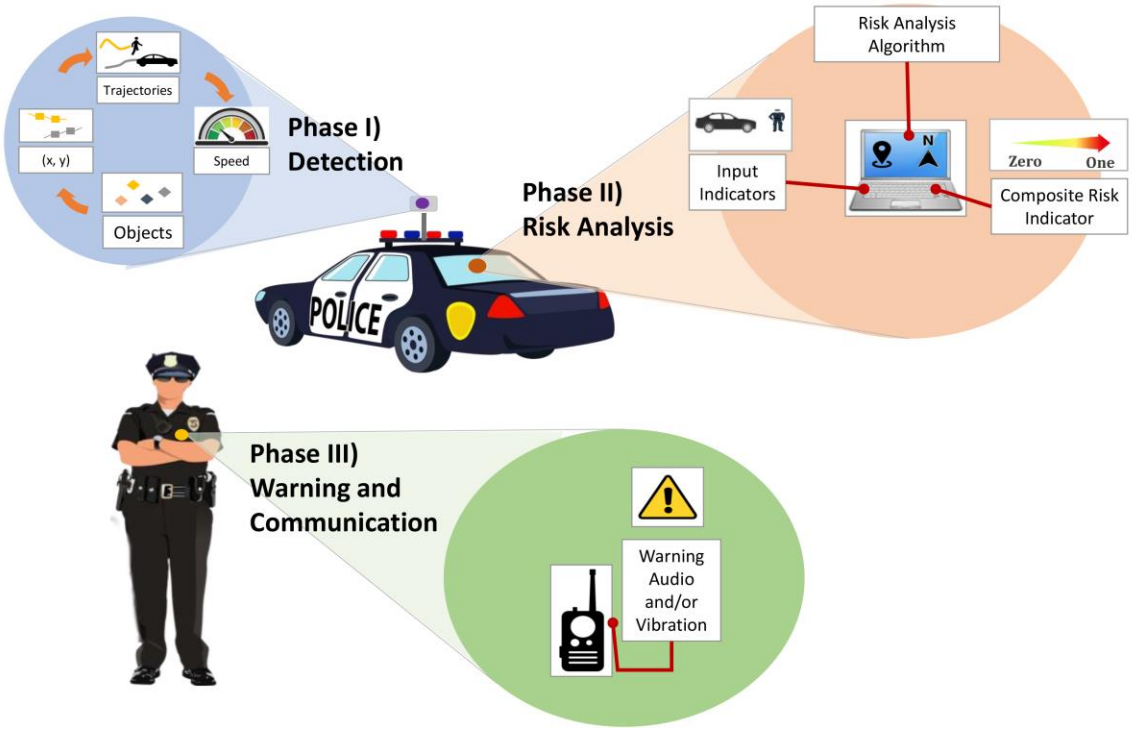


Figure 3-1: Concept of the proposed smart protection system

Although we aim to develop a single-unit machine that can perform all three phases of tasks real-time and automatically, this study focuses primarily on Phase II (risk analysis). As a conventional radar system can detect input information such as the second by second position of police officers and vehicles, the technology needed for Phase I already exists and this study adopted a radar system similar to the system applied by Mukherjee et al., (2013). Phase III (warning and communication) will be investigated in future work.

3.3. Research Approach

The risk analysis phase comprises of five distinct sub-tasks:

1) **Select input indicators** - determine appropriate input indicators for capturing the possible collision risks associated with a police officer's field activities and traffic environment;

2) **Apply a risk analysis algorithm -** select and use an appropriate risk analysis algorithm (a fuzzy rule-based algorithm in this study) for combining the input indicators into a single composite indicator;

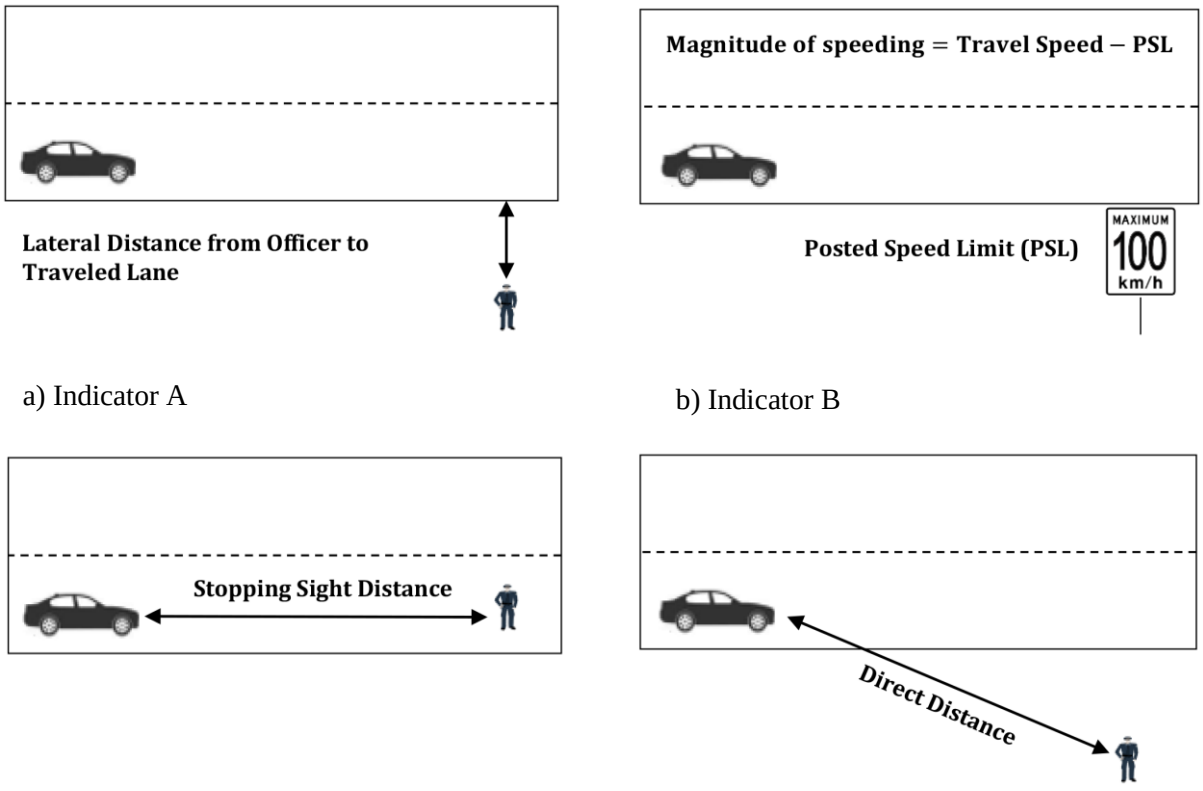
3) **Calibrate the algorithm** - collect real-world data and use the data on input indicators to calibrate the risk analysis algorithm to mimic the real-world;

4) **Estimate the composite risk indicator** - apply the calibrated algorithm to a scenario that reflects a typical situation for police officers to illustrate how the algorithm can be used to estimate the risk associated with the scenario; and

5) **Validate the algorithm** - compare the accuracy estimated by the proposed algorithm with the past study to demonstrate how well the algorithm estimates the risk associated with a police officer's field activities and traffic environment.

3.3.1. Task 1: Select Input Indicators

To select appropriate input indicators, the study investigated several collisions in which first responders including police officers were struck by a vehicle (U.S. Fire Administration, 2008; U.S. Fire Administration, 2012). We also thoroughly reviewed surrogate road safety indicators used in past studies (Mahmud, et al., 2017; Johnsson et al., 2018). We then selected and/or developed the four indicators shown in Figure 3-2. The four indicators are: 1) Indicator A (lateral distance from officer to traveled lane), 2) Indicator B (magnitude of speeding), 3) Indicator C (stopping sight distance (SSD)), and 4) Indicator D (direct distance).



a) Indicator A

b) Indicator B

c) Indicator C

d) Indicator D

Figure 3-2: The four input indicators

1) Indicator A (lateral distance from officer to traveled lane): When a pedestrian (a police officer in our case) is working on a roadway, there is a possibility of being hit by a passing vehicle.

Transportation engineers try to provide a safe lateral distance (minimum 0.6 m) between pedestrians walking along the side of a roadway and the traveled lane (US Department of Transportation, 2020). Figure 3-2a illustrates the concept of Indicator A. As the lateral distance from the officer to the traveled lane decreases, the situation's risk increases.

2) Indicator B (magnitude of speeding): Aggressive driving (e.g., speeding, tailgating, running a red light, etc.) is known to be a key factor in collisions, and speeding (i.e., excessive speed) is known to be the most important aggressive driving behavior. Many North American studies have argued that reducing speeding is one of the most promising ways to reduce any type of collision (Hauer 2009; Schroeder et al., 2013; Castignani et al., 2015; Eftekhari and Ghatee 2018; NHSTA 2021) and it appears very reasonable to assume that aggressive driving will contribute to increasing the probability of a collision with a police officer on a roadway (Castignani et al., 2015). This study used the magnitude of speeding, i.e., the speed differential between a vehicle's travel speed and the posted speed limit, as a primary measure to assess an approaching vehicle's aggressiveness. Equation (3-1) and Figure 3-2b explain the concept of indictor B. As the magnitude of speeding increases, the risk increases.

$$MS_i = TS_i - PSL \tag{3-1}$$

where,

MS_i = magnitude of speeding for vehicle *i* (km/h);

TS_i = travel speed for vehicle *i* (km/h); and

PSL = posted speed limit on the segment considered (km/h).

Note that Indicators A and B do not assess any spatial interaction between an officer and a passing vehicle. These interactions can be assessed using the following two indicators.

3) Indicator C (stopping sight distance (SSD)): SSD is an estimate of the minimum distance at which a driver notices an object (in our case an officer) on the roadway and can make a complete stop to

avoid the collision with the object (AASHTO, 2018; TAC, 2019). See Figure 3-2c. SSD is an important highway design parameter that considers a variety of safety elements including perception-reaction time, speed and friction coefficient (see Equation (3-2)). In this study, the SSD indicator is presented as a binary indicator. If the distance between the vehicle and the nearest point of conflict is less than the SSD, the situation is considered risky. If the distance is longer than the SSD, the situation is considered safe.

$$SSD = Vt_{pr} + \frac{V^2}{2\mu(f)} \quad (3\text{-}2)$$

where,

SSD = stopping sight distance (m);

V = vehicle speed (m/s);

t_{pr} = perception-reaction time (s);

μ = 9.8 m/s^2; and

f = friction coefficient.

4) Indicator D (direct distance): In certain situations, indicators A, B, and C cannot capture the potential risk and will not trigger a warning that a vehicle is approaching. For instance, if a police officer is located outside the threshold lateral distance from the roadway, an approaching vehicle is traveling at less than the posted speed limit, and the vehicle's estimated SSD for appears adequate, the system may not trigger a warning. However, there could be a rare but real circumstance when a police officer may suddenly move into the roadway without noticing an approaching vehicle. To take such a circumstance into account, this study used the direct distance between a vehicle and a police officer as an input indicator (see Figure 3-2d). As the direct distance decreases, the situation's risk increases.

Although this study did not use the traveling speed of vehicles as a separate and independent indicator, both Indicators C (SSD) and D (direct distance) indirectly capture the risk associated with an

approaching vehicle's travel speed. Equation (3-2) shows, for instance, that as a vehicle's approaching speed increases, the SSD required for a complete stop increases. It is clear that higher speeds increase the level of risk. In addition, as the proposed system analyzes the situation second by second, Indicator D (direct distance) also takes an approaching vehicle's traveling speed into account. As an approaching vehicle's speed increases, the direct distance rapidly decreases, and the estimated level of risk rapidly increases. In summary, we think it is redundant (double dipping) to consider approaching vehicles' travel speed as an independent indicator in our proposed algorithm.

3.3.2. Task 2: Apply a Risk Analysis Algorithm

Figure 3-3 shows a diagram of fuzzy rule-based algorithm with two main modules: 1) the knowledge base, and 2) the fuzzy inference unit (Mirabbasi et al., 2008). The knowledge base is the information that a modeler needs to provide for a fuzzy rule-based algorithm. This information can be collected using past studies and/or experts' opinions. The fuzzy inference unit refers to the consecutive processes which the algorithm follows to generate output. The fuzzy inference unit has three main stages: 1) fuzzification, 2) rule evaluation, and 3) defuzzification (Mendel, 1995). In the fuzzification stage, the input indicators (in this study, the four input indicators shown in Figure (3-2) are converted into input fuzzy sets using input membership functions. The number and shape of the membership functions for each input indicator are defined in the knowledge base module. In the rule evaluation stage, the four input indicators in the form of fuzzy input sets are combined using "IF-THEN" rule constructions. The output of this stage is fuzzy output sets. The IF-THEN rules should be defined in the knowledge base module. For example, a rule could be defined as: if lateral distance from officer to traveled lane (Indicator," and A) is "safer magnitude of speeding (Indictor B) is "safer," and stopping sight distance (Indicator C) is "safer," and direct distance (Indictor D) is "safer," then output fuzzy set is "very low."

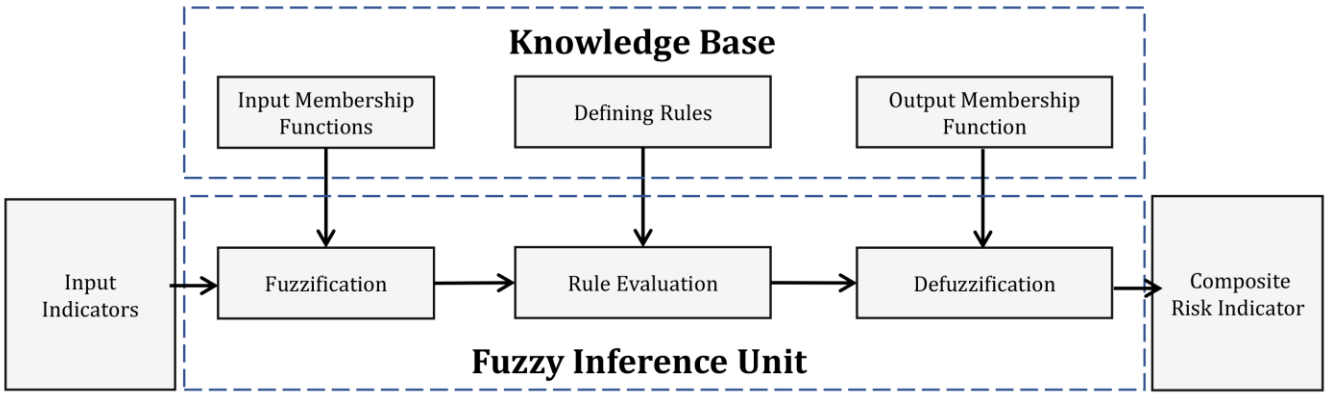


Figure 3-3: Fuzzy rule-based algorithm

In the defuzzification stage, the fuzzy output sets are converted to estimated composite risk indicator using output membership functions. The number and shape of the membership functions for composite risk indicator are defined in the knowledge base module. A more detailed description of a fuzzy rule-based algorithm is provided by Mendel (1995).

3.3.3. Task 3: Calibrate the Algorithm

As the proposed smart protection system has not yet been implemented in a real-world circumstance, the study collected study data (e.g., second by second x- and y-coordinates and speed data of pedestrians, traffic, etc.) from a radar system installed on a section of Highway 416 near Kemptville, Ontario. The radar covers a 1.4 km of section of a rural highway with four lanes (two lanes in each direction) and a grass median. The lane widths are 3.5 m and the shoulder widths are 3.5 m. The posted speed limit is 100 km/h. The static radar scans 360° and can detect any moving objects within line of sight up to a maximum radius of 700 meters. The proposed smart protection system will have similar detection capabilities (see Mukherjee et al., 2013).

For the algorithm calibration, we used 51 real-world events collected on the Highway 416 section. In each situation, one or more people were performing activities outside of the vehicle (e.g., moving along a shoulder lane or even crossing a highway). Figure 3-4 shows a real-world event that can be considered

a risky circumstance. The left part of Figure 3-4 shows a truck driver (the white box) outside the truck while a vehicle is approaching (the blue arrow). The right part of Figure 3-4 shows the bird's eye view of the situation with the truck driver standing on the roadway by the front of the truck with the vehicle about to pass him. From the moment that the truck driver stepped out of the driving cabin, it took 13 seconds for the approaching vehicle to pass the driver.

We think that this situation is comparable to a police officer ticketing a stopped vehicle. For this particular situation, we were able to access recorded data showing 13 consecutive seconds of second by second trajectories of the driver moving in the area by the truck and the vehicles passing by.

Figure 3-4: Real-world situation showing a pedestrian at risk

Figure 3-5 shows the three-step calibration process required to localize the fuzzy rule-based algorithm to a study location: Step 1) Fuzzification, Step 2) Rule Evaluation, and Step 3) Defuzzification.

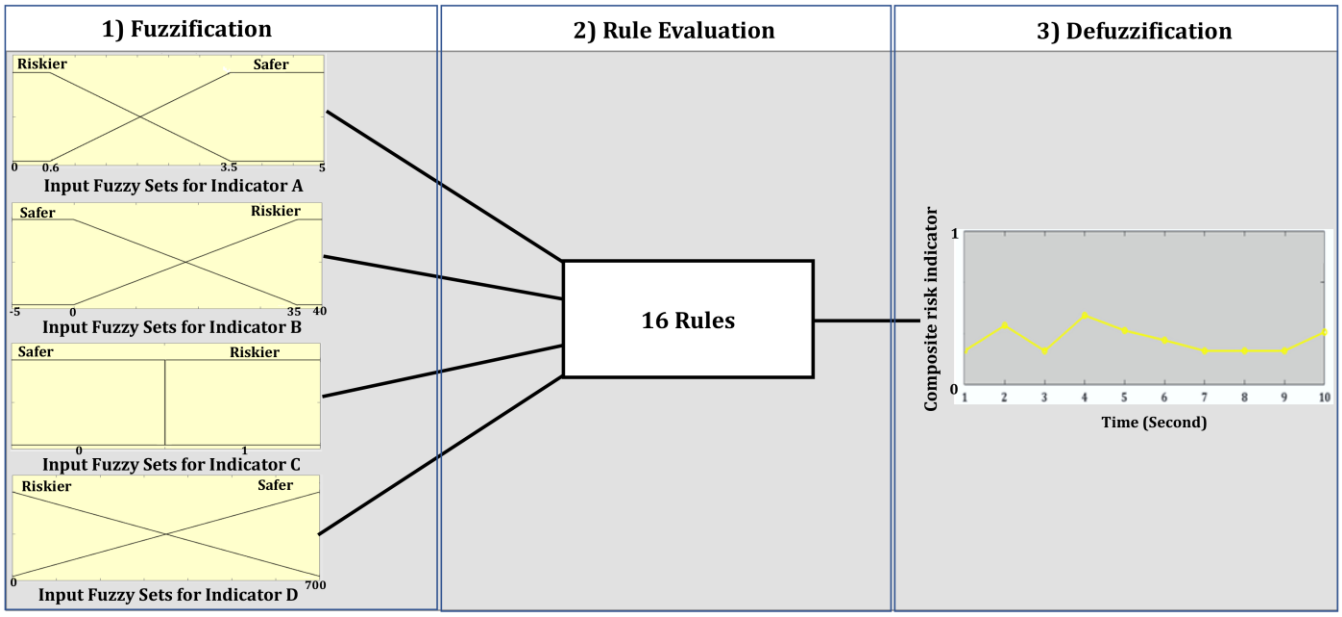


Figure 3-5: Proposed fuzzy rule-based algorithm

Step 1) Fuzzification

This step includes the following three sub-steps.

1) Determination of Input Fuzzy Set: We assigned the most basic number (two) of input fuzzy sets for each indicator, i.e., each indicator could be "safer" or "riskier."

2) Determination of Membership Function for Input Fuzzy Set: Several alternative types of membership function can be adopted including triangular, trapezoidal or bell-shape. As the trapezoidal shape function can uniformly discriminate between riskier situation and safer situation and was therefore considered the most suitable function for this study.

3) Determination of Boundary Values for Membership Function: The boundary values of a membership function play an important role in a fuzzy rule-based algorithm. In this study, the boundaries were determined by using existing knowledge obtained either from the relevant literature or from analyzing data in our study area.

For the boundary values for Indicator A, we used the lateral buffer space regulation suggested by the US Department of Transportation *(2020)*. The lateral buffer space is a lateral area that separates traffic flow from the workspace or collision site (providing safe space for workers such as police officers to perform their duties on the road). The minimum lateral buffer space is usually 0.6 m and the maximum value is usually approximately 3.5 m (one full lane width) (Owens et al., 2010, US Department of Transportation, 2013; US Department of Transportation, 2019). This study set the boundary values of the input fuzzy sets of Indicator A as 0.6 m and 3.5 m, respectively.

The boundary values for indicator B were estimated from the data for vehicle travel speeds at the study site. Figure 3-6 is a boxplot showing the distribution of the travel speeds. As the posted speed limit of the study section is 100 km/h, the boundary values for the fuzzy sets for Indicator B are 0 km/h and 35 km/h (= 135 km – 100 km), i.e., the magnitude of speeding varies from 0 km/h to 35 km/h.

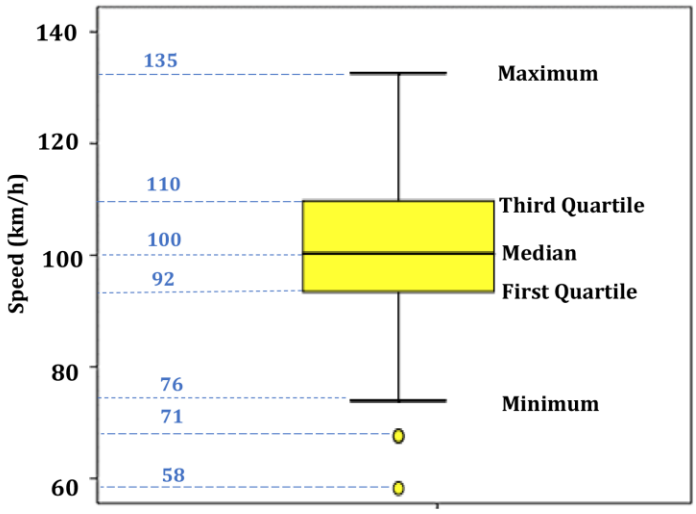


Figure 3-6: Distribution of vehicle travel speeds

The boundary value for Indictor C was based on the SSD presented in the most recent geometric guideline (AASHTO, 2018). We estimated the distance from an approaching vehicle to the projected location of police officer into the path of the vehicle and compared this distance with the SSD. If the

distance obtained was positive (higher than or equal to zero), the situation was considered safe. If the distance obtained was negative (lower than zero), the situation was considered risky. Therefore, SSD has a binary value of zero represents safer and one represents riskier.

To determine the boundary values for Indicator D, we used the maximum and minimum distance between a police officer and a vehicle. The maximum distance was the maximum distance covered by the radar, i.e., approximately 700 m. The minimum distance was zero (the distance at which a collision occurs).

Step 2) Rule Evaluation

The rules provide an important bridge between the four indicators and the composite risk indicator and allow us to apply "IF-THEN" rule constructions. As there are four indicators and each indicator has two input fuzzy sets (safer and riskier), the total number of combinations becomes 16 (= 2^4). Each combination presents a distinct rule.

Step 3) Defuzzification

For each of the 16 rules, we needed to define the output fuzzy sets associated with the composite risk indicator. This study followed the procedure suggested in previous studies to generate pre-defined output fuzzy sets (Kouloumpis et al., 2008, Phillis and Davis, 2009). We assigned an integer value 0 (safer) or 1 (riskier) to the two input fuzzy sets for each indicator. Each rule maps each input indicator (A, B, C, D) to the output fuzzy set for composite risk indicator (φ,) where $(A, B, C, D) \in \{0,1\}$, and $\varphi = A + B + C + D$. The minimum φ value of 0 (= $0 + 0 + 0 + 0$) occurs when all four input fuzzy sets are assigned as "safer", and the maximum φ value of 4 (= $1 + 1 + 1 + 1$) occurs when all four input fuzzy sets are assigned as "riskier." In our case, we had five output fuzzy sets as φ could have one of five values {0, 1, 2, 3, 4}. The five output fuzzy sets were interpreted as ranging from "very low {0}" risk to "very high {4}" risk. We mapped these five output fuzzy sets evenly to produce the composite risk indicator. The

value of composite risk indicator varies from 0 to 1.

3.3.4. Task 4: Estimate the Composite Risk Indicator

We used data from the real-world situation shown in Figure 3-4 to illustrate the proposed approach. The data included 13 consecutive seconds of trajectories of a human and an approaching vehicle. Figure 3-7 shows the four input indicators and the estimated composite risk indicator during the 13 seconds. We used the Matlab Simulink package to produce the composite risk indicator.

1) Indicator A (lateral distance from officer to traveled lane). Figure 3-7a shows that the person is positioned between 1.0 m and 1.5 m from the shoulder-side traveled lane for the entire 13 seconds. Indicator A is expected to contribute to an increase in the value of composite risk indicator if the person moves towards the traveled lane.

2) Indicator B (magnitude of speeding). Figure 3-7b shows that the magnitude of speeding is 0 km/h at each of 13 seconds except second 10 when the vehicle increases its speed to 102 giving a magnitude of speeding of 2 km. Indicator B can be expected to contribute to an increase in xrisk indicator at time 10.

3) Indicator C (stopping sight distance). Figure 3-7c shows that the SSD increases from 0 for the first 7 seconds, but increases to 1 from time 8 onwards. The value of 1 for SSD means that the vehicle could not stop in time if the person moves into the traveled lane or if the driver loses control of his vehicle. Indicator C can be expected to contribute to an increase in the value of composite risk indicator from time 8.

4) Indicator D (direct distance) Figure 3-7d shows that the direct distance between the person near the truck and the approaching vehicle decreased until time 11 and then increased after passing the person. Indicator D is expected to contribute to gradually increasing the value of composite risk

indicator as the vehicle came closer to the person up to time 11, to become 7 meters (highest risk) at time 11 and then decrease as the vehicle increases distance after passing the person at time 11.

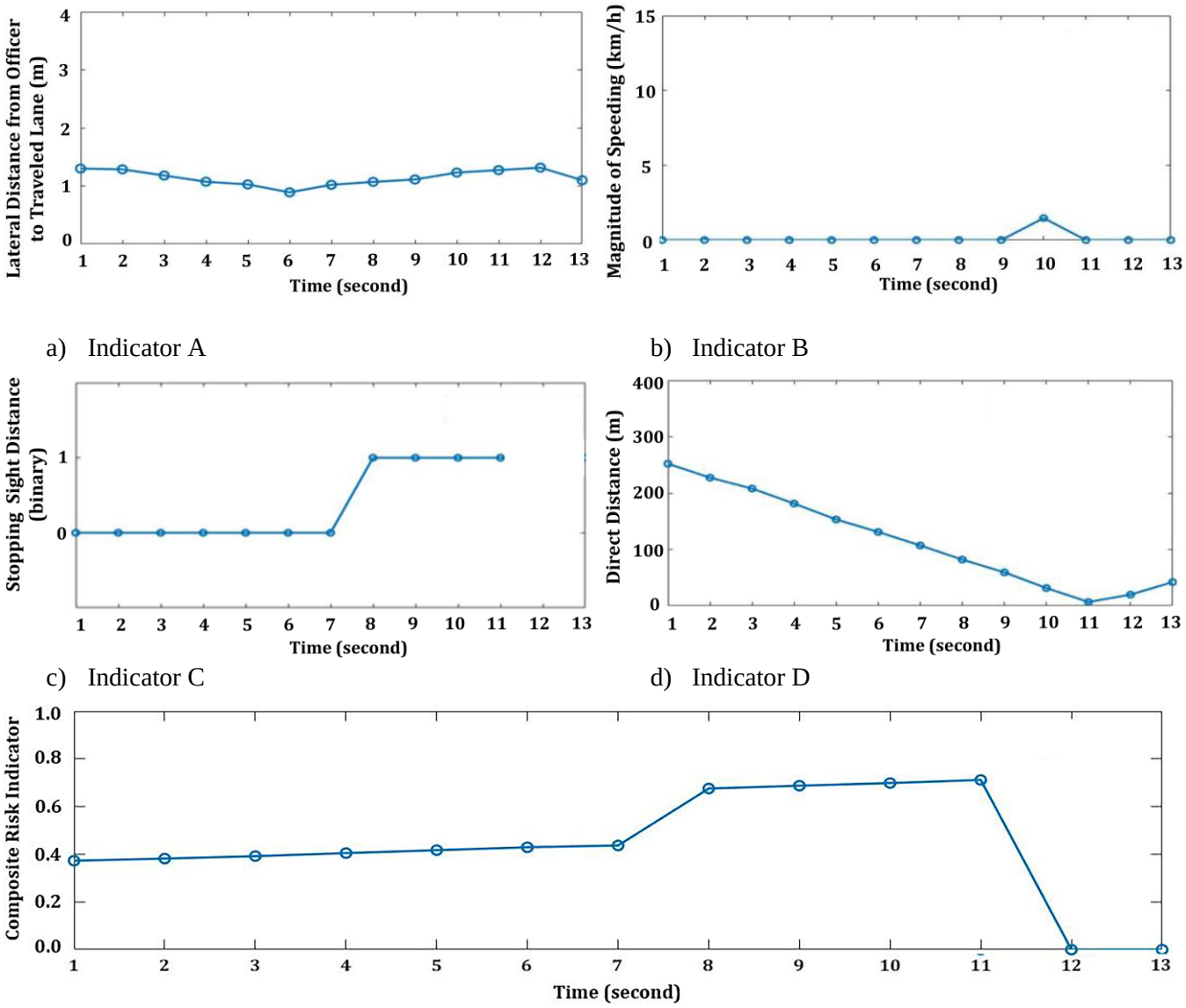


Figure 3-7: Changes in the four indicators and the estimated composite risk indicator during the 13 seconds of the event

Figure 3-7e shows the estimated composite risk indicator for the person near the truck as the vehicle passes during the 13 seconds. As expected, the composite risk indicator increases for the first 11

seconds. At time 8, the composite risk indicator jumps from 0.43 to 0.67. At this point, a collision may occur if the passing vehicle cannot stop in time to miss someone positioned within the traveled lane or if the driver loses control of the vehicle (see Indicator C). The highest value of composite risk indicator was estimated as 0.69 at time 11 when the vehicle is passing the person. At this second, the human is only 1.3 m away from the roadway.

The appropriate interpretation of the composite risk indicator and the matching appropriate warning strategy are part of Phase III (warning and communication system) and not a focus of this study, but can be briefly considered here. A warning message can be delivered to a police officer if the value of composite risk indicator is higher than a certain pre-set threshold (e.g., 0.67 or higher). The warning can be made using audio and/or vibration equipment. If the level of risk increases, audio and/or vibration equipment level can be increased.

In the particular example described in this paper, a warning would be initiated at around time 8 only if the threshold value is 0.67. Phase III will need to deepen and broaden the analysis of the questions of the warning threshold and warning level and will need to take into account additional issue such as police officers' perception and reaction times.

3.3.5. Task 5: Validate the Algorithm

The algorithm validation investigates how well the proposed system can estimate the risk involved in real-world situations (e.g., near miss collisions). Many past studies (Wang et al., 2015; Kluger et al., 2016; Xiong et al., 2019) argued that a high-risk situation can be identified by investigating vehicles' braking behavior. A hard braking behavior (indicated by a severe deceleration rate) generally indicates the possibility of risk. For example, if a passing vehicle decelerates sharply when approaching a police officer on duty, it can be assumed that the driver of the vehicle realized that there was a high level of risk of hitting the officer and braked to avoid a collision. Different studies have suggested different thresholds

for defining hard braking in a high-risk situation. For example, Wang et al. (2015), Moon et al. (2009) and Xiong et al. (2019) suggested a deceleration rate of 1.5 m/s^2, 2.0 m/s^2, 3.0 m/s^2 respectively. This study applied 2.17 m/s^2 (the average of the three values suggested) as the threshold for validating a high-risk situation using the proposed model.

To validate our algorithm's ability to identify high-risk situations, we analyzed the 51 real-world situations collected and found that 29 suggested a potentially risky situation, e.g., a vehicle traveling on the shoulder-side travel lane while a person is moving on the same lane ahead. Following Xiong et al.'s (2019) approach, we further investigated these 29 risky situations to see whether a vehicle had shown a hard brake (i.e., 2.17 m/s^2 or higher) in the last two seconds before reaching to a person to avoid collisions. Almost half (14) of the 29 situations included a vehicle that hard braked. According to the braking behavior model, these 14 real-world situations could be described as high-risk situations.

To determine the threshold for defining a situation as high risk using our fuzzy rule-based algorithm, we followed an approach used by Wang et al. (2015) and Xiong et al. (2019). Both studies used three categories (low, medium and high-risk). The estimated composite risk indicator generated by the fuzzy rule-based algorithm varies from 0 and 1. We divided this range into three equally sized categories (low (0 to 0.33), medium (0.33 to 0.67) and high (0.67 to 1.00)) and applied the categories to the 14 high risk situations identified by the braking behavior model.

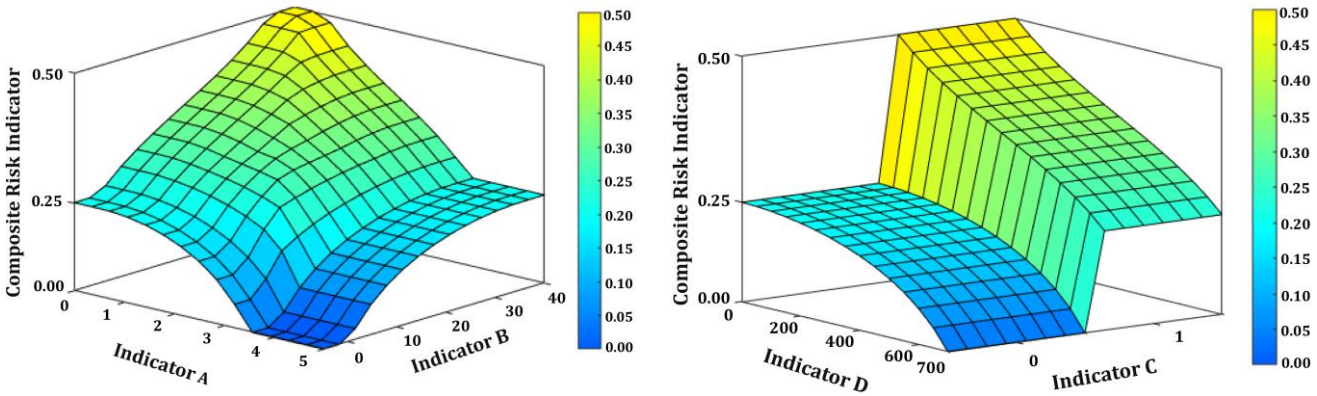
The fuzzy rule-based algorithm identified 11 of the 14 as high risk, i.e., 78%. This level of validity compares well with the results of other studies that validated their models' assessments of the risk of a collision. For instance, Wang et al. (2015) applied a K-Means clustering approach to analyze the risk of collisions and detected 66% of high-risk situations. Kluger et al. (2016) used the discrete Fourier transform algorithm combined with a K-Means clustering approach. Their study detected 78% of high-

risk situations. Xiong et al. (2019*)* proposed a fuzzy rule-based algorithm combined with K-Means clustering and detected 87% of high-risk situations.

We are aware, however, that the braking behavior model does not identify all high-risk situations. For example, some drivers suddenly and/or rapidly change lane to avoid a collision with an obstacle ahead (e.g., a pedestrian) without braking severely. This is also a high-risk situation, but it is missed by the driver braking behavior model. We were unable to include such situations in our algorithm's validation exercise, and must leave the issue of the fuzzy rule-based algorithm's detection of additional sources of risk to future research.

3.4. Discussion and Recommendations

To further illustrate the impact of the four input indicators on the estimated composite risk indicator, this study created two 3-D surfaces (See Figure 3-8). For each surface, only two of the four input indicators vary while the other two remain constant. For instance, Figure 3-8a shows how the estimated composite risk indicator changes as Indicators A (lateral distance from officer to traveled lane) and B (magnitude of speeding) change (constant values of Indicator C = 0 and Indicator D = 700 m).



a) Estimated composite risk indicator as Indicators A and B change

b) Estimated composite risk indicator as Indicators C and D change

Figure 3-8: Varying values of composite risk indicator in 3D surfaces

When Indicator A decreases and/or Indicator B increases, the composite risk indicator increases. This yields what we expect**.** Figure 3-8b shows how the composite risk indicator changes as Indicators D (direct distance) and C (SSD) change (constant values of Indicator A = 5 m and Indicator B = 0 km/h).

When Indicator D decreases, the composite risk indicator increases. Note that indicator C is binary. We see a sudden increase in the value of composite risk indicator when Indicator C changes from zero to one. These results are also what we expect.

The main focus of this study was on improving the safety of police officers, but the proposed approach can easily be expanded to target other groups of people whose duties require activities on public roadways. Such groups include, but are not limited to, public roadway maintenance crews in temporary work zones and heavy commercial vehicle drivers who park at the side of roadways for vehicle checks or short-term rests. Police officers have numerous duties on the roadway and these duties may absorb their attention and increase the risk of a collision with an approaching vehicle. This study developed a three-phase real-time risk assessment framework for a smart protection system designed to improve the safety of police officers performing duties on a roadway. This study focused on the risk analysis phase (Phase II) and used a fuzzy rule-based algorithm to combine four input indicators to generate a single composite risk indicator. Although the algorithm was based on only a limited amount of data, a validation exercise showed that the approach successfully identified many high-risk situations (defined as situations that included a hard braking vehicle).

If more data on the four input indicators can be collected, it would be possible to apply a more advanced method than the fuzzy rule-based algorithm used in this study. For example, future work may require collecting different types of high-risk situations (e.g., sudden lane change) in the field. If such data are available, a more advanced method (e.g., machine learning) might significantly improve the performance of the proposed smart protection system. Such a study might consider applying different

weights to the different input indicators instead of the equal weights used in this paper. This study collected the data for pedestrians who parked their vehicle and did activities similar to police officers; however, presence of police officers may largely affect the behavior of approaching driver. Future studies need to collect real-world data where police officers doing activities on the roadway. Searching for optimal weight values for the different input indicators would be an important objective of the study. The use of additional input indicators (e.g., weather conditions, time of day (nighttime/daytime), and work zone information regarding presence/absence of temporary traffic control devices) may increase the accuracy of risk estimates. Although our study did not consider the impact of environmental factors (e.g., weather and road condition) on the risk directly; however, the proposed risk analysis algorithm can indirectly capture the environmental factors using Indicator C (SSD). Weather and road condition will affect Indicator C (SSD) indirectly by affecting the road friction coefficient and perception and reaction time of driver. However, we did not conduct any direct or indirect analysis to capture the impact of weather and road condition due to data unavailability. Identifying a valid method for an early warning system designed to save the lives of police officers on duty on the roadway is a critically important issue. This study focuses on braking behavior, but additional data analysis from possible other data source (e.g., dashboard camera video clips showing near-miss collisions) would allow for broadening the validation approach to include, for example, steering and fast lane changing. Such studies would, however, require extensive and costly data collection. This study used 700m as the maximum distance detectable by radar. Radar coverage is largely associated with the cost of the radar system. In general, a larger detection area involves higher cost.

Future research could consider some of the limitations of this study. For example, the Phase II analysis assumed that Phase I (detection) used an advanced radar system that delivered flawless data. In reality, occlusions of moving objects due to blocking of lines of sight are a well-known challenge

affecting the continuous tracking of multiple moving objects by any existing radar system. Improvements in Phase I detection are important as our system requires reliable and simultaneous detection of the presence and approaching speed of one or more vehicles. Future work needs to resolve the many issues related to Phase III (warning and communication system). These issues include the selection of the most effective warning technologies to disseminate the varying levels of warning, and the determination of appropriate composite risk indicator thresholds for each level of warning. If the threshold is too low, a high number of warnings could cause unnecessary distraction/stress for police officers, but if the threshold is too high, the system may miss an emergency situation that requires a police officer to react. Eventually, the communication capability of the proposed smart protection system should be enhanced to create a real V2I system that could, for example, communicate with approaching vehicles and automatically reduce their speed and/or change path to avoid possible collisions with police officers working on a road.

3.5. Chapter 3 References

Arun, A., Haque, M. M., Bhaskar, A., Washington, S., and Sayed, T. (2021). A systematic mapping review of surrogate safety assessment using traffic conflict techniques. Accident Analysis & Prevention. 153, 106016.

American Association of State Highway and Transportation Officials (AASHTO). (2018). A Policy on Geometric Design of Highway and Streets.

Castignani, G., Derrmann, T., Frank, R., and Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. IEEE Intelligent Transportation Systems Magazine. 7(1), 91-102.

Dayal, S., Mortazavi, A., Huynh, K. H., Gerges, R. L., and Shynk, J. J. (2013). A cooperative high-accuracy localization algorithm for improved road workers' safety. In 2013 Asilomar Conference on Signals, Systems and Computers (pp. 1709-1713). IEEE.

Eftekhari, H. R., and Ghatee, M. (2018). Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition. Transportation Research Part F: Traffic Psychology and Behaviour, 58, 782-796.

Federal Bureau of Investigation (FBI). (2019). Law enforcement officers accidentally killed- type of accident and activity of victim officer at time of incident, 2015–2019. Available at: https://ucr.fbi.gov/leoka/2019/tables/table-65.xls [Accessed June, 2022].

Foerster, K. T., Gross, A., Hail, N., Uitto, J., and Wattenhofer, R. (2014). Spareeye: enhancing the safety of inattentionally blind smartphone users. In Proceedings of the 13th international conference on mobile and ubiquitous multimedia. (pp. 68-72).

Hassein, U., Diachuk, M., and Easa, S. M. (2018). In-vehicle passing collision warning system for two-lane highways considering driver characteristics. Transportation Research Record. 2672(37), 101-112.

Hauer, E. (2009). Speed and safety. Transportation Research Record. 2103(1), 10-17.

Johnsson, C., Laureshyn, A., and De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. Transport Reviews. 38(6), 765-785

Ka, D., Lee, D., Kim, S., and Yeo, H. (2019). Study on the framework of intersection pedestrian collision warning system considering pedestrian characteristics. Transportation Research Record. 2673(5), 747-758.

Kluger, R., Smith, B. L., Park, H., and Dailey, D. J. (2016). Identification of safety-critical events using kinematic vehicle data and the discrete fourier transform. Accident Analysis & Prevention. 96, 162-168.

Kouloumpis, V. D., Kouikoglou, V. S., and Phillis, Y. A. (2008). Sustainability assessment of nations and related decision making using fuzzy logic. IEEE Systems Journal. 2(2), 224-236.

Lee, D., Pietrucha, M. T., and Donnell, E. T. (2006). Hierarchical fuzzy inference system to evaluate expert opinions on median safety. Transportation Research Record. 1961(1), 34-43.

Li, Y. Z., Hu, H., and Huang, D. Z. (2012). Dynamic fuzzy logic model for risk assessment of marine crude oil transportation. Transportation Research Record. 2273(1), 121-127.

Lin, C. H., Chen, Y. T., Chen, J. J., Shih, W. C., and Chen, W. T. (2016). psafety: A collision prevention system for pedestrians using smartphone. In 2016 IEEE 84th Vehicular Technology Conference. (pp. 1-5). IEEE.

Liu, M., Lu, G., Chen, Y., Zhang, X., and Li, B. (2017). Modeling of vehicles merging movement at unsignalized intersections considering drivers' risk perception. Transportation Research Record. 2663(1), 73-81.

Mahmud, S. S., Ferreira, L., Hoque, M. S., and Tavassoli, A. (2019). Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment. IATSS research, 43(1), 27-36.

Mathworks. (2016). MATLAB & Simulink: Simulink reference r2015b.

Mendel, J. M. (1995). Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE, 83(3), 345-377.

Mirabbasi, R., Mazloumzadeh, S. M., and Rahnama, M. B. (2008). Evaluation of irrigation water quality using fuzzy logic. Research Journal of Environmental Sciences. 2(5), 340-352.

Moon, S., Moon, I., and Yi, K. (2009). Design, tuning, and evaluation of a full-range adaptive cruise control system with collision avoidance. Control Engineering Practice. 17(4), 442-455.

Moudon, A. V., Lin, L., Hurvitz, P., and Reeves, P. (2008). Risk of pedestrian collision occurrence: case control study of collision locations on state routes in King County and Seattle, Washington. Transportation Research Record. 2073(1), 25-38.

Mukherjee, A., Stolpner, S., Liu, X., Vrenozaj, U., Fei, C., and Sinha, A. (2013). Large animal detection and continuous traffic monitoring on highways. Sensors, IEEE (pp. 1-3).

Mullakkal-Babu, F. A., Wang, M., Farah, H., van Arem, B., and Happee, R. (2017). Comparative assessment of safety indicators for vehicle trajectories on highways. Transportation Research Record. 2659(1), 127-136.

National Highway Traffic Safety Administration (NHSTA). 2013. Get the facts: Move Over: It's the Law, NHTSA. Available at: https://www.nhtsa.gov/staticfiles/communications/pdf/MoveOver_QA.pdf [Accessed June, 2022].

National Highway Traffic Safety Administration (NHSTA). (2021). Traffic Safety Facts: Speeding. National Highway Traffic Safety Administration, Washington, D.C., 2021. Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813194 [Accessed June, 2022].

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., and Trego, T. (2010). Traffic incident management handbook. No. FHWA-HOP-10-013.

Phillis, Y. A., and Davis, B. J. (2009). Assessment of corporate sustainability via fuzzy logic. Journal of Intelligent and Robotic Systems. 55(1), 3-20.

Roofigari-Esfahan, N., White, E., Mollenhauer, M., and Talledo Vilela, J. P. (2021). Development of a connected smart vest for improved roadside work zone safety, Virginia Tech Transportation Institute.

Schroeder, P., Kostyniuk, L., and Mack, M. (2013). National survey of speeding attitudes and behaviors. No. DOT HS 811 865. United States. National Highway Traffic Safety Administration. Office of Behavioral Safety Research.

Teodorović, D. (1994). Fuzzy sets theory applications in traffic and transportation. European Journal of Operational Research. 74(3), 379-390.

Tong, Y., Jia, B., Wang, Y., and Yang, S. (2018). Detecting pedestrian situation awareness in real-time: algorithm development using heart rate variability and phone position. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting. (Vol. 62, No. 1, pp. 1579-1583). Sage CA: Los Angeles, CA: SAGE Publications.

Tong, Y., and Jia, B. (2019). An Augmented-reality-based warning interface for pedestrians: user interface design and evaluation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 63, No. 1, pp. 1834-1838). Sage CA: Los Angeles, CA: SAGE Publications.

Transportation Association Canada (TAC), Geometric Design Guide (2019). Available at: https://www2.gov.bc.ca/assets/gov/driving-and-transportation/transportation-infrastructure/engineering-standards-and-guidelines/highway-design-and-survey/tac/tac-2019-supplement/bctac2019-chapter-0000-front.pdf [Accessed June, 2022]

US Fire Administration. (2008). Traffic incident management systems. Federal Emergency Management Agency.

US Fire Administration. (2012). Traffic incident management systems. Federal Emergency Management Agency.

US Department of Transportation. (2013). National traffic incident management responder training program, Federal Highway Administration.

US Department of Transportation. (2020). Manual on uniform traffic control devices for streets and highways. US Department of Transportation, Federal Highway Administration.

Wang, T., Cardone, G., Corradi, A., Torresani, L., and Campbell, A. T. (2012). Walksafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads. 12th Workshop on Mobile Computing Systems & Applications. (pp. 1-6).

Wang, J., Zheng, Y., Li, X., Yu, C., Kodaka, K., and Li, K. (2015). Driving risk assessment using near-crash database through data mining of tree-based model. Accident Analysis & Prevention. 84, 54-64.

Xiao, J., Kulakowski, B. T., and EI-Gindy, M. (2000). Prediction of risk of wet-pavement accidents: Fuzzy logic model. Transportation Research Record. 1717(1), 28-36.

Xiong, X., Wang, M., Cai, Y., Chen, L., Farah, H., and Hagenzieker, M. (2019). A forward collision avoidance algorithm based on driver braking behavior. Accident Analysis & Prevention, 129, 30-43.

Zadeh, L. A. (1965). Fuzzy sets. Information and control. 8(3), 338-353.

CHAPTER FOUR

4. DEVELOP A SEMI-SUPERVSIED MACHINE LEARNING ALGORITHM FOR A SMART PROTECTION SYSTEM TO PREVENT COLLISIONS WITH POLICE OFFICERS ON DUTY ON THE ROADWAY

Abstract: Current smart protection systems designed to prevent vehicle collisions with police officers, work zone workers, and other pedestrians working on roadways have major limitations: the systems use only one input indicator; they apply a simple risk analysis algorithm based on a fixed threshold to identify risky situations; they cannot simultaneously capture a variety of important collision contributing factors; and the simple risk analysis algorithm cannot combine multiple contributing factors into a single composite risk indicator. This study screened and selected four input indicators to capture officer behavior, driver behavior, and the spatial and temporal interaction between an officer and a vehicle. The study used real-world data from Highway 416 in Ontario, Canada to develop a semi-supervised machine learning algorithm. We successfully used driver braking behavior to detect the small number of probable near miss collisions in the dataset. These near miss collisions provided labeled data. We used a heuristic approach based on a fuzzy rule-based algorithm to give labels to the unlabeled data in our dataset. The semi-supervised algorithm then combined the labeled and unlabeled data and estimated the level of risk (low, medium and high). The study also developed an unsupervised machine learning algorithm to compare with the developed semi-supervised algorithm. The results showed that the semi-supervised machine learning algorithm accurately estimated 94% of the risk labels, successfully identified 88% of the near miss collisions, and provided warning of the risk of a collision three to four seconds in advance.

Keywords: *Collisions with police officers, Semi-supervised machine learning, Work zone, Risk analysis, Pedestrian-to vehicle collisions.*

4.1. Introduction

This study is concerned with the development of a semi-supervised machine learning algorithm as part of a smart protection system designed to prevent collisions between vehicles and police officers performing duties on a roadway. The study was conducted using real-world data from Highway 416 in Ontario, Canada. This study relies on the results of Chapter 3.

4.1.1. Number of Collisions between Vehicles and Pedestrians Working on Roadways in the United States

Police officers are exposed to the risk of being hit by a passing vehicle while they perform duties such as speed enforcement, overseeing work zones, assisting/investigating vehicle collisions, etc. Other workers exposed to similar risks include work zone workers, first responders and mining workers. In only two years, (2018 and 2019), the United States recorded 252 worker deaths in work zones (126 deaths per year) (NHSTA, 2020).

Being struck by a vehicle is the second most common hazard affecting police officers' lives, i.e., a very serious issue (Tofighi et al., 2021). During the ten years from 2011 to 2020, 92 police officers were struck and killed in the United States while performing road related duties. Traffic stops accounted for 24% of the 92 deaths and assisting/investigating stopped vehicles accounted for 20% (FBI, 2021).

4.1.2. Existing Systems for Preventing Vehicle Collisions with Pedestrians Working on Roadways

Research has shown that detecting drivers who take risks and providing early warning to people working on foot on roadways are important ways to reduce the number of collisions (Montgomery 2019; Mollenhauer et al., 2019). Smart protection systems such as smart vests, smart cones and smart helmets can be equipped with sensors (e.g., radar, lidar and Bluetooth) which can identify risk taking drivers and

disseminate early warning to pedestrian roadway workers (Nnaji et al., 2018).

Existing systems use only one input indicator (e.g., an intrusion of a passing vehicle into a work zone/safety area) to identify risk taking passing drivers and to disseminate early warning. For example, Mollenhauer et al. (2019) developed "Work Zone Intrusion Alarm Technology" which aims to warn workers in work zones about a potentially dangerous passing vehicle. The system relies on smart cones to identify intrusions by passing vehicles into a work zone. Another example is the "Traffic Accident Warning System" which aims to warn police officers at the site of a collision about a potentially dangerous passing vehicle in order to prevent a secondary collision (Confucian, 2022). If a smart cone is hit by a passing vehicle, the cone immediately sends out a signal which alerts the police officer. In the mining industry, sensors (e.g., Bluetooth) embedded into mining equipment and a worker's smart helmet detect intrusions by a potentially dangerous passing vehicles such as an excavators or dump trucks into defined safety areas (e.g., a circle with radius of 10m). The system immediately alerts affected pedestrian mining workers (Baek and Choi 2018).

4.1.3. Gaps in Current Protection Systems and Research

Existing systems use one input indicator (intrusion of a passing vehicle into an area of pedestrian workers) and apply a simple risk analysis algorithm (i.e., compare input indicator with a fixed threshold) to differentiate a safe situation from a risky situation. There are clear gaps in the research into input indicator selection and into risk analysis algorithms.

4.1.3.1. Gaps in Input Indicator Selection

Existing systems use one input indicator (intrusion of a passing vehicle into an area of pedestrian workers) as inputs for risk analysis. This input indicator is an example of a surrogate safety indicator. Surrogate safety indicators are used to estimate the risk of a collision between road users in real-time (Zheng et al., 2021). A typical transportation engineering surrogate safety indicator is time-to-collision (TTC) (Li et

al., 2022). TTC represents the time required for two road users to collide if they continue at their present speed and on the same path. Researchers assign different risk thresholds to the surrogate safety indicator to classify the risk of a collision as, for example, low, medium or high (Nadimi et al., 2016; Xiong et al., 2019). Li et al. (2022), for example, classified a collision risk as high if TTC was less than 1.1 s, medium if TTC was 1.9 to 2.9 s, or low if TTC was greater than 2.9 s.

Transportation engineers have created numerous surrogate safety indicators to consider different aspects of a collision and assess the level of collision risk in real-time. Surrogate safety indicators can be divided into two groups:

1. Surrogate safety indicators that attempt to capture the **spatial and temporal interactions** between two road users by measuring the distance or time between the road users (Johnsson et al., 2018). These indicators include TTC, post encroachment time, braking time, lateral distance, etc.
2. Surrogate safety indicators that attempt to capture the **behavior of road users** (e.g., aggressiveness) (Terzi, 2018). These indicators include magnitude of speeding (the speed differential between the actual speed and the posted speed limit), sudden lane change, sudden brake force during deceleration, etc.

It is clear that no single indicator can simultaneously capture the various aspects of a collision, i.e., the spatial and temporal interactions between two road users in distance or time and the road behavior of the road users (Johnsson et al., 2018). Although a pedestrian roadside worker is at risk from different types of collision including a head-on collision, a crossing the road collision, and side-swipe collision, the use of a single indicator in existing protection systems means that the systems can capture only one type of collision (Johnsson et al., 2018). For example, in the case of a rear-end or head-on collision, TTC, time headway, braking time are useful (Mahmud et al., 2019). In the case of a pedestrian crossing

collision, PET and time to line crossing are useful (Mahmud et al., 2019). In the case of a side-swipe collision, lateral distance between two road users is useful.

An additional issue is that simply introducing multiple indicators and conducting a separate risk for each indicator does not solve the problem as indicators can result in different conclusions (Chen et al. 2015). The assessment of collision risk requires a sophisticated algorithm that can integrate multiple input indicators into a single composite risk indicator. The next sub-section addresses gaps in existing risk analysis algorithms.

4.1.3.2. Gaps in Existing Risk Analysis Algorithms

Existing systems apply a simple risk analysis algorithm and use a fixed threshold. This approach is straightforward, but it cannot combine multiple input indicators.

To find a risk analysis algorithm that could combine input indicators, we investigated risk analysis algorithms that have been used to combine multiple input indicators for other types of collision. Many studies have applied statistical algorithms to combine multiple input indicators. These algorithms include the binary logit algorithm (Xing et al., 2019), the ordered logit algorithm (Tageldin and Sayed, 2019), the multinomial logit algorithm (Dimitriou et al., 2018), and the nonlinear regression algorithm (Gu et al., 2019). The algorithms use a statistical function (applied to a logistic/normal distribution) to link the independent variables (e.g., input indicators) to a dependent variable (composite risk indicator). Although statistical algorithms are popular, easy to perform, and simple to understand, they need a large amount of data including actual collision and near miss collision data to calibrate the algorithm, and they cannot capture uncertainty.

In our study, we faced data uncertainties and limited availability as no dataset (including our dataset from Highway 416) provides position data (x- and y-coordinates) or the speed of road users in the seconds leading up to a collision or near miss collision. Our dataset also lacked data for a large number of

collisions or near miss collisions. As we did not know the independent variables (e.g., input indicators) and the dependent variable (composite risk indicator), we could not simply use a statistical algorithm for our study.

Several studies have developed machine learning algorithms. Machine learning algorithms learn from past historical data, and can be used to combine several input indicators into a single composite risk indicator to estimate the risk associated with collisions, but the studies available focused on vehicle-to-vehicle collisions (Wang et al., 2015; Xiong et al., 2019).

The historical data used in machine learning algorithms are classified into labeled and unlabeled data. Data are considered labeled if the outcome of certain combinations of input indicators can reasonably be considered to indicate a certain risk level (e.g., low, medium or high). Data are considered unlabeled data if the risk level is unknown.

Machine learning can be supervised, unsupervised or semi-supervised (Wang et al., 2015).

In supervised machine learning, all data are labeled. The algorithms create a risk analysis by finding relationships between the labels and the input indicators. The most widely used algorithms are Artificial Neural Networks (ANNs) (Firat et al., 2010).

In unsupervised machine learning, the data are unlabeled. The algorithms classify the data for the input indicators into multiple risk labels based on patterns/structure (e.g., closeness to the centre) in the data. Unsupervised algorithms include K-Nearest Neighbor, K-Means, and Fuzzy C-Means (Tang et al., 2020).

Semi-supervised machine learning uses a combination of a limited amount of labeled data (e.g., near miss collision data) and a larger amount of unlabeled data (Cusick et al., 2021), but as the labeled data are insufficient for training the machine learning algorithm, researchers use a heuristic approach to

give labels to the larger amount of the unlabeled data. When all the data have labels, it is possible to train the machine learning algorithm and develop the risk analysis.

As we had a combination of a small amount of labeled data (e.g., for near miss collisions) and a larger amount of unlabeled data, we included an evaluation of the possibility of developing a semi-supervised machine learning algorithm. To the best of our knowledge, there has been no research into the development of a semi-supervised machine learning algorithm that can provide the risk analysis required for a smart protection system that can alert pedestrian roadway workers to the risk of being hit by a passing vehicle.

4.1.4. Developing a Smart Protection System for Police Officers Working on Foot

Figure 4-1 shows the three high level tasks involved in a smart protection system designed to provide early warning for police officers working on foot on a roadway:

I. Detection: a radar system installed in a police officer's car detects the positions (e.g., x- and y-coordinates) of passing vehicles and police officers in real-time;

II. Risk analysis: a machine learning algorithm uses multiple input indicators to estimate a composite risk indicator which has three risk labels (low, medium, high); and

III. Warning and communication: the system provides early warning of high-risk situations to police officers working on foot on a roadway.

Although our ultimate aim is the development of a single-unit machine that performs all three tasks simultaneously and automatically, this study focused primarily on Task II, the risk analysis. For Task I, conventional radar systems that can detect required input indicators such as the second by second positions of police officers and vehicles are available. We adopted a radar system that is very close to the system

developed by Mukherjee et al. (2013). We consider that Task III requires additional separate studies as warning and communication are outside the scope of this study.

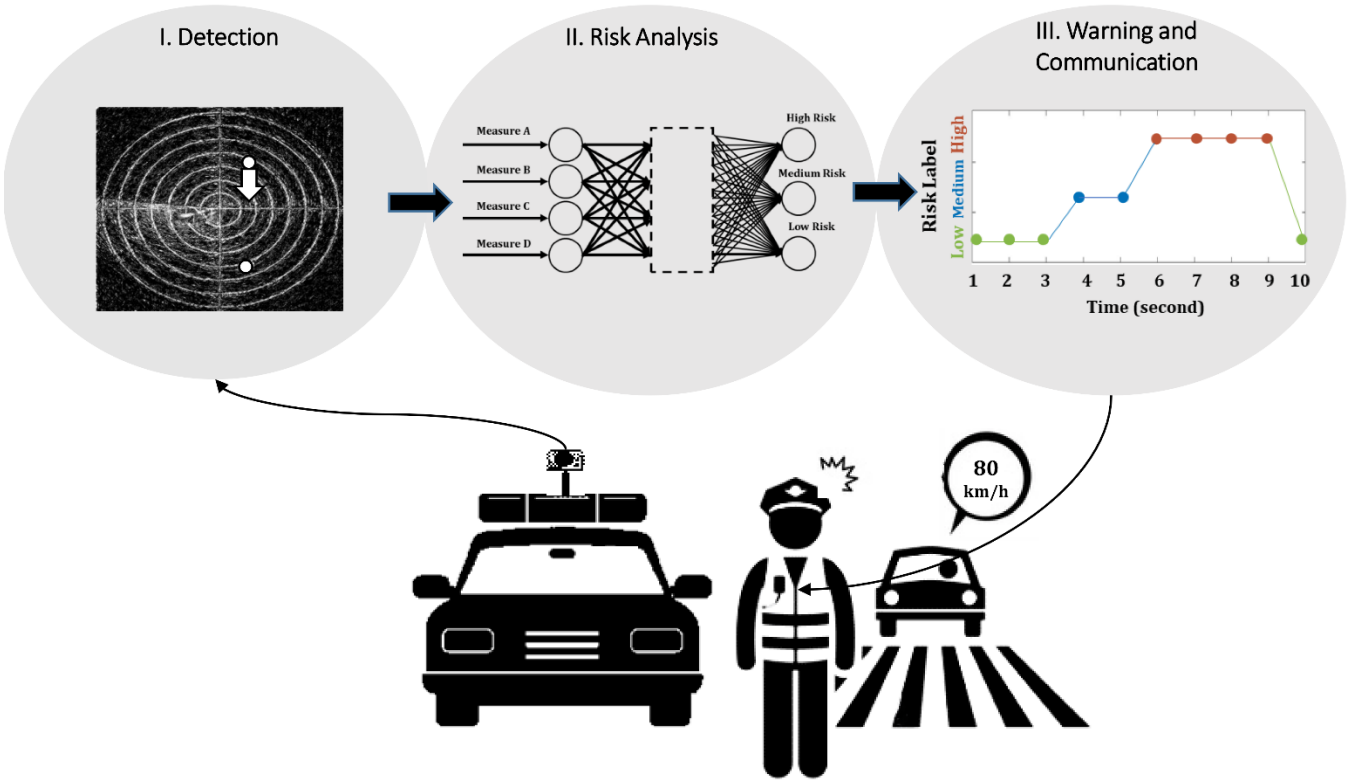


Figure 4-1: Proposed smart protection system

4.1.5. Study Goal and Objectives

The goal of this study was to develop a semi-supervised machine learning algorithm can produce the risk analysis required for an advanced smart protection system that provides police officers working on foot on a roadway with an early warning that they are at risk being hit by a vehicle.

This study had three objectives:

1) Perform the labeling for a portion of the data in a dataset of unlabeled data. This approach identifies near miss collisions which can be used as small number of labeled data in the risk analysis;

2) Develop a semi-supervised machine learning algorithm that combines labeled and unlabeled data and estimates levels of composite risk indicator into three categories (low, medium and high); and

3) Apply an unsupervised machine learning algorithm that classifies data based on their similarities into three risk categories (low, medium and high). Using a real-world dataset, evaluate the performance of the semi-supervised machine learning algorithm and compare the performance to the performance of an unsupervised machine learning algorithm.

4.2. Study Data

We used second by second position data (x- and y-coordinates) for passing vehicles and pedestrians obtained from a radar system installed on a section of Highway 416 near Kemptville, Ontario, Canada. The study used three months of data collected from December 2019 to February 2020.

The section of Highway 416 is 1.4 km of a rural highway with four lanes (two lanes in each direction), a grass median and shoulder lanes. The lanes are 3.5m wide, the grass median is 25 m wide, and the shoulder lanes are 3.5 m wide. The posted speed limit is 100 km/h.

The radar scans 360° and can detect any moving object within the line of sight up to a maximum 700 meters of radius. We expect that the proposed smart protection system will have similar capabilities in terms of the detection area.

In this study, an “event” was defined as a situation in which the simultaneous movement of one (or more) pedestrian(s) and the movement of one (or more) vehicle(s) was detected within the detection area. During the data collection period, the radar system detected 70 events which produced 931 data points (one data point referred to one second of detection period).

The four thermal images on the left side of Figure 4-2 show examples of potentially risky events. The right side shows a bird’s eye view of the detection area. In the thermal images, vehicle drivers are walking along the shoulder lane to look into possible mechanical issues in their vehicles. We assumed that

pedestrian movements in these situations were comparable to police officers' activities on the roadways, for example, investigating and ticketing a stopped vehicle, etc.

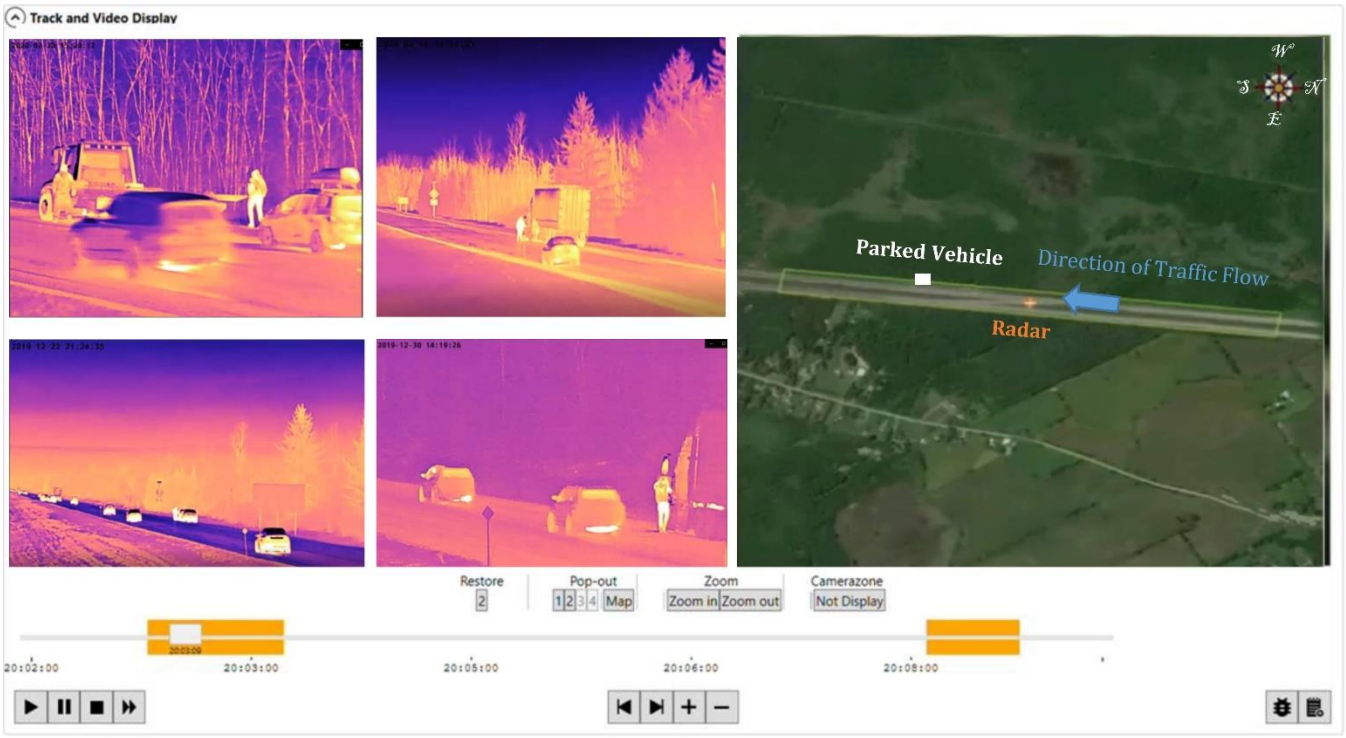


Figure 4-2: Pedestrian activities on the road/roadside in the study area

4.3. Methodology

Figure 4-3 shows the study's four sub-tasks: 1) identify labeled and unlabeled data, 2) screen, select, estimate, and normalize input indicators 3) develop a semi-supervised machine learning algorithm and apply an unsupervised machine learning algorithm, and 4) propose evaluation measures.

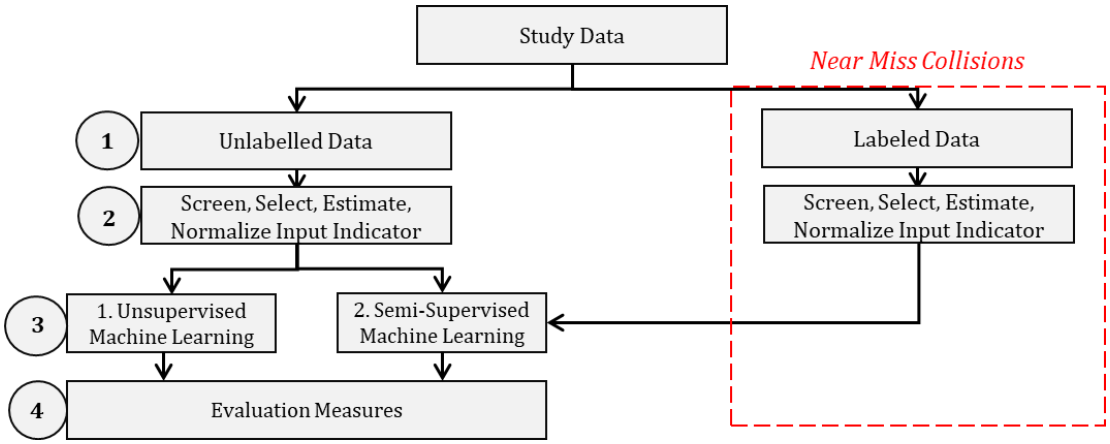


Figure 4-3: Study methodology

4.3.1. Identify Labeled and Unlabeled Data

As mentioned in section 4.1. Introduction, data are considered to be labeled if the data can be used to assess risk levels. To identify labeled data, we used driver braking behavior to identify near miss collisions. Driver braking behavior was also used to identify near miss collisions by, for example, Moon et al. (2009) and Wang et al. (2020).

A hard brake (indicated by a severe deceleration rate) generally indicates the possibility of high risk. Indeed, Bagdadi (2013) showed a high correlation between the frequency of hard braking events and the frequency of actual collisions. Studies have suggested various thresholds for defining hard braking behavior. For example, Moon et al. (2009), Xiong et al. (2019), and Wang et al. (2020) suggested a deceleration rate of 2.0 m/s^2, 3.0 m/s^2, and 1.5 m/s^2 respectively. In this study, we used at least 2.1 m/s^2 (the average deceleration rate suggested by these three studies) as the threshold defining a hard brake.

We identified two conditions for identifying a near miss collision:

1) A vehicle was traveling on the shoulder-side travel lane while a pedestrian was walking along the same lane or shoulder lane ahead; and

2) A driver took a hard brake (at least 2.1 m/s^2) in the last two seconds before passing the pedestrian. This condition assumed that the hard-braking behavior was related to pedestrian activity on the roadway.

We analyzed the 70 real-world events in our dataset and found that 15 of the events met our conditions. These 15 events were identified as near miss collisions and were considered as labeled data representing high-risk.

Each of 15 identified events included approximately 12 seconds of data for the position of passing vehicles, the speed of passing vehicles, and the speed of pedestrians walking on the roadway. We need to select the seconds (datapoints) in which the near miss collision happened and give high-risk label to the datapoints. Drivers perceive a near miss collision and react by hard brake. According to AASHTO (2018), the perception and reaction times for a normal driver is usually considered to be 2.5 seconds. This value can accommodate 90% of drivers. In fact, a driver feels the possibility of a near miss collision about 2.5 seconds before a hard brake. Thus, we consider the moment of hard braking and 2 seconds before the hard brake.

The 15 events provided 46 high-risk labeled data points.

The remaining 55 (70 – 15 = 55) events presented 885 unlabeled data points. We later on used a heuristic approach (a fuzzy rule-based algorithm) to give labels (low, medium, and high) to the unlabeled data points.

4.3.2. Screen, Select and Normalize Input Indicators

Input indicators were screened by reviewing the surrogate safety indicators discussed in the literature (e.g., Johnsson et al., 2018; Arun et al., 2021). As road user behaviors have been shown to be involved in 90% of collisions on roadways (Terzi, 2018), we selected two road user behaviors as input indicators

We also selected two input indicators based on the spatial and temporal interactions between a pedestrian

and a vehicle (see Indicator C and D in Figure 4-4). Indicator C could also capture environmental factors (e.g., road conditions).

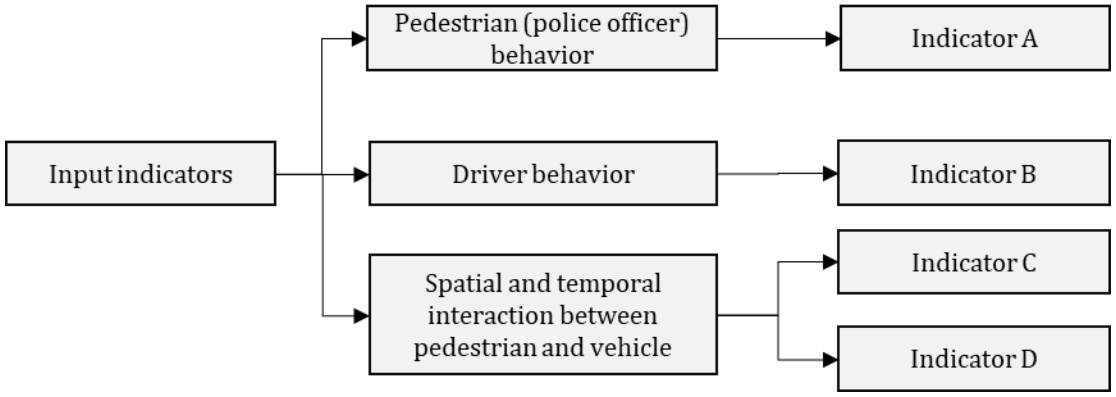


Figure 4-4: The four input indicators

1) Indicator A (lateral distance from officer to traveled lane) referred to pedestrian (police officer) behavior which was measured as the perpendicular lateral distance between an officer and a traveled lane. When a pedestrian is working on a shoulder lane, there is a possibility of being hit by a passing vehicle. Highway designers try to provide a safe lateral distance between pedestrians working on roadways and the traveled lane (US DOT, 2020). This distance is based on the concept of lateral buffer space which refers to the lateral area that separates traffic flow from the workspace or collision site and provides a safe space for workers such as police officers to perform their duties on the road. The minimum lateral buffer space is usually 0.6m and the maximum value is usually approximately 3.5m (one full lane-width) (US DOT, 2020). As the lateral distance decreases, the risk increases.

2) Indicator B (magnitude of speeding) referred to aggressive driving behavior, a major concern in road safety research especially due to the association between aggressive driving behavior and severe collisions (Fountas et al., 2019). An aggressive driver increases a collision chance by 55% (Song et al., 2021). The magnitude of speeding was used to measure aggressive driving behavior, a commonly used approach (Castignani et al. 2015). The magnitude of speeding is the speed differential between a vehicle's

travel speed and the posted speed limit. As the magnitude of speeding increases, the risk increases.

3) Indicator C (Stopping Sight Distance (SSD)) provided an estimate of the spatial and temporal interactions between a pedestrian and a vehicle by considering the minimum distance at which a driver notices an object (in our case a police officer) on the roadway and can make a complete stop to avoid a collision with the object (AASHTO, 2018). Stopping Sight Distance (SSD) was used to measure this distance. SSD is an important highway design parameter that considers a variety of safety elements including perception-reaction time, speed, and the friction coefficient (see Equation (4-1)). In this study, the SSD indicator was interpreted as a binary indicator. If the distance between a vehicle and the nearest point of conflict was less than the SSD, the situation was considered risky. If the distance was longer than the SSD, the situation was considered safe. Equation (4-1) shows that

$$SSD = Vt_{pr} + \frac{V^2}{2\mu(f)} \quad (4\text{-}1)$$

where:

SSD = stopping sight distance (m);

V = vehicle speed (m/s);

t_{pr} = perception-reaction time (s);

μ = 9.8 m/s^2; and

f = friction coefficient.

Like Indicator C, Indicator D (direct distance) provided an estimate of the spatial and temporal interactions between a pedestrian and a vehicle, but Indicator D was simply the direct distance between the pedestrian and the vehicle. This indicator was selected because in certain situations, Indicators A, B, and C cannot capture the potential risk of a situation and would not trigger a warning that a vehicle was approaching. For instance, if a police officer was located outside the threshold lateral distance from the roadway, an approaching vehicle was traveling at less than the posted speed limit, and the vehicle's estimated SSD appeared adequate, the system would not trigger any warnings. As the direct distance

decreases, the risk increases.

Interquartile range filtering was used to detect and remove outliers in our data (Coifman and Kim, 2009). The upper and lower limits of the valid data range for each input indicator were determined as $Upper\ Limit = Q_3 + IQR \times 1.5$ and $Lower\ Limit = Q_1 - IQR \times 1.5$ where Q_1 and Q_3 are the first and third quartiles of the datapoint respectively. IQR is the interquartile range. Datapoints which lay outside of the upper and lower limits were considered outliers.

Figure 4-5 provides a histogram and box plot of each indicator. The blue boxes are box plots showing the IQR.

For Indicator A (lateral distance from officer to traveled lane), a value higher than 15m was considered an outlier. Removing values higher than 15m is logical as this threshold is high enough for a police officer to be safe.

For Indicator B (magnitude of speeding), a value higher than 30km/h was considered an outlier. The posted speed limit in this corridor was 100 km/h and Figure 4-5.b shows that few drivers in this corridor drove at speeds of greater than 130 km/h.

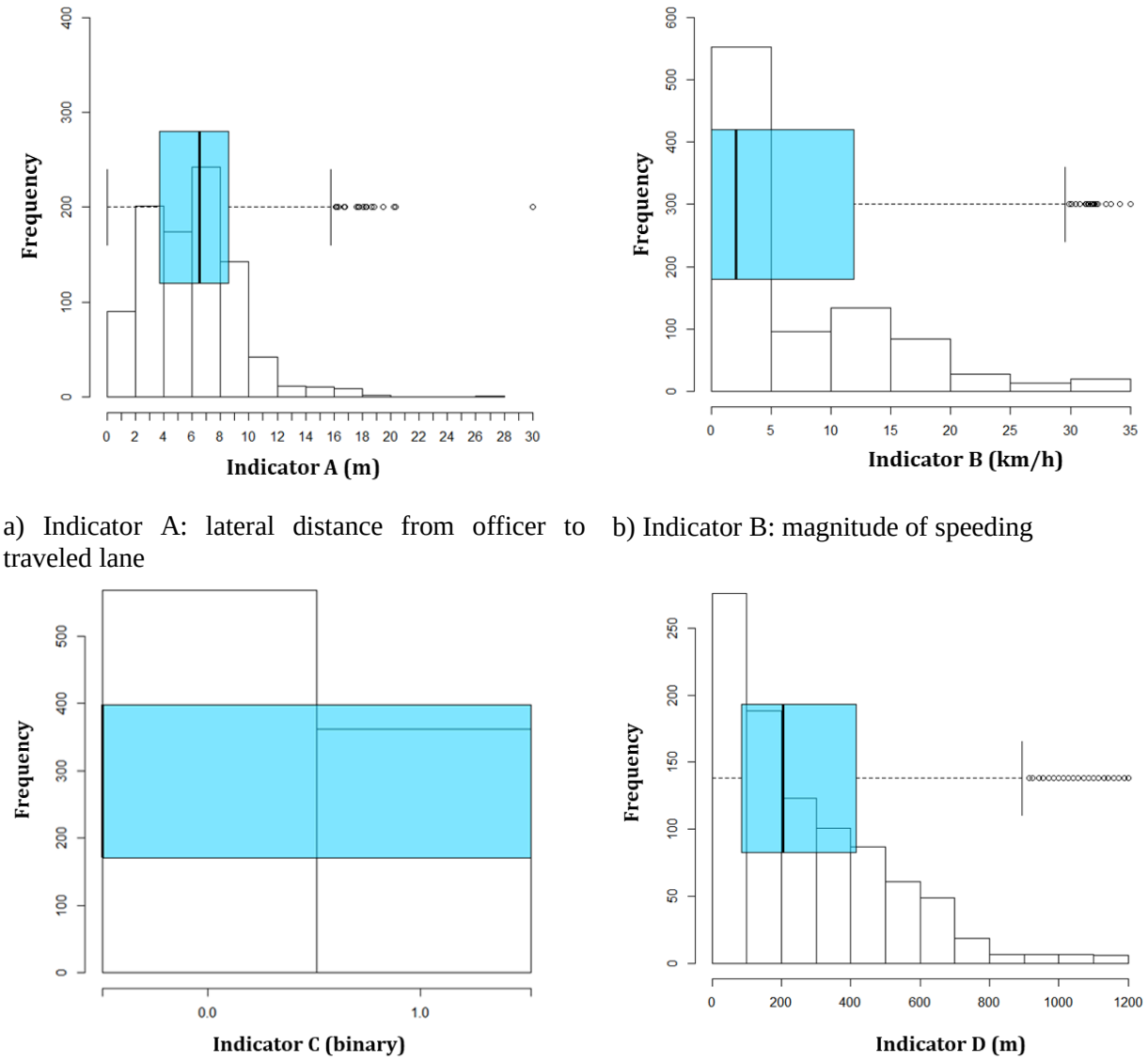
For Indicator C (stopping sight distance (SSD)), the assessment was a binary value and there were no outliers.

For Indicator D (direct distance), a value higher than 900m was considered an outlier. Removing values higher than 900m is logical as this threshold is high enough for a police officer to be safe.

Singh (2020) noted that data normalization could improve the performance of machine learning algorithms. We used linear the Max-Min method which is the most widely used method of normalization (Vafaei et al., 2018). Equation (4-2) shows the equation for the Max-Min method:

$$Normalized\ Measure = \frac{Measure\ value - Min\ value}{Max\ value - Min\ value} \quad (4\text{-}2)$$

where Min Value is the minimum value of the input indicator, and Max Value is the maximum value of the input indicator. Each normalized indicator has a value between zero and one.



a) Indicator A: lateral distance from officer to traveled lane

b) Indicator B: magnitude of speeding

c) Indicator C: stopping sight distance (SSD)

d) Indicator D: direct distance

Figure 4-5: Estimating input indicators and removing outliers

4.3.3. Develop Semi-Supervised and Unsupervised Machine Learning Algorithms

4.3.3.1. Semi-Supervised Machine Learning

A semi-supervised ANN is a cellular information processing system designed and developed on the basis of perceived notions of the human brain and brain's neural system (Firat et al., 2010). The system has flexible mathematical structures that can identify complex nonlinear relationships or patterns.

We developed a semi-supervised ANN that could identify relationships between our four input indicators and the composite risk indicator. Figure 4-6 is divided into six sections to show the multiple processes involved in developing our semi-supervised ANN.

To give labels to the large amount of unlabeled data, we adopted a heuristic approach based on a fuzzy rule-based algorithm. Figure 4-6.a represents our original unlabeled data. The letter under each graph refers to the input indicator. Note that input indicator C is a binary value (safer (0), and riskier (1)). Figure 4-6.b represents our fuzzy rule-based algorithm. A fuzzy rule-based algorithm is based on the concept of fuzzy logic and a rule-based system that can take uncertainty in limited data into account (Dimitriou et al., 2015). The fuzzy logic concept captures uncertainty by defining a "membership degree" that ranges from 0 to 1 rather than the usual "true or false" binary (1 or 0) approach. For example, in terms of input indicator A, highway designers determine a minimum of 0.6m and a maximum of 3.5m for the lateral distance. Fuzzy logic could define two fuzzy sets, one for safer and one for riskier, and assign a membership degree to each fuzzy set. For example, 2.05m (the average value for the lateral distance) can be considered a member of the "riskier" fuzzy set with the membership degree of 0.5, and considered a member of the "safer" fuzzy set with the membership degree of 0.5.

A rule-based system captures limited data availability by condensing the opinions of experts into IF-THEN statements that can be in quantitative of linguistic form. For example, IF "indicator A" is "riskier" and, "indicator B" is "riskier," and "indicator C" is "riskier," and "indicator D" is "riskier,"

THEN “the composite risk indicator (output)” is “very high.” A fuzzy rule-based algorithm has four main stages: fuzzifier, inference engine, rules, and defuzzifier. In the fuzzifier stage, we converted the four input indicators into fuzzy sets.

Each indicator provided two input fuzzy sets (safer and riskier). In the inference engine stage, we determined the shape of the membership functions for each fuzzy set and assigned a membership degree to each input indicator. A trapezoidal shaped function was considered the most suitable function for this study because it can uniformly discriminate between riskier and safer situations. In the rules stage, we defined 16 linguistic IF-THEN rule constructions. For example, we defined fuzzy rule #1 as: if Indicator A is “safer,” and Indicator B is “safer,” and Indicator C is “safer,” and Indicator D is “safer,” then the composite risk indicator (output) is “very low.” In the defuzzifier stage, we converted the fuzzy relative risk outputs to numerical values describing the estimated composite risk indicator. The values of composite risk indicator ranged from 0 (lowest relative risk) to 1 (highest relative risk). Figure 4-6.b represents the fuzzy rule-based algorithm we developed.

Figure 4-6.c shows the three labels assigned to the composite risk indicator: low (0, 0.33), medium (0.33,0.67), and high (0.67, 1). Figure 4-6.d shows the unlabeled data changed to labeled data. The heuristic approach has assigned a low-risk (green dot), medium-risk (blue dot) or high-risk (red dot) label to each unlabeled data point. Figure 4-6.e shows an example of a near miss collision identified by driver braking behavior although only a small amount of labeled data was available in the original dataset. Figure 4-6.f shows the use of the semi-supervised ANN to progress from the four indicators we used in the input stage through a hidden layer to the three collision risk levels of the output layer.

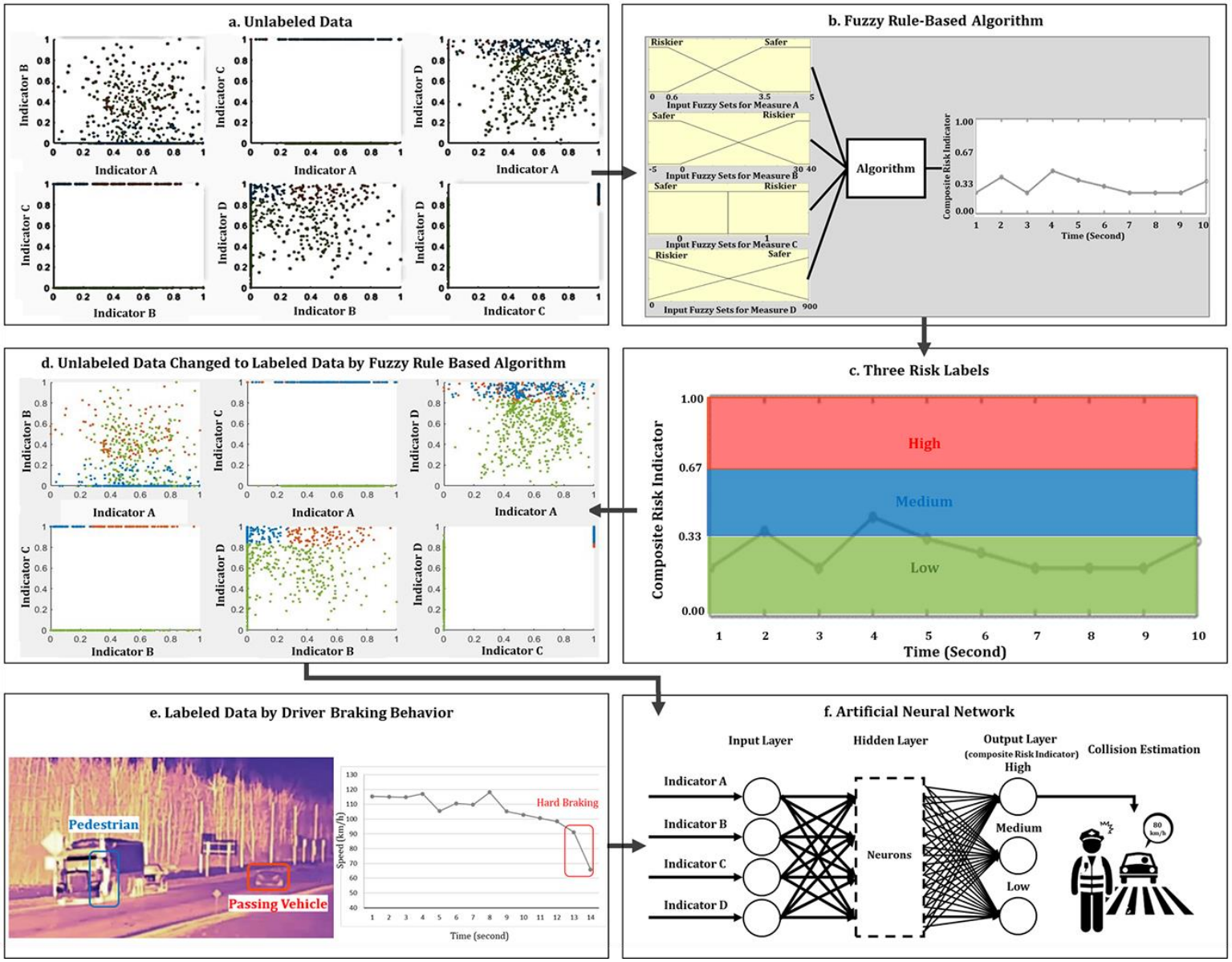


Figure 4-6: Semi-supervised ANN based on driver braking behavior

An ANN is usually composed of the three layers (i.e., stages) shown (Basheer and Hajmeer, 2000) in Figure 4-6.f. The hidden layer includes several neurons and the output layer includes three neurons (as we have three labels). A neuron is a computational component composed of a weighted sum of the input indicators, a bias (similar to the bias in a regression model), and a transfer to the next layer using an activation function. The activation functions can have different shapes (e.g., sigmoid or SoftMax). We needed to determine the activation function for the neurons in the hidden and output layers. Figure 4-7 shows the structure of a neuron in the hidden layer (Figure 4-7a) and in the output layer (Figure 4-7b).

In the hidden layer, we used the sigmoid activation function (as it performs rapidly resulting in less training time (Liu et al., 2018). See Equation (4-3):

$$f(\xi) = 1/(1 + \exp(-\xi)) \tag{4-3}$$

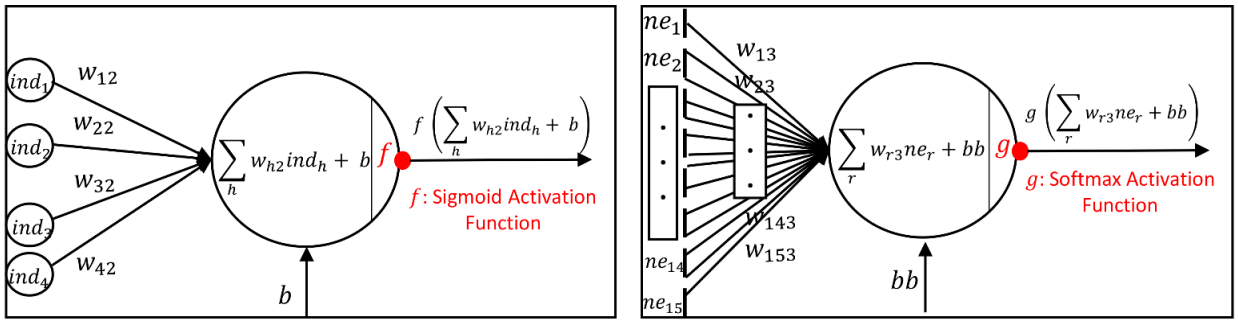
Where $\xi = \sum_h w_{h2} ind_h + b.w_{l2}$, the weight between input indicator *h* and a neuron in the hidden layer; ind_h is the value of input indicator *h*; and b is the bias. In the hidden layer, the neuron received a weighted sum of the four input indicators, a bias (similar to the bias in a regression model), and the sigmoid activation function.

In the output layer, we used the SoftMax activation function (Dunne and Campbell, 1997). This function is usually used when we have a multiple number of labels (e.g., low, medium, and high) and is shown in Equation (4-4):

$$g(\delta) = \exp(\delta) \Big/ \sum_{c=1}^{3} \exp(\delta_c) \tag{4-4}$$

where $\delta = (\sum_r w_{r3} ne_r + bb)$, w_{r3} is the weight applied between the neuron with index *r* in layer 2 (the hidden layer), ne_r is the value of neuron *r*, and bb is the bias.

Figure 4-7.b shows the structure of a neuron in the output layer neuron that used a SoftMax activation the neuron received a weighted sum of the 15 neurons in the hidden layer, a bias (similar to the bias in a regression model) and the SoftMax activation function.



a) Structure of a neuron in hidden layer b) Structure of a neuron in output layer

Figure 4-7: Structure of a neuron in hidden layer and output layer

The learning process in an ANN is equivalent to minimizing cross entropy (the error between the risk label and the estimated risk label). Equation (4-5) shows cross-entropy:

$$Cross\ Entropy\ = -\sum_{c=1}^{3} t_c \log(p_c) \tag{4-5}$$

where t_c is the binary truth state for label c (if the label estimated by the semi-supervised ANN is matched with labeled data, the binary truth state is 1; otherwise, 0), and p_c is the SoftMax probability (the output value obtained from Equation (4-5)). A major concern regarding semi-supervised ANNs is the overfitting problem in which the network seems to improve considerably during training, but becomes worse during the actual application. To overcome the overfitting problem, the total available data are randomly divided into a training set (60%), a validation set (20%) and a test set (20%) (Kocadagli, 2015). The training set is used in the ANN model development, the validation set is used in the determination of the stopping point of the training process to avoid the overfitting problem, and the test set is used to evaluate the estimation ability of the ANN.

In our study, we used a maximum of 100 iterations to train the network, and the training stopped when the ANN's performance failed to improve for six consecutive validation runs which prevent the network begins to overfit the data, memorising the training data rather than to maximize the generalization capability of the ANN (Lei et al., 2014; Slavici et al., 2016; Yang et al., 2019).

In the ANN training process, the weights of a network are initialized by random numbers between -1 and +1. The weights are modified iteratively to minimize the cross-entropy, but the procedure does not guarantee an optimal solution (Kocadagli, 2015). To alleviate this problem, we trained the network 30 times using different initial random weights. We then selected the most desirable solutions for the weights on the basis of evaluation measures such as accuracy.

We determined the number of neurons by analyzing the result. See Section 4.4, Results of Analysis.

4.3.3.2. Unsupervised Machine Learning

We selected Fuzzy C-Means for our unsupervised machine learning algorithm. Fuzzy C-Means is based on fuzzy logic and a distance based objective function. The algorithm discovers significant patterns in the input indicators by classifying the input indicators into labels such that the datapoints assigned to the same label are as similar as possible and each label is as a different as possible (Nayak et al., 2015). Fuzzy C-Means first considers the number of labels (three risk labels in our study) and then randomly assigns the initial label centers (i.e., the initial mean value of the labels). A membership degree is then assigned to each datapoint in each label. For example, a datapoint could belong to a low-risk label with 0.3 membership degree and belong to a medium-risk label with 0.7 membership degree. The label centers and corresponding membership degrees

of each datapoint are updated iteratively by minimizing the objective function of Fuzzy C-Means as shown in Equation (4-6):

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^{m} d_{ij}^{2} \tag{4-6}$$

where U is the matrix of membership degree, V denotes the matrix of label center, n is the total number of datapoints, c is the number of labels which is three (low, medium and high) in our study, m is the fuzziness parameter (uncertainty degree), v_i is the center of label i, μ_{ij} represents the membership degree of the datapoint x_j to label v_i, d_{ij}^2 is the Euclidean distance of the datapoint x_j to the label center v_i, and $d_{ij}^2 = ||x_j - v_i||^2$.

The iterative procedure updates the membership degree v_{ij} and the label centers v_i using Equations (4-7) and (4-8):

$$\mu_{ij} = 1 \Big/ \sum_{k=1}^{c} (d_{ij}/d_{jk})^{2/(m-1)} \tag{4-7}$$

$$v_i = \sum_{j=1}^{n} \mu_{ij}^{m} x_j \Big/ \sum_{j=1}^{n} \mu_{ij}^{m} \tag{4-8}$$

where $\mu_{ij} \epsilon$ [0,1]. The iterative process continues until the objective function of the previous iteration and the current iteration are equal.

A value has to be selected for m, the fuzziness parameter (uncertainty degree). Little theoretical guidance and few generally accepted criteria are currently available to support the selection of m. A range of m based on experience, $1.1 \leq m \leq 5.0$ was proposed by Bezdek (1981). As all input indicators in the Fuzzy C-means algorithm should be continuous values, we changed the binary indicator for input indicator C (SSD) to used continuous values. We determined

the most appropriate parameter of m (the fuzziness parameter) by analyzing the results. See Section 4.4, Results of Analysis.

4.3.4. Propose Evaluation Measures

We proposed evaluation measures by using a confusion matrix to compare the performance of the semi-supervised machine learning algorithm and the unsupervised machine learning algorithm.

Table 4-1 shows a confusion matrix for a three-label (risk) classification problem. The matrix shows how a machine learning algorithm approaches the problem of correctly estimating three risk labels (Ballesteros and Riquelme, 2014).

Table 4-1: A confusion matrix for a three-label classification problem

		Estimated risk labels by machine learning algorithm			**Recall**
		High-risk	**Medium-risk**	**Low-risk**	
Risk label	**High-risk**	a	b	c	$a/(a+b+c)$
	Medium-risk	d	e	f	$e/(d+e+f)$
	Low-risk	g	h	i	$i/(g+h+i)$
Precision		$a/(a+d+g)$	$e/(b+e+h)$	$i/(c+f+i)$	

The confusion matrix shown in Table 4-1 includes the letters "a" to "i." For example, the letter "a" represents the number of high-risk labels that a machine learning algorithm correctly estimated as high risk, the letter "b" shows the number of high-risk labels that the machine learning algorithm incorrectly estimated as medium-risk, and so on.

We used three common evaluation measures to evaluate the performance of machine learning algorithms. The evaluation measures were: accuracy, precision, and recall.

Accuracy is the most widely used evaluation measure (Ballesteros and Riquelme 2014). Accuracy in our study was defined as the ratio of correctly estimated labels to total datapoints. A higher value indicated a better network. The equation for accuracy is shown in Equation (4-9):

$$Accuracy = (a + e + i)/(a + b + c + d + e + f + g + h + i) \quad (4\text{-}9)$$

Precision was defined for each estimated label in our study, and was defined as the ratio of correctly estimated labels to total estimated labels. A higher value indicated a better network.

Recall was defined for each risk label and was the ratio of correctly estimated labels to total risk labels. A higher value indicated a better network.

Precision and recall are usually averaged into an F1 score. A higher value indicates a better network (Ballesteros and Riquelme 2014). See Equation (4-10):

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4\text{-}10)$$

When assessing the performance of the algorithms, the calculation of accuracy considered all the labels together. The calculation of precision, recall, and the F1 score considered each label separately.

4.4. Results of Analysis

4.4.1. Sensitivity Analysis

We conducted a sensitivity analysis on the number of neurons in the semi-supervised ANN algorithm and the parameter m in the unsupervised Fuzzy C-Means algorithms. We developed six algorithms in the sensitivity analysis: (1) ANN (5 neurons), (2) ANN (15 neurons), (3) ANN (25 neurons), (4) Fuzzy C-Means (m=2), (5) Fuzzy C-Means (m=3.5), and (6) Fuzzy C-Means (m=5). We compared the six algorithms by calculating the proposed evaluation measures and selecting the algorithm that performed best.

Table 4-2 shows the evaluation measures for the six algorithms**.** The Table shows the six algorithms, each algorithm's accuracy, and each algorithm's precision, recall and F-1 scores for each of the three risk labels.

Table 4-2: The six algorithms and the performance measures

Algorithms	Specification	Accuracy		High-Risk	Medium-Risk	Low-Risk
Algorithm 1	ANN (neurons = 5)	91%	Precision	89%	93%	94%
			Recall	75%	92%	98%
			F1-Score	81%	92%	96%
Algorithm 2	**ANN (neurons = 15)**	**94%**	Precision	91%	90%	96%
			Recall	74%	93%	98%
			F1-Score	**81%**	**93%**	**97%**
Algorithm 3	ANN (neurons = 25)	90%	Precision	91%	96%	91%
			Recall	81%	82%	95%
			F1-Score	85%	88%	92%
Algorithm 4	Fuzzy C-Means (m=2)	33%	Precision	100%	2%	48%
			Recall	11%	4%	92%
			F1-Score	21%	2%	63%
Algorithm 5	Fuzzy C-Means (m=3.5)	31%	Precision	88%	3%	46%
			Recall	10%	5%	83%
			F1-Score	18%	3%	58%
Algorithm 6	Fuzzy C-Means (m=5)	30%	Precision	91%	1%	44%
			Recall	12%	3%	81%
			F1-Score	21%	1%	56%

The best algorithm was algorithm 2 (ANN 15 neurons) which is shown in bold in Table 4-2. This algorithm had an accuracy of 94%, and F1 score of 81% for the high-risk label, 93% for the medium-risk label, and 97% for the low-risk label. The learning time for the algorithm was 4 seconds which shows rapid learning.

There was a big difference in terms of the evaluation measures for the ANN algorithms (algorithms 1-3) compared to the Fuzzy C-Means algorithms (algorithms 4-6). The more successful algorithms were the semi-supervised ANN machine learning algorithms which considered safety considerations (i.e., by using the fuzzy rule-based algorithm and the identification of near miss collisions). The less successful algorithms were the Fuzzy C-Means

algorithms which simply attempted to classify datapoints on the basis of the similarities between the datapoints (i.e., in terms of distance from center) and did not take safety considerations into account.

The precision of the three Fuzzy C-Means algorithms was high when estimating the high-risk labels, but far less successful when estimating the low-risk labels, and very low when estimating the medium-risk labels. This means that the three Fuzzy C-means algorithms tended to estimate many events as high-risk. Estimating many events as high-risk would generate a warning for a large number of events resulting in police officers possibly coming to neglect or ignore the system's warnings.

Figure 4-8 shows the cross entropy (error) curve of the best algorithm (algorithm 2) for the training, validation, and test data. The Figure shows that gradual learning (i.e., minimizing the error) and improvement occurred over 35 iterations.

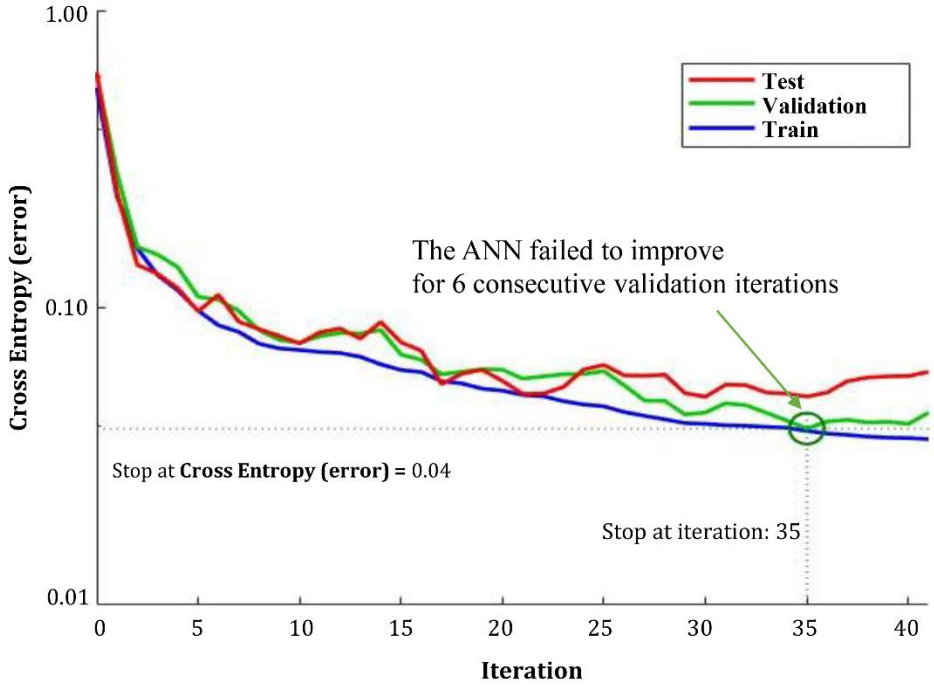
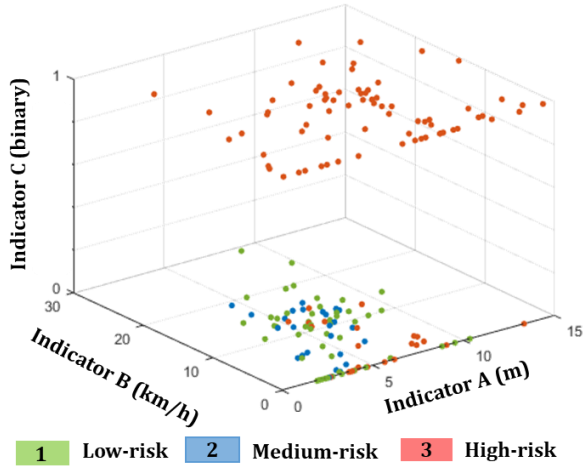


Figure 4-8: Cross entropy (error) of the best algorithm (ANN with 15 neurons)

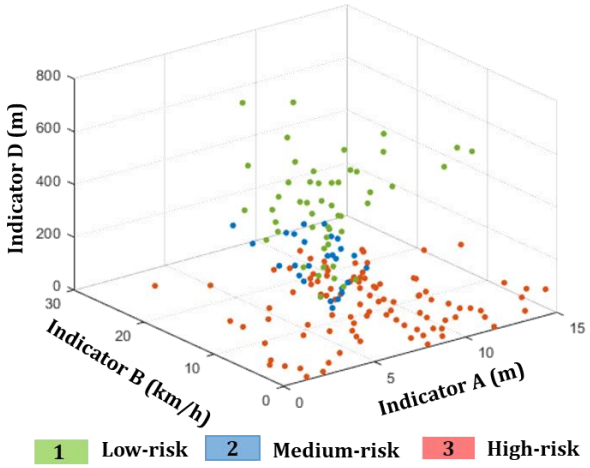
The training stopped on iteration 35 when using the validation data. At this point, the lowest error for the validation data was obtained (0.04), and the algorithm's performance failed to improve for six consecutive iterations. The gap between the curves for the testing, validation and training data was not large indicating that the best algorithm did not have an overfitting problem (Kocadagli, 2015).

Figure 4-9 provides four three-dimensional (3-D) representations of the results showing how the data points were classified into low, medium and high-risk labels by indicators A, B, C, and D by one of the unsupervised Fuzzy C-Means algorithms (algorithm 4) and by the best performing semi-supervised machine learning algorithm, (algorithm 2).

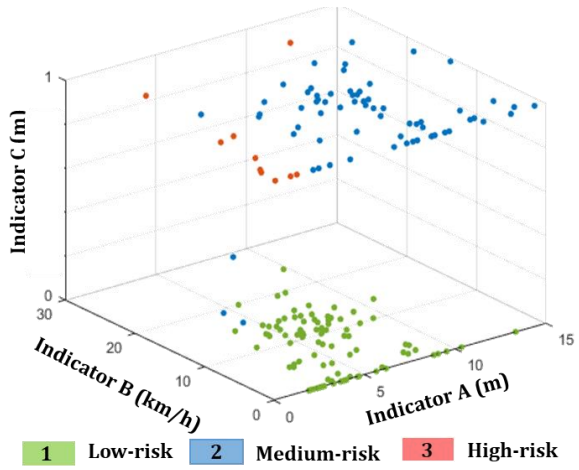
The Fuzzy C-Means algorithm (Figure 4-9.a and Figure 4-9.b) tried to identify patterns in the datapoints and classified the datapoints on the basis of similarities between the datapoints. Many datapoints were classified as high-risk (the red points). The Fuzzy C-Means algorithm incorrectly estimated many low-risk and medium-risk datapoints as high-risk resulting a score of 11% for recall, but this algorithm scored 100% for precision (see algorithm 4 in Table 4-2). These results indicate that the high-risk labels estimated by the algorithm included all the genuinely risky situations such as the near miss collisions. The results also showed that datapoints were assigned a high-risk label when there was an indication of risk because the direct distance was small (Indicator D was less than 200 m) and SSD was inadequate (Indicator C had a value of one). The values of Indicators A (lateral distance from officer to traveled lane) and B (magnitude of speeding), however, could vary indicating that the algorithm did not associate a decrease in Indicator A (pedestrian behavior measured as the lateral distance from officer to traveled lane) with an increase in risk, and did not associate an increase in Indicator B aggressive driving measured by the magnitude of speeding) with an increase in risk.



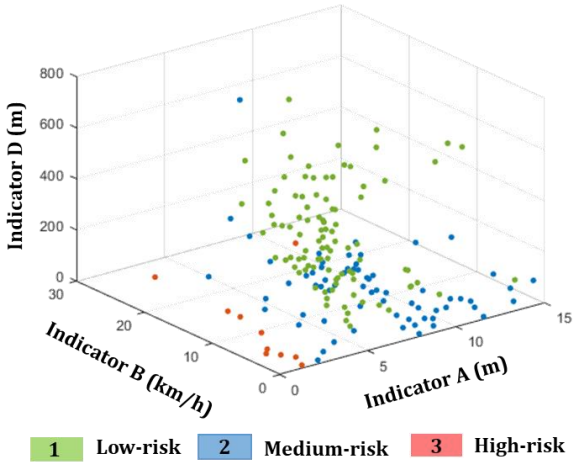
a) Classification by unsupervised Fuzzy C-Means showing on Indicators A, B, and C



b) Classification by unsupervised Fuzzy C-Means showing on Indicators A, B, and D



c) Classification by semi-supervised ANN showing on Indicators A, B, and C



d) Classification by semi-supervised ANN showing on Indicators A, B, and D

Figure 4-9: Examples of classification of data points into low, medium and high-risk by an unsupervised Fuzzy C-Means algorithm (algorithm 4) and by the best algorithm, a semi-supervised ANN (algorithm 2)

The results for the semi-supervised ANN were very different. For example, Figure 4-9.c and Figure 4-9.d show that almost all the datapoints with a high-risk label (the red points) refer to situations in which a pedestrian was close to the traveled lane (Indicator A was less than 3.6 m), the magnitude of speeding was positive (Indicator B > 0), SSD (Indicator C) implied risk (a value

of one), and Indicator D showed that the pedestrian and vehicle were very close to each other (Indicator D < 100 m apart).

The results clearly show that the semi-supervised ANN could capture pedestrian behavior (Indicator A), driver behavior (Indicator B), and spatial and temporal interactions between pedestrian and vehicle (Indicators C and D).

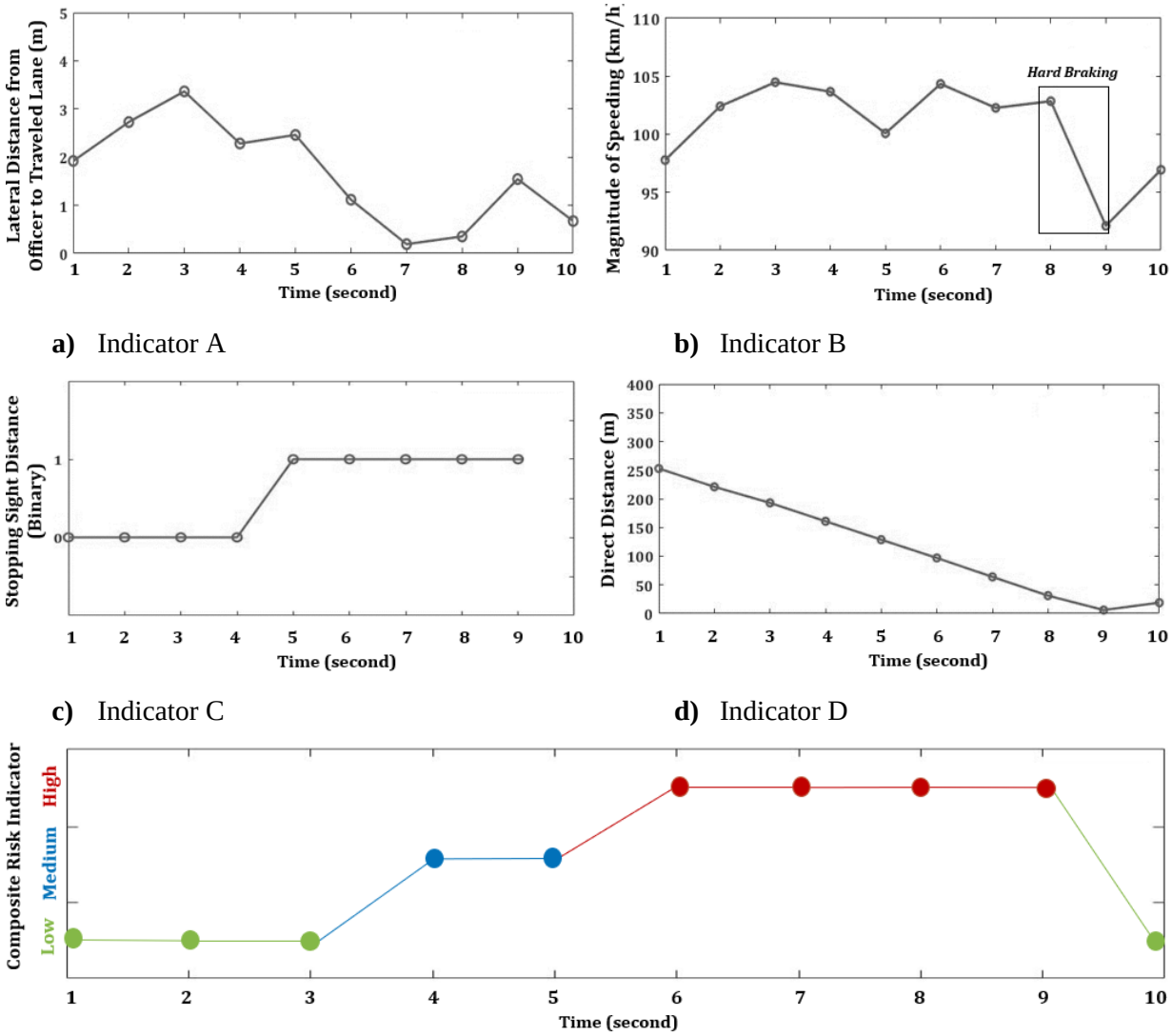
The number of datapoints assigned a high-risk label by the semi-supervised ANN was significantly less that the number of datapoints assigned a high-risk label by the Fuzzy C-Means algorithm. This is because the semi-supervised ANN was only alerted to warn police officers when there was a high-risk event (e.g., a near miss collision).

4.4.2. Performance of Best Semi-Supervised ANN

The study included using a real-world dataset to evaluate the performance of the semi-supervised machine learning algorithm. For the semi-supervised machine learning algorithm, we used the best performing algorithm (algorithm 2). We randomly selected a near miss collision event which included hard braking by the driver and analyzed how algorithm 2 used the four indicators to estimate the event's risk label over a period of 10 consecutive seconds. The hard braking occurred at second 8. Figure 4-10 shows the data for the four input indicators for the 10 seconds of the event selected and the combination of the four input indicators into the estimated composite risk indicator.

Indicator A (lateral distance from officer to traveled lane). Figure 4-10a shows that the pedestrian was positioned between 0.20 m and 3.30 m from the shoulder-side traveled lane for the entire 10 seconds. At second 8, Indicator A was only 0.35 m. Indicator A is expected to have contributed to an increase in risk when the pedestrian moved towards the traveled lane.

Indicator B (magnitude of speeding). Figure 4-10b shows that the magnitude of speeding had a positive value (i.e., the vehicle's speed exceeded 100 km/h) until the moment of hard braking (i.e., second 8). Just after this moment, the magnitude of speeding became zero. Indicator B can be expected to have contributed to an increase in risk for the seconds before second 8.



a) Indicator A **b)** Indicator B

c) Indicator C **d)** Indicator D

e) Composite risk indicator (estimated risk label)

Figure 4-10: Analysis of algorithm 2's combination of the four input indicators into the estimated composite risk indicator (risk label)

Indicator C (stopping sight distance). Figure 4-10c shows that the SSD increased from 0 for the first 4 seconds to 1 from second 5 onwards. The value of 1 for SSD means that the vehicle could not stop in time if the pedestrian moved into the traveled lane or if the driver lost control of his vehicle. Indicator C can be expected to have contributed to an increase in risk from second 5.

Indicator D (direct distance) Figure 4-10d shows that the direct distance between the pedestrian and the approaching vehicle decreased until second 9 and then increased after the vehicle had passed the pedestrian. Indicator D is expected to have contributed to a gradually increasing level of risk as the vehicle came closer to the pedestrian up to second 9 (at which point the direct distance was 15 m) and then to have contributed to a decrease in the level of risk after passing the pedestrian and traveling away.

Figure 4-10e shows the risk labels estimated by algorithm 2 during the 10 seconds. As expected, the risk label was low (green) for the first 3 seconds while the approaching vehicle was not close to the pedestrian (Indicator D, direct distance, was 186 m at second 3). At second 3, a warning that a vehicle was approaching was not required as a warning would distract the police officer for no real reason.

The risk label changed from low to medium (blue) at second 4 when the approaching was closer (Indicator D, direct distance, was 186 m at second 3, but 156 m at second 4). At second 5, the SSD increased from 0 to 1 indicating that the vehicle could not stop in time if the pedestrian moved into the traveled lane or if the driver lost control of his vehicle. At second 5, the algorithm continued to assign a medium-risk label as the pedestrian was not close to the traveled lane (Indicator A, lateral distance from officer to traveled lane, was 2.5 m) and the direct distance remained adequate (Indicator D, direct distance, was 125 m at second 5).

At second 6, the risk label changed from medium to high (red). This was when the

pedestrian moved toward the traveled lane (Indicator A, lateral distance from officer to traveled lane, decreased from 2.5 m in second 5 to 1.1 m to second 6), and the vehicle's driver had a positive magnitude of speeding (105 km/h), but did not change his behavior. The algorithm was able to identify a high-risk driver two seconds prior to a hard brake. At second 7, the pedestrian was 0.2 m from the traveled lane, the driver had a magnitude of speeding of 3 km/h, and the direct distance was only 47 m. From second 8 to second 9, the driver applied a hard brake decreasing the vehicle's speed from 103 km/h in second 8 to 92 km/h in second 9. The risk label remined high from second 6 to second 9.

At second 10, the risk label is shown as low although the algorithm assigned no risk to the situation. The driver had passed the pedestrian and the algorithm responded immediately.

If we consider second 9 as the second of a potential collision, the algorithm could identify an approaching vehicle as high-risk at second 6. At second 6, the direct distance (Indicator D) was 99 m. Figure 4-10 shows that the ANN could give a warning three to four seconds before a potential collision.

A warning given three to four seconds before a potential collision sounds appropriate. For example, Hines et al. (2015) analyzed collisions that involved pedestrians who were working on roadways and suggested that 2.5s provided enough early warning time to avoid a collision. Lin et al. (2016) developed a pedestrian protection system (''pSafety'') which provided a warning 3s before a potential collision.

4.4.2. Performance of Semi-Supervised ANN when Identifying Near Miss Collisions

The semi-supervised ANN had a high score for accuracy (94%) when estimating all three risk labels, and performed well when identifying near miss collisions. Our test data included nine

datapoints for near miss collisions. The semi-supervised ANN identified eight of the nine (88%) near miss collisions correctly which shows that the algorithm performed well for validity.

This level of validity compares well with the results of other studies that validated a machine learning algorithm's assessment of the risk of a collision. For instance, Wang et al. (2015) applied a K-Means clustering approach to analyze the risk of collisions and detected 66% of near miss collisions. Kluger et al. (2016) used the discrete Fourier transform algorithm combined with a K-Means clustering approach. Their study detected 78% of near miss collisions. Xiong et al. proposed a fuzzy rule-based algorithm combined with K-Means clustering and detected 87% of high-risk situations (2019).

4.5. Study Conclusions, Limitations and Recommendations

The goal of this study was to develop a semi-supervised machine learning algorithm that can produce the risk analysis required for an advanced smart protection system that provides police officers working on foot on a roadway with an early warning that they are at risk being hit by a vehicle. A police officer working as a pedestrian on a roadway may be distracted by numerous duties which may prevent the officer from paying attention to a risk brought about by a passing vehicle.

Smart protection systems that provide a warning of possible risk from a passing vehicle are available for pedestrians working on roadways, but these systems are based solely on a single input indicator (for example, intrusion of work zone/safety area by a passing vehicle). The systems simply apply a fixed threshold to the single input indicator to differentiate a safe event from a risky event and cannot capture the variety of situations that may create a risk for a police officer. A more sophisticated approach that could combine more than one input indicator is clearly needed.

The first objective was to find an approach to label a portion of the data in a dataset of unlabeled data. This approach identifies near miss collisions which can be used as small number of labeled data in the risk analysis. No existing dataset for collisions between pedestrian roadway workers and vehicles provides data for pedestrian and vehicle trajectories in the seconds leading up to a collision, and as the system we propose has not yet been implemented, we cannot use our system to obtain the data. Our system was developed by collecting a large amount of data from a radar installation which included data for vehicle drivers who were walking along the road or shoulder lane (similar activities for police officers). We successfully used driver braking behavior to detect probable near miss collisions in the dataset. As the data for these near miss collisions were associated with high-risk, we considered the data for near miss collisions to be labeled data. The rest of the data obtained from the radar system remained unlabeled at this stage, i.e., the level of risk was completely unknown.

The second objective was to develop a semi-supervised machine learning algorithm that combines labeled and unlabeled data and estimates levels of risk into three categories (low, medium and high). As the labeled data were insufficient for algorithm development or testing, we used a heuristic approach based on a fuzzy rule-based algorithm to give labels to the unlabeled data in our dataset. We developed three semi-supervised ANN algorithms. By developing our semi-supervised machine learning algorithms, we were able to create a composite risk indicator with the three risk categories (low, medium and high).

The third objective was to apply an unsupervised machine learning that classifies data based on their similarities into three risk categories (low, medium and high). For this purpose, we developed three unsupervised Fuzzy C-Means algorithms. In addition to the three semi-supervised ANN algorithms, we developed three unsupervised Fuzzy C-Means algorithms. We evaluated and

compared the performance of the six machine learning algorithms. This objective used a real-world dataset.

The results showed that the best semi-supervised ANN algorithm could combine four input indicators, classify risk into low, medium and high, and successfully capture police officer behavior, driver behavior, and spatial/temporal interactions between pedestrians and vehicles. It showed a high level of performance, i.e., 94% accurate when estimating the three risk labels, and 88% successful when identifying near miss collisions. The Fuzzy C-Means algorithms did not perform well. They failed to simultaneously capture the four selected collision contributing factors, and the best Fuzzy C-Means algorithm was only 33% accurate when estimating the three risk labels. The semi-supervised ANNs gave only a small number of datapoints a high-risk label, i.e., the semi-supervised ANNs only warned warn police officers when there was a high-risk event (e.g., a near miss collision). The Fuzzy C-Means algorithms tended to give many datapoints a high-risk label, i.e., they generated a large number of warnings which could lead police officers to stop taking notice of the alarms. The best semi-supervised ANN algorithm showed 88% success to detect near miss collisions. It means that the algorithm is able to detect a near miss collision (the moment of the hard braking and 2 seconds before the hard brake) with 88% success rate. Therefore, the algorithm could generate a warning three to four seconds before a collision. This warning period is in line with the literature and provides an acceptable amount of time for a police officer react to potential collision risk.

We noted the following limitations in our study and make the following recommendations for future research:

- The proposed semi-supervised machine learning algorithm was based on a small numbers of near miss collisions identified by driver braking behavior. Other evasive actions

conducted by a driver (e.g., fast steering to avoid a collision) or by a police officer (e.g., rapidly changing position to avoid collision) could also provide appropriate means for identifying near miss collisions. Future research could consider different ways to collect data on these evasive actions to identify additional near miss collisions and improve the accuracy of the algorithm.

- We selected the most efficient functions/optimization algorithms for the proposed semi-supervised ANN. These algorithms showed rapid learning (the learning time was less than 5 seconds for all of our algorithms), but the time required for training may increase as the number of datapoints increases, and this time increase might delay a warning of a possible collision. Further improvement of the training time is not possible unless we collect a larger number of datasets. A future study could use naturalistic driving or virtual reality technology to simulate real-world situations and collect the large amount of data required to improve training by designing meta-heuristic algorithms such as genetic or particle swarm optimization algorithms for use in the semi-supervised ANN learning process.
- We developed and compared one semi-supervised machine learning algorithm with another unsupervised machine learning algorithm. Future studies may consider the possibility of developing other semi-supervised machine learning algorithms (e.g., Random Decision Forest and Linear Regression) and other unsupervised machine learning algorithms (K-Nearest Neighbor and K-Means) to investigate which risk analysis algorithm is most appropriate for use in a smart protection system designed to protect police officers working as pedestrians on highways.

4.6. Chapter 4 References

Arun, A., Haque, M. M., Washington, S., Sayed, T., and Mannering, F. (2021). A systematic review of traffic conflict-based safety measures with a focus on application context. Analytic methods in accident research, 32, 100185.

American Association of State Highway and Transportation Officials (AASHTO) (2018). A Policy on Geometric Design of Highway and Streets. AASHTO, Washington, D.C.

Baek, J., and Choi, Y. (2018). Bluetooth-beacon-based underground proximity warning system for preventing collisions inside tunnels. Applied Sciences, 8(11), 2271.

Bagdadi, O. (2013). Assessing safety critical braking events in naturalistic driving studies. Transportation research part F: Traffic Psychology and Behaviour, 16, 117-126.

Ballesteros, A. J. T., and Riquelme, J. C. (2014). Data mining methods applied to a digital forensics task for supervised machine learning. In Computational Intelligence in Digital Forensics: Forensic Investigation and Applications (pp. 413-428). Springer, Cham.

Basheer, I. A., and Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods, 43(1), 3-31.

Bezdek J C. (1981). Pattern recognition with fuzzy objective function algorithms. New York: Plenum Press,

Castignani, G., Derrmann, T., Frank, R., and Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. IEEE Intelligent transportation systems magazine, 7(1), 91-102.

Chen, F., Wang, J., and Deng, Y. (2015). Road safety risk evaluation by means of improved entropy TOPSIS–RSR. Safety science, 79, 39-54.

Coifman, B., and Kim, S. (2009). Speed estimation and length-based vehicle classification from freeway single-loop detectors. Transportation research part C: emerging technologies, 17(4), 349-364.

Confucian. (2022). "Traffic Accident Warning System" Available at: https://en.confucian.com.tw/autopage_detail/4/23 [Accessed June, 2022].

Cusick, M., Adekkanattu, P., Campion Jr, T. R., Sholle, E. T., Myers, A., Banerjee, S., and Pathak, J. (2021). Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. Journal of psychiatric research, 136, 95-102.

Dimitriou, L., and Vlahogianni, E. I. (2015). Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions. Analytic Methods in Accident Research, 5, 59-71.

Dimitriou, L., Stylianou, K., and Abdel-Aty, M. A. (2018). Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions. Accident Analysis & Prevention, 118, 221-235.

Dunn, J.C., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, Journal of Cybernetics, 3:3, 32-57.

Dunne, R. A., and Campbell, N. A. (1997). On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function. In Proc. 8th Aust. Conf. on the Neural Networks, Melbourne (Vol. 181, p. 185).

El-Basyouny, K., and Sayed, T. (2013). Safety performance functions using traffic conflicts. Safety science, 51(1), 160-164

Federal Bureau of Investigation (FBI) (2021). the Law Enforcement Officers Killed and Assaulted (LEOKA) Program; U.S. Department of Justice: Washington, DC, USA, 2021, Available in: https://crime-data-explorer.app.cloud.gov/pages/downloads [Accessed June, 2022]

Firat, M., Turan, M. E., and Yurdusev, M. A. (2010). Comparative analysis of neural network techniques for predicting water consumption time series. Journal of hydrology, 384(1-2), 46-51.

Fountas, G., Pantangi, S. S., Hulme, K. F., and Anastasopoulos, P. C. (2019). The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: A correlated grouped random parameters bivariate probit approach. Analytic methods in accident research, 22, 100091.

Gu, X., Abdel-Aty, M., Xiang, Q., Cai, Q., and Yuan, J. (2019). Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. Accident Analysis & Prevention, 123, 159-169.

Hines, K., Lages, W., Somasundaram, N., and Martin, T. (2015). Protecting workers with smart e-vest. In Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (pp. 101-104).

Johnsson, C., Laureshyn, A., and De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. Transport Reviews, 38(6), 765-785.

Kluger, R., Smith, B. L., Park, H., and Dailey, D. J. (2016). Identification of safety-critical events using kinematic vehicle data and the discrete fourier transform. Accident Analysis & Prevention, 96, 162-168.

Kocadagli, O. (2015). A novel hybrid learning algorithm for full Bayesian approach of artificial neural networks. Applied Soft Computing, 35, 52-65.

Laureshyn, A., de Goede, M., Saunier, N., and Fyhri, A. (2017). Cross-comparison of three surrogate safety methods to diagnose cyclist safety problems at intersections in Norway. Accident Analysis & Prevention, 105, 11-20.

Lei, Z., Zhenpo, W., Xiaosong, H., & Dorrell, D. G. (2014). Residual capacity estimation for ultracapacitors in electric vehicles using artificial neural network. IFAC Proceedings Volumes, 47(3), 3899-3904.

Li, X., Guo, Z., and Li, Y. (2022). Driver operational level identification of driving risk and graded time-based alarm under near-crash conditions: A driving simulator study. Accident Analysis & Prevention, 166, 106544.

Lin, C. H., Chen, Y. T., Chen, J. J., Shih, W. C., and Chen, W. T. (2016). Psafety: A collision prevention system for pedestrians using smartphone. In 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall) (pp. 1-5). IEEE.

Liu, J., Boyle, L. N., and Banerjee, A. G. (2018). Predicting interstate motor carrier crash rate level using classification models. Accident Analysis & Prevention, 120, 211-218.

Mahmud, S. S., Ferreira, L., Hoque, M. S., and Tavassoli, A. (2019). Micro-simulation modelling for traffic safety: A review and potential application to heterogeneous traffic environment. IATSS research, 43(1), 27-36.

Mendel, J. M. Fuzzy Logic Systems for Engineering: A Tutorial. Proceedings of the IEEE, Vol. 83, No. 3, 1995, pp. 345–377.

Mohammadi, A., Park, P. Y., Asgary, A., Mukherjee, A., and Liu, X. (2022). Internet-of-Things system to protect police officers from collisions while on duty on the roadway. Transportation Research Record, DOI: 03611981221076848.

Mollenhauer, M. A., White, E., and Roofigari-Esfahan, N. (2019). Design and evaluation of a connected work zone hazard detection and communication system for connected and automated vehicles (CAVs).

Montgomery, B. (2019) "Improving Officer Safety on the Roadways," National Institute of Justice, Issue number 280, Available at: https://nij.ojp.gov/topics/articles/improving-officer-safety-roadways [Accessed June, 2022]

Moon, S., Moon, I., and Yi, K. (2009). Design, tuning, and evaluation of a full-range adaptive cruise control system with collision avoidance. Control Engineering Practice, 17(4), 442-455.

Mukherjee, A., Stolpner, S., Liu, X., Vrenozaj, U., Fei, C., and Sinha, A. (2013). Large animal detection and continuous traffic monitoring on highways. In Sensors, 2013 IEEE (pp. 1-3).

Nadimi, N., Behbahani, H., and Shahbazi, H. (2016). Calibration and validation of a new time-based surrogate safety measure using fuzzy inference system. Journal of traffic and transportation engineering (English edition), 3(1), 51-58.

Nayak, J., Naik, B., and Behera, H. (2015). Fuzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014. Computational intelligence in data mining-volume 2, 133-149.

National Highway Traffic Safety Administration (NHSTA) (2020), Fatality Analysis Reporting System (FARS), U.S. Department of Transportation, Available at: https://www-fars.nhtsa.dot.gov/People/PeopleAllVictims.aspx [Accessed June, 2022]

Sacchi, E., Sayed, T., and Deleur, P. (2013). A comparison of collision-based and conflict-based safety evaluations: The case of right-turn smart channels. Accident Analysis & Prevention, 59, 260-266

Singh, D., and Singh, B. (2020). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97, 105524.

Slavici, T., Maris, S., & Pirtea, M. (2016). Usage of artificial neural networks for optimal bankruptcy forecasting. Case study: Eastern European small manufacturing enterprises. Quality & Quantity, 50(1), 385-398.

Song, X., Yin, Y., Cao, H., Zhao, S., Li, M., and Yi, B. (2021). The mediating effect of driver characteristics on risky driving behaviors moderated by gender, and the classification model of driver's driving risk. Accident Analysis & Prevention, 153, 106038.

Tageldin, A., and Sayed, T. (2019). Models to evaluate the severity of pedestrian-vehicle conflicts in five cities. Transportmetrica A: transport science, 15(2), 354-375.

Tang, J., Zheng, L., Han, C., Yin, W., Zhang, Y., Zou, Y., and Huang, H. (2020). Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review. Analytic Methods in Accident Research, 27, 100123.

Terzi, R., Sagiroglu, S., and Demirezen, M. U. (2018). Big data perspective for driver/driving behavior. IEEE Intelligent Transportation Systems Magazine, 12(2), 20-35.

Tofighi, M., Asgary, A., Tofighi, G., Podloski, B., Cronemberger, F., Mukherjee, A., and Liu, X. (2021). Applying Machine Learning Models to First Responder Collisions Beside Roads: Insights from “Two Vehicles Hit a Parked Motor Vehicle” Data. Applied Sciences, 11(23), 11198.

US Department of Transportation (2020). Manual on Uniform Traffic Control Devices; for Streets and Highways. US Department of Transportation, Federal Highway Administration, Washington, D.C.

Vogel, K. 2003. Modelling driver behavior-A control theory-based approach. Doctoral thesis Dissertation, University of Linkoping.

Wang, J., Zheng, Y., Li, X., Yu, C., Kodaka, K., and Li, K. (2015). Driving risk assessment using near-crash database through data mining of tree-based model. Accident Analysis & Prevention, 84, 54-64.

Wang, X., Liu, J., Qiu, T., Mu, C., Chen, C., and Zhou, P. (2020). A real-time collision prediction mechanism with deep learning for intelligent transportation system. IEEE transactions on vehicular technology, 69(9), 9497-9508.

Xing, L., He, J., Abdel-Aty, M., Cai, Q., Li, Y., and Zheng, O. (2019). Examining traffic conflicts of up stream toll plaza area using vehicles’ trajectory data. Accident Analysis & Prevention, 125, 174-187.

Xiong, X., Wang, M., Cai, Y., Chen, L., Farah, H., and Hagenzieker, M. (2019). A forward collision avoidance algorithm based on driver braking behavior. Accident Analysis & Prevention, 129, 30-43.

Yang, X. I. A., Zafar, S., Wang, J. X., & Xiao, H. (2019). Predictive large-eddy-simulation wall modeling via physics-informed neural networks. Physical Review Fluids, 4(3), 034602.

Zheng, L., Sayed, T., and Mannering, F. (2021). Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. Analytic methods in accident research, 29, 100142.

CHAPTER FIVE

5. DEVELOP A RISK ANALYSIS APPROACH TO EVALUATE PHYSICAL DISTANCING ON URBAN SIDEWALKS

Abstract: The COVID-19 outbreak has fundamentally changed how people move around and live in a city. It is reported that the virus spreads through transmission via large droplets and/or small aerosols when an infected individual coughs or sneezes or even breathes or talks near an uninfected individual. This study considers the implications of physical distancing requirements for pedestrian-to-pedestrian conflict using urban sidewalks. Pedestrian physical distancing can be described as a mobility intervention designed to limit virus infection by keeping people, for example, 2 meters apart. Physical distancing assumes that the risk of transmission declines with the distance between people, but a more sophisticated risk analysis algorithm than the current binary algorithm which is simply based on a single distance may help health agencies to evaluate and monitor how well physical distancing is being achieved. The study developed pedestrian physical distancing indicators to quantitatively evaluate different levels of physical distancing and proposed levels of pedestrian physical distancing that can be used to select and implement appropriate mobility interventions. This study also developed a fuzzy rule-based algorithm to estimate the relative risk of viral transmission between pedestrians under different pedestrian walking conditions. Finally, we demonstrated the application of the proposed study in a simulated virtual walkway environment using PTV Viswalk. We hope the study can help decision-makers in health agencies to select the most appropriate pedestrian mobility interventions for limiting spread of the virus on urban sidewalks.

Keywords: *Pandemic, Levels of pedestrian physical distancing, Mobility interventions, Pedestrian-to-pedestrian conflict, Risk analysis algorithm*

5.1. Introduction

As of June, 2022, the COVID-19 virus has infected almost 540 million people around the world and killed nearly 6.3 million people. The case and death toll are still growing (WHO, 2022). Interventions implemented to limit the pandemic include individual level interventions such as mask-wearing and hand-washing with soap or sanitizer, and community level interventions such as education, restricting public gatherings, maintaining physical distancing, and enforcement. To reduce the possibility of COVID-19 spread, public health agencies strongly recommend that people avoid crowded areas and maintain physical distancing (WHO, 2020; ECDC, 2020; Chen et al., 2020; Fong et al., 2020; Zhang et al, 2020a; Varotsos and Krapivin, 2020; Huynh, 2020).

This study proposes a measure to assess the level of pedestrian physical distancing (LPPD) on urban sidewalks during a pandemic. We also develop a risk analysis algorithm using a fuzzy rule-based algorithm (hereafter fuzzy algorithm) to help decision-makers understand how the risk of COVID-19 transmission between pedestrians may be affected by physical distancing between pedestrians.

It is reported that the virus spreads when an infected individual coughs or sneezes or even breathes or talks near an uninfected individual (Leung et al., 2020; Stadnytskyi et al., 2020). The virus is transmitted via large droplets and/or airborne transmission via small aerosols (Zhang et al., 2020b). If the infected person is not wearing a mask, transmission via large droplets typically occurs over a distance of up to 2 meters, but transmission via small aerosols can occur over a distance of up to 8 meters under certain temperature and moisture conditions (Bourouiba, 2020; Parshina-Kottas et al., 2020). Physical distancing assumes that the risk of transmission declines with the distance between people. Policies are implemented largely to avoid virus transmission via large droplets, but public health agencies apply different physical distancing thresholds. For

example, the United Kingdom, the United States and Canada use 2 meters (6 ft), but Germany and Australia use 1.5 meters, and France and Demark use 1.0 meter (Shukman, 2020). Some health scientists have expressed concern about using a fixed value and have pointed out the wide range of circumstances that affect the distance threshold for COVID-19 transmission (Chu et al., 2020; Feng et a., 2020). Exposure time also plays an important role in the possibility of viral transmission. Increased exposure time increases the risk of virus transmission (Vuorinen et al., 2020; CDC, 2020; Parshina-Kottas et al., 2020). When modeling pedestrian physical distancing on urban streets, we can reasonably assume that pedestrian density and pedestrian exposure time are the two most important indicators affecting pedestrians' risk of contracting COVID-19.

5.1.1. Pedestrian Physical Distancing as a Mobility Intervention

Pedestrian physical distancing can be described as a "mobility intervention" designed to keep people apart. Public health authorities in North America have introduced several mobility interventions that support pedestrian physical distancing. The idea of "Slow Streets" is one of the most well-known programs in North America (NACTO, 2020; SFMTA, 2020; CoT, 2020). The Slow Streets program attempts to provide additional space for active transportation, i.e., pedestrians and bicyclists, by increasing the surface infrastructure capacity available. For example, the program might temporarily or permanently expand the space available to pedestrians and bicyclists by replacing curbside parking spaces or curbside lanes with pedestrian walkways and/or bicycle lanes. The Slow Streets program also suggests even stronger mobility interventions such as completely prohibiting non-essential vehicular traffic on designated urban streets to provide significantly more space for active transportation. Many North American municipalities have implemented similar programs, for example, Slow Streets (Denver), ActiveTo (Toronto), Open Streets (Brooklyn), and Stay Healthy Streets (Seattle) (NACTO, 2020). For trails, beaches and

public spaces where the lack of space is not necessarily a major issue, public health agencies typically encourage physical distancing through other types of mobility intervention. These include education (e.g., social media advertisements and outdoor signage) and enforcement (e.g., tickets and fines for physical distancing violators) (de Bruin et al., 2020).

Increasing sidewalk capacity does not always guarantee continuous physical distancing over time among pedestrians. If pedestrians violate physical distancing requirements (whether intentionally or not), it is clear that mobility interventions designed to provide additional capacity cannot be regarded as sustainable solutions to the problem and we need a better understanding of pedestrian space and the level of risk implied by pedestrians failing to maintain physical distance.

5.1.2. Understanding Levels of Pedestrian Physical Distance

We assume that the risk of viral infection increases with higher density and/or a lower rate of compliance with physical distancing guideline, but the distance between pedestrians is constantly changing. We need to understand how these (second by second) spatial and temporal changes affect the interventions we use to reduce the risk of virus transmission. Figure 5-1 provides a graphical example of how pedestrian density and compliance rate (inverse of exposure) are constantly changing.

To demonstrate the implications of differences in pedestrian density and compliance rate associated with the physical distancing, we developed four scenarios, one for each box with a colored border in Figure 5-1:

- In Scenario 1 (purple box), each pedestrian has enough walking space to maintain the exact physical distancing required (e.g., 2 meters). In this scenario, the space is operating at capacity, i.e., the whole space is being occupied by the maximum number of pedestrians

who can maintain physical distancing from each other.

- In Scenario 2 (green box), each pedestrian has more than enough walking space (e.g., 8 meters) to maintain physical distancing.
- In Scenario 3 (blue box), some pedestrians cannot maintain physical distancing as pedestrian volume has exceeded the capacity of the space.
- In Scenario 4 (pink box), some pedestrians do not maintain the required physical distancing although the space has more than enough space available.

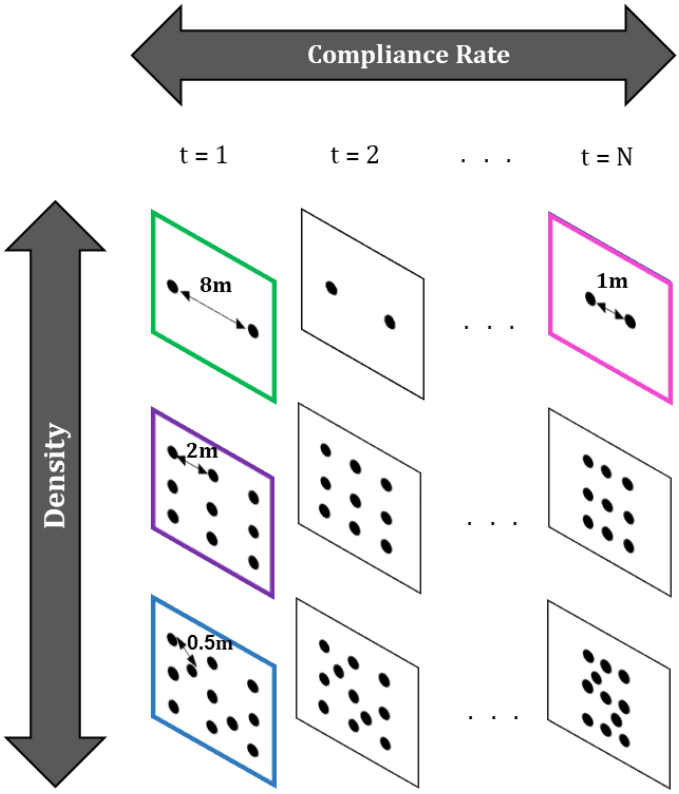


Figure 5-1: Examples of changes in density/compliance rate over time

Understanding the different walking conditions shown in the four scenarios is important when choosing a mobility intervention. For example, Scenario 3 (blue box) clearly demonstrates a situation that requires increased capacity. The surface infrastructure changes proposed in the

Slow Streets program might be appropriate mobility interventions. On the other hand, Scenario 4 (pink box) would not require changes in surface infrastructure. A stronger education and/or enforcement program would probably be more appropriate.

5.1.3. Study Goal and Objectives

The goal of this study is to evaluate how well physical distancing is being achieved on urban sidewalks during a pandemic. We propose a risk analysis that can evaluate and monitor the sidewalks using a sophisticated algorithm rather than the current binary distance threshold approach (2-meter physical distancing) which is based only on a simple distance cut off. The study has three objectives:

1) Develop pedestrian physical distancing indicators that can quantitatively evaluate different levels of physical distancing;
2) Propose levels of pedestrian physical distancing (LPPD) that can be used to select and implement appropriate mobility interventions; and
3) Develop a risk analysis algorithm to estimate the relative risk of viral transmission between pedestrians under different pedestrian walking conditions.

The study involved two major assumptions:

1) None of the pedestrians on the sidewalk are not members of a household unit (e.g., family members).
2) Pedestrians are not using face masks.

Every pedestrian should, therefore, maintain the required physical distance.

Section 5.2 discusses literature review. Section 5.3 presents the three-step approach in relation to LPPD and risk analysis algorithm. Section 5.4 presents a simulation study to

demonstrate the results of our three-step approach. Section 5.5 discusses the findings of this study as well as the recommendations for future studies.

5.2. Literature Review

In the literature review, we focus on two existing approaches to measuring the quality of pedestrian walking environments: the pedestrian level of service (PLOS) set out in the US Highway Capacity Manual (HCM) (TRB, 2016), and the pedestrian comfort level (PCL) developed by the City of London, England (Finch et al., 2010). We adapted and combined some useful concepts in the PLOS and the PCL to develop our LPPD.

5.2.1. Pedestrian Level of Service (PLOS)

The HCM's PLOS is the measure most widely used to describe the operational conditions of pedestrian flows on urban sidewalks. The HCM's PLOS is based on two indicators, 1) pedestrian space (meter^2/pedestrian) and 2) pedestrian flow rate (pedestrian/min/m), and lists six levels ranging from A (best condition) to F (worst condition). If the number of pedestrians per minute decreases, the space per pedestrian increases and the operational walking conditions for pedestrians improve. Transportation engineers use PLOS to improve pedestrian infrastructure by, for example, widening the sidewalk.

LOS A describes a walking condition in which all pedestrians can choose and vary their walking speed, and conflicts between moving pedestrians are unlikely. LOS B to D describe walking conditions ranging from minimal restriction on movement to increasingly restricted movement due to reduced space and slower walking speeds. LOS E describes a walking condition in which pedestrian volume is at capacity. Pedestrians have no space for passing other pedestrians, and often need to adjust their walking speed due to slow moving pedestrians ahead. LOS F describes a walking condition in which pedestrian flows and walking speeds are severely restricted

and queues develop on the walkway (TRB, 2016).

5.2.2. Pedestrian Comfort Level (PCL)

The City of London, England has very high numbers of pedestrians using some of its walkways for considerable periods of time each day. To assess whether a walkway was adequate for the pedestrian volume, the City developed a measure of pedestrian flow conditions known as the pedestrian comfort level (PCL).

The PCL is based on two indicators: 1) pedestrian crowding, and 2) restricted movement. As pedestrian crowding is based on the number of pedestrians per minute per meter (pedestrian/min/m) (i.e., same as the pedestrian flow indicator in HCM's PLOS). Restricted movement is a measure of comfort derived from the percentage of pedestrians observed to experience restricted movement. If the percentage is high, the conditions are not comfortable as higher number of pedestrians need to change their route and speed to avoid problems like shoulder brushing and bumping into other pedestrians. The PCL lists five main levels ranging from A (very comfortable) to E (very uncomfortable). Levels A, B and C can be sub-divided (+/0/-) giving a total of 11 PCL levels (Finch et al., 2010).

5.3. Methodology

The study used three steps to propose a LPPD and estimate the risk of contracting a virus infection at different physical distances.

1) Developing Indicators: As pedestrian density and pedestrian exposure time appear to be the two most important indicators affecting the risk of pedestrians contracting COVID-19. To provide a quantitative measure each indicator, it was necessary to develop a mathematical expression for each indicator.

2) Developing LPPD: We develop LPPD similar to the PLOS and PCL using the two indicators developed. We explain the relationship between the two indicators and the proposed LPPD.

3) Developing Risk Analysis Algorithm: We develops a mathematical process that can measure the relative risk of viral infection due to the different physical distances between pedestrians. We then demonstrate the use of this algorithm using a simulation.

5.3.1. Developing Indicators

1) Developing Indicator I (pedestrian space): It refers to pedestrian space (PS), the average space available to each pedestrian area regardless of hitting other pedestrians expressed as square meters per pedestrian. Equation (5-1) shows the formula for PS:

$$PS = \frac{S_p}{V_p} \tag{5-1}$$

where

PS is Pedestrian space (m^2/p),

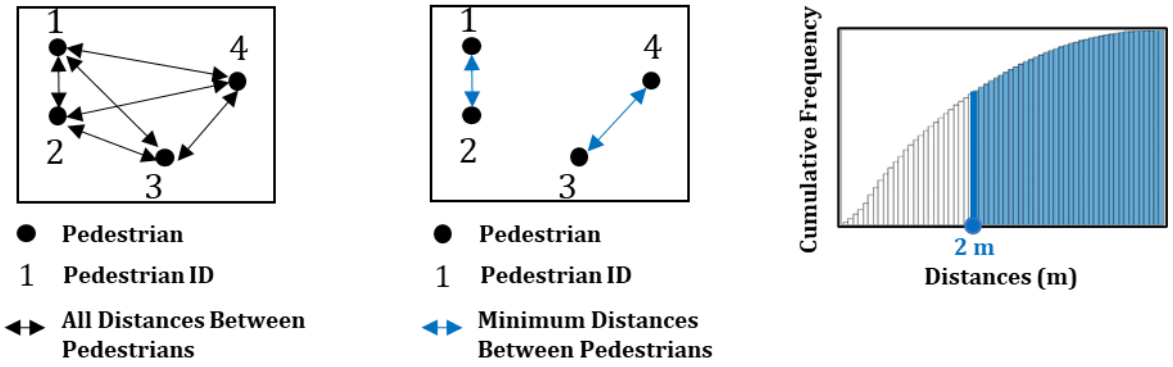
S_p is Pedestrian speed (m/min), and

V_p is Pedestrian flow per unit width (pedestrian/min/m).

Pedestrian space is an indicator used in the HCM to define PLOS. The indicator involves estimating pedestrian speed in the study area for one minute and estimating the flow rate for one minute. The pedestrian speed is then divided by the flow rate to estimate the average space for each pedestrian. Predefined thresholds are used to determine the different PLOS. The PLOS thresholds are based on pedestrian comfort when walking on sidewalks and do not consider physical distancing. In this study, we used the pedestrian space indicator developed by HCM PLOS, but we used physical distancing to define the thresholds for LPPD.

2) Developing Indicator II (compliance rate): It is the compliance rate. The rate incorporates the physical distance threshold and can be viewed as a surrogate measure for exposure time: a higher compliance rate implies a lower exposure time. Figure 5-2a shows a space with four pedestrians. At any moment, there are six physical distances between the pedestrians in this particular example. We consider the minimum distance for each pedestrian. As an example, for pedestrian #1, the physical distance with pedestrian #2 (blue color line) is the minimum distance out of three distances need to be considered. The distance also becomes the minimum distance for pedestrian #2. For pedestrian #3 and #4, the distance between themselves (blue color line) is the minimum distance for these two pedestrians. See Figure 5-2b. We repeat this observation each second for all pedestrians during a preset study period (e.g., 1 minute). The minimum distance each second obtained for all the pedestrians in the study area during the study period can then be used to build a cumulative frequency histogram. See Figure 5-2c. The total area under the cumulative frequency histogram represents the total time spent by all pedestrians in the study area regardless of physical distance between them. Using the histogram, we can then estimate the percentile of pedestrians who have maintained the required physical distancing (e.g., 2m) over the study period at the study area. See blue area in Figure 5-2c. We can argue that the greater the percentile values (that means larger the number of pedestrians who maintained the required physical distancing during a study period), less the possibility of viral transmission between pedestrians (e.g., less number of persons who exposed to the less than required physical distancing). Compliance rate is an indicator that can be viewed as the inverse of exposure time. This study defined the exposure time as the time span used to identify pedestrians who were continuously violating 2 meters of physical distancing. For example, exposure time might be defined as two pedestrians violating 2 meters of physical distancing continuously for 9 seconds

out of 10 seconds of the total trip duration of the two pedestrians in a study area. In this case, exposure time would be calculated as 0.9 (90%), and the compliance rate as 0.1 (10%), indicating that pedestrians could maintain physical distancing for 10% of the total trip duration (one second in this example).



a) All distances among four pedestrians

b) Minimum distance for four pedestrians

c) Cumulative frequency histogram

Figure 5-2: The concept of indicator II (compliance rate)

5.3.2. Developing LPPD

Figure 5-3 shows the concepts incorporated into the proposed LPPD. As an example, if we use 2 meters as our threshold for physical distancing, each pedestrian then requires a circle of space with a radius of 1 meter to ensure compliance with physical distancing requirements.

We developed six primary levels of LPPD ranging from A to F. These levels were based on Indicator I and are similar to the six levels in the HCM's PLOS. We also developed three sub-levels (+, 0, -) for each LPPD (except level F). The based sub-levels were based on Indicator II and are similar to the three sub-levels used in the City of London's PCL (Finch et al., 2010).

1) Creating six Primary Levels based on Indicator I

Each primary level can be interpreted as follows:

- LPPD A represents a pedestrian walking environment in which pedestrians can maintain at least 8 meters of physical distance from other pedestrians without needing to alter their walking speed or path. LPPD A implies that pedestrians minimize the possibility of contracting with virus as they are maintaining a distance that exceeds the distance over which transmission can occur via small aerosols under certain temperature and moisture conditions (Parshina-Kottas et al., 2020). If a sidewalk can provide PS $\geq$ 50.24 (m^2/p), we consider the space to be operating under LPPD A.

- LPPD E represents a pedestrian walking environment in which pedestrians can maintain the required physical distancing, but the sidewalk is operating at or near capacity in terms of the pedestrian flows. Under LPPD E, pedestrians often need to restrict their normal walking speed and adjust their gait to maintain physical distancing. It is extremely hard for pedestrians to overtake slow moving pedestrians without violating the physical distancing requirement. If a sidewalk can provide 0.78 (m^2/p) < PS $\leq$ 3.14 (m^2/p), we consider the space to be operating under LPPD E.

- LPPD F represents a pedestrian walking environment that is operating beyond capacity. Some pedestrians need to restrict their walking speeds severely and it is not possible for the pedestrians to progress forward without violating the physical distancing requirement. If a sidewalk provides PS $\leq$ 0.78 (m^2/p), we consider the space to be operating under LPPD F, and the pedestrians appear to be queuing rather than moving forward.

- To adopt the HCM's PLOS that describes six levels of pedestrian walking conditions, we have derived LPPD B to LPPD D by gradually changing PS constraint for each level (see Figure 5-3 for the proposed PS for each level). As an example, the LPPD D presents a situation when each pedestrian can maintain 2 meter (r = 1 meter) to 4 meter (r = 2 meter)

of physical distance between each other and corresponding PS ranges between 3.14 (m^2/p) < PS ≤ 12.56 (m^2/p).

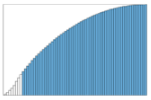
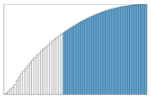
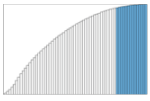
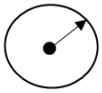
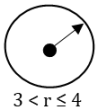
Radius (m) = r	Pedestrian Space (m^2/p) = πr^2	Compliance Rate		
		85 %	15% ~ 85 %	15%
				
 4 < r	50.24 < PS	A+	A0	A-
 3 < r ≤ 4	28.26 < PS < 50.24	B+	B0	B-
 2 < r ≤ 3	12.56 < PS < 28.26	C+	C0	C-
 1 < r ≤ 2	3.14 < PS < 12.56	D+	D0	D-
 0.5 < r ≤ 1	0.78 < PS < 3.14	E+	E0	E-
 0.5 ≥ r	PS < 0.78	F		

Figure 5-3: Level of pedestrian physical distance (LPPD)

2) Creating 3 Sub-levels based on Indicator II

We used Indicator II to develop three sub-levels of pedestrian exposure. Transportation engineering often uses the 85th and the 15th percentile values as a threshold for design and operational parameters (Hou et al., 2012). This is especially true when safety (i.e., saving lives) is the main consideration. For example, the theoretical rationale for defining the maximum posted speed limit for highway segments is based on the 85th percentile operating speed (Milliken, 1998).

The 85th percentile values are also used to determine minimum stopping sight distance in relation to important geometric design parameters such as perception and reaction time (AASHTO, 2018). The 15th percentile walking speed is used to determine pedestrian signal timing at pedestrian crossings (Knoblauch, 1996) and also provides the theoretical rationale for defining the minimum posted speed limit on high-speed highways in the United States that maintain a minimum speed limit as well as a maximum speed limit (Mussa, 2004).

As Figure 5-3 showed, if 85th or greater percentile of pedestrians have complied with the required physical distancing over the preset study period (10 minutes), the sub-level will be specified as "+." If 15th percentile or less percentile of pedestrians have complied with the required physical distance during the study period, the sub-level will be specified as "-." Lastly, if the percentile values are positioned between these two thresholds, we defined the sub-level as "0." Note that the 85th and 15th percentile values of compliance rate can be estimated from the cumulative frequency histogram of pedestrians in Figure 5-2. In summary, we applied the three sub-levels (high (+), medium (0) and low (-)) to LPPD A to E (See Figure 5-3).

From LPPD A to LPPD E, pedestrians maintain physical distancing of at least 2 meters. Under LPPD F, it is not possible for all pedestrians to maintain this level of physical distancing and the compliance rate is not calculated.

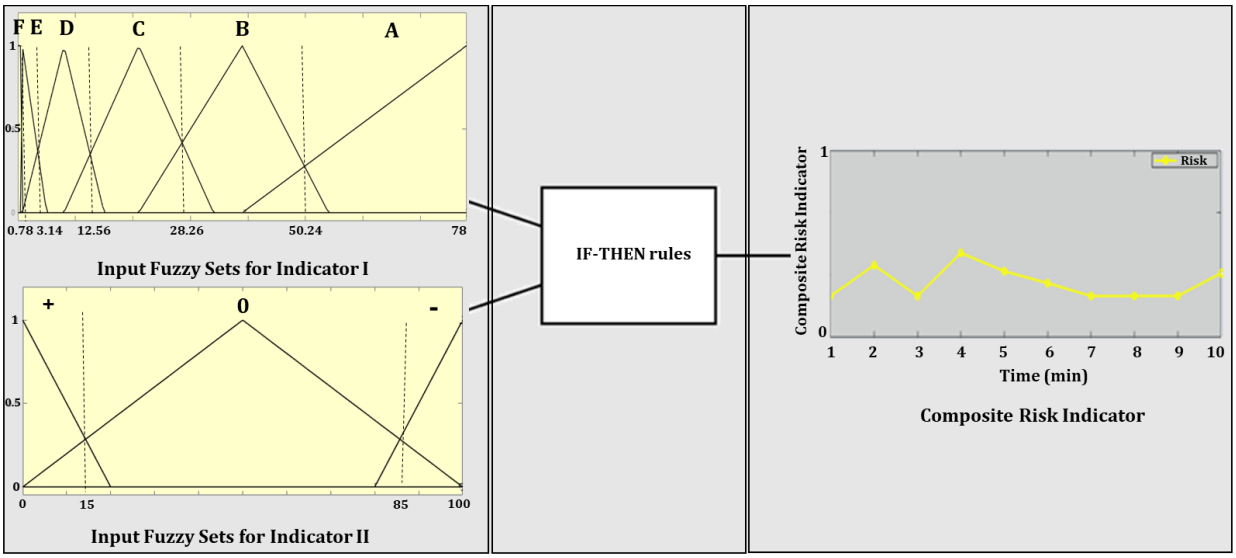
5.3.3. Developing Risk Analysis Algorithm

In the risk analysis algorithm, we use the two indicators explained in section 5.3.1 to quantitatively estimate the relative risk of viral infection (i.e., composite risk indicator). As the indicators involve uncertain, vague and/or subjective sets of data, we used a fuzzy algorithm for our algorithm development (Zadeh, 1965).

A fuzzy algorithm uses a membership function to determine a membership value between

0 and 1 for each input indicator. The membership function can be changed into a “fuzzy set” designed to capture the uncertainty of the input and output variables. These fuzzy sets can be defined as categorical variables (levels A to F in our case) or as binary situations such as yes/no, true/false, etc. (Mendel, 1995). A fuzzy algorithm has three main stages: a) fuzzification, b) rule evaluation, and c) defuzzification (Mendel, 1995) (see Figure 5-4).

In the fuzzification stage, we used fuzzy linguistic variables to convert the two input indicators into fuzzy sets. Indicator I provided six input fuzzy sets (A to F). Indicator II provided three additional fuzzy sets (+, 0, -). We also determined the shape of the membership functions for each fuzzy set. We used a “triangular” form for the membership function as this is the most commonly used functional form (Barua et al., 2013).



a) Fuzzification c) Rule evaluation d) Defuzzification

Figure 5-4: A fuzzy rule-based algorithm for risk analysis

In the rule evaluation stage, we defined 16 linguistic IF-THEN rule constructions. For example, we defined fuzzy rule #1 as: if Indicator I is “LPPD A” and Indicator II is "+," then the relative value of composite risk indicator (output) is “very low.” In the defuzzification stage, we

converted the fuzzy relative value of composite risk indicator outputs to numerical values describing the estimated composite risk indicator. These values ranged from 0 (lowest relative risk) to 1 (highest relative risk).

5.4. Simulation Study

5.4.1. Virtual Walking Environment

To demonstrate the application of the proposed LPPD and risk analysis algorithm, we created a virtual pedestrian walkway to mimic pedestrians' movement along an urban sidewalk. We investigated downtown Toronto sidewalks (e.g., Queen Street East between Parliament Street and Sackville Street along) where the local jurisdiction implemented mobility interventions. As we found that most downtown Toronto sidewalks are about 3 meters wide and 20 meters long, we used 3 meters as the width and 20 meters as the length of the study's sidewalk study area. We used PTV Viswalk (see Figure 5-5) for the simulation. PTV Viswalk which is a popular pedestrian simulation tool used to track pedestrian walking trajectories (the x-y coordinates of pedestrians in a study area) each second over a preset period of time (Das et al., 2014).

We used PTV Viswalk for the simulation. PTV Viswalk is a popular pedestrian simulation tool that can be used to track pedestrian walking trajectories (the x-y coordinates of pedestrians in a study area) each second over a pre-set period of time (Das et al., 2014). PTV Viswalk formulizes pedestrian motion using Equation (5-2):

$$m_a \times \frac{dv_a}{dt} = F_a \tag{5-2}$$

Equation (5-2) calculates the force required for pedestrian a with mass of m_a and speed of v_a to change speed. However, Equation (5-2) is very simple and only considers the direction and desired speed of each pedestrian.

In addition to Equation (5-2), PTV Viswalk uses a mathematical equation known as the social force model to formulize other pedestrian behaviors. These behaviors include avoiding hitting other objects such as other pedestrians as shown in Equation (5-3) (Helbing and Molnar, 1995; PTV, 2020):

$$F_{\alpha\beta} = A \times exp\left[\frac{(r_{\alpha\beta} - d_{\alpha\beta})}{B}\right] \times n_{\alpha\beta} \quad (5\text{-}3)$$

In Equation (5-3),

$F_{\alpha\beta}$ = the social force of pedestrian β to pedestrian α.

A and B = the strength coefficients.

$n_{\alpha\beta}$ = the unit vector of participant β pointing to participant α,

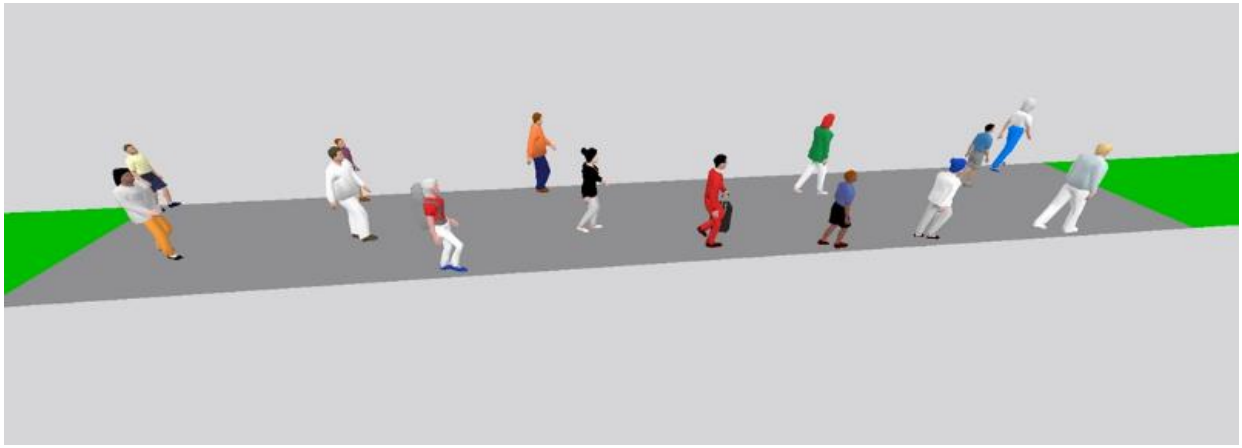
$d_{\alpha\beta}$ = the distance between pedestrian β to pedestrian α (body surface to body surface)

$r_{\alpha\beta}$ = the additional required distance in meters between pedestrian β to pedestrian α.

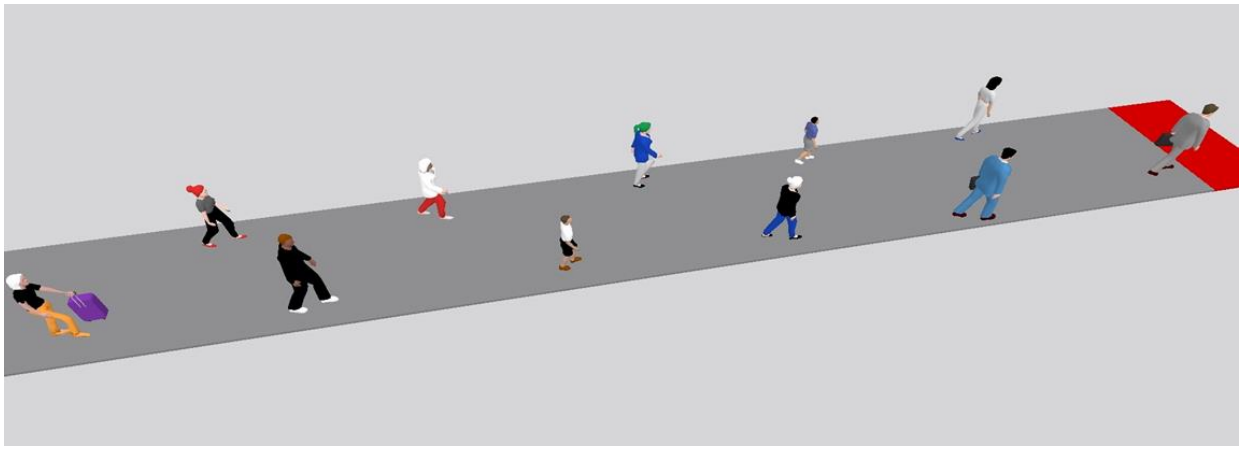
If $r_{\alpha\beta} = 0$, the social force model ensures that pedestrians do not to collide with each other. If we set $r_{\alpha\beta}$ to 2, the social force model ensures that pedestrians pass each other with more than 2 meters of distance, i.e., setting $r_{\alpha\beta}$ to 2 enables the model to apply the 2 meters of physical distancing rule. By multiple trial and error, we gave A a value of 11, and B a value of 0.66. $d_{\alpha\beta}$ and $n_{\alpha\beta}$ were determined by the simulation in each unit of time (i.e., in each second). With these parameters, we found that on average, 90% of the pedestrians maintained 2 meters of physical distancing while walking along the sidewalk

The number of pedestrians (the pedestrian flow rate) is directly associated with the space available to each pedestrian (the pedestrian space). As the pedestrian flow rate increases, the pedestrian space decreases. The pedestrians can maintain the required physical distancing under LPPD A to C as long as all pedestrians are willing to make the necessary speed and direction

adjustments. On the other hand, it is not possible for pedestrians to maintain the required physical distancing under LPPD F due to lack of capacity.



a) Pedestrians walk randomly without attempting to maintain the required physical distancing



b) All pedestrians walk while maintaining the required physical distancing

Figure 5-5: Pedestrian simulation environment in PTV Viswalk

In this study, we simulated pedestrian flows for LPPD D (25 p/min/m) and LPPD E (35 p/min/m). Notice, however, that LPPD E is not considered a sustainable or stable walking environment as many pedestrians under LPPD E cannot continuously and consistently maintain the exact physical distancing required, and thus LPPD E can easily transition into LPPD F. We created three scenarios:

- Base Scenario: The base scenario was a default scenario that we could compare with scenarios 1 and 2. In the base scenario, pedestrians could walk randomly without attempting to maintain the required physical distancing. The flow rate set at 25 p/min/m.
- Scenario 1: Scenario 1 was created to demonstrate the impact of change in Indicator I (pedestrian space). The flow rate was (35 p/min/m) which allowed all pedestrians to walk randomly without attempting to maintain the physical distancing.
- Scenario 2: Scenario 2 was created to demonstrate the impact of change in Indicator II (compliance rate). The flow rate was 25 p/min/m (the same as the base scenario). All pedestrians in this scenario were forced to maintain 2 meters of physical distancing.

We simulated each scenario twice for 10 minutes per simulation. All the pedestrians in each scenario walked in the same direction (movement to one direction) and each pedestrian's average desired walking speed was set at 1.3 meter per second with 0 standard deviation. In reality, pedestrians exhibit a wide range of desired walking speed, varying between 0.8 meter per second to 1.8 meter per second (TRB, 2020). We set a value of 1.3 meter per second (average value of 0.8 m and 1.8 m) for pedestrian's desired walking speed. In the analysis, we extracted the x-y coordinates for each pedestrian's second by second movement along the virtual walkway. This produced 3,600 raw datasets (10 minutes × 60 seconds × 3 scenarios × 2 simulations per scenario). These datasets were converted into input indicators and used in the LPPD analysis and risk estimation. We used the Matlab Simulink package for the analysis (Mathworks, 2016).

5.4.2. Results of Analysis

Table 5-1 shows the average estimates Indicators I and II each over the 10 minutes of simulation for each of the three scenarios and the corresponding LPPD (see Figure 5-6).

Table 5-1: Estimation of indicators and LPPD

Scenario	Indicator I (pedestrian space)	Indicator II (compliance rate)	LPPD
Base Scenario	3.23 m^2/p	39th Percentile	D0
Scenario 1	2.36 m^2/p	22th Percentile	E0
Scenario 2	3.16 m^2/p	87th Percentile	D+

If we compare the base scenario with scenario 1, the increase in the pedestrian flow rate (from 25 p/min/m in the base scenario to 35 p/min/m in scenario 1 has decreased Indicator I from 3.23 m^2/p in the base scenario to 2.36 m^2/p in scenario 1. Indicator II has decreased from the 39th percentile in the base scenario to the 22th percentile in scenario 1. The LPPD for the base scenario is D0, but the decrease in Indicator I reduces the LPPD to E0 in scenario 1.

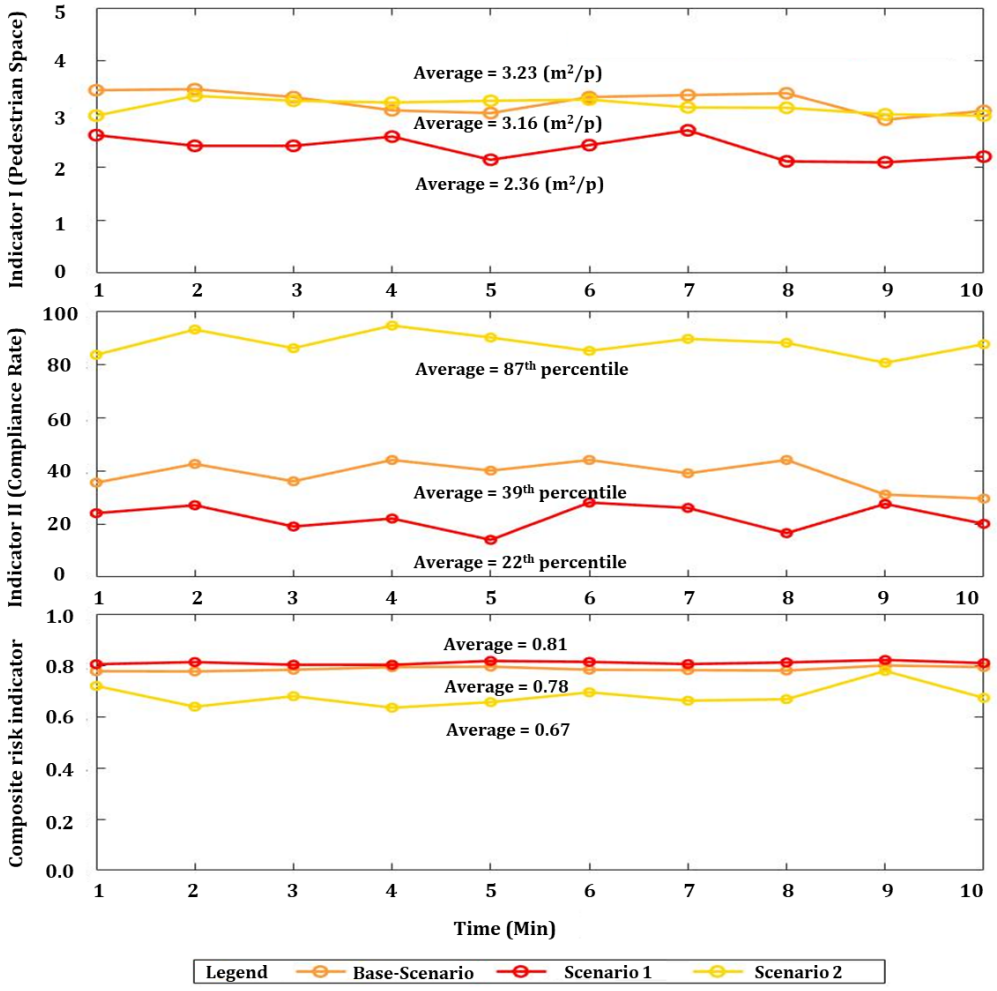


Figure 5-6: Input indicators and estimated composite risk indicator for scenarios

The change in Indicator II does not change the LPPD as both percentiles are treated the same (i.e., both fall within the 15% to 85% area). If we compare the base scenario with scenario 2, Indicator I has decreased slightly (from 3.23 m^2/p in the base scenario 1 to 3.16 m^2/p in scenario 2, but Indicator II has increased significantly (from the 39^{th} percentile in the base scenario to the 87^{th} percentile in scenario 2.

The LPPD has changed from D0 in the base scenario to D+ in scenario 2 due to the increase in Indicator II. The change in Indicator I did not not change the LPPD. The estimates of Indicators I and II (see Table 5-1) were used as inputs by the fuzzy algorithm to output the estimated relative risk during the 10-minute simulation period for each scenario. Figure 5-6 shows the results. Scenario 1 had the highest average risk (0.81) followed by the base scenario (0.78) followed by scenario 2 (0.67).

5.5. Discussion and Recommendations

In this section, we discuss the implications of the proposed LPPD and the implications of the risk analysis algorithm.

5.5.1. Implications of LPPD

The purpose of the LPPD is to help decision makers select the most appropriate mobility interventions. In our simulation study, for example, to improve the LPPD of scenario 1 from E0 to the base scenario's D0, we can modify the surface infrastructure (e.g., the cross-sectional design of urban streets) to increase the space available (capacity) for pedestrians/bicyclists. The Slow Streets program mentioned in Section 5.1.1 includes various appropriate mobility interventions (temporarily or permanently replacing curbside parking spaces or curbside lanes with pedestrian

walkways and/or bicycle lanes, completely prohibiting non-essential vehicular traffic on designated urban streets, etc.).

If, however, we wish to improve the LPPD of the base scenario (D0) to scenario 2's D+, increasing walking space may not be the best approach. In this case, we need to focus on mobility interventions that influence pedestrian walking behavior. Such interventions include educational programs (e.g., advertisements via outdoor signage and social media) and/or enforcement programs (e.g., ticketing and fining people who violate the physical distancing requirement). Increasing the available pedestrian space is an appropriate solution to improve the level of pedestrian physical distancing only if pedestrian flow does not increase. Thus, it would be more effective to control pedestrian flow to prevent excessive pedestrian conflicts in the limited space.

The HCM's PLOS does not consider 2-meter physical distancing. Table 5-2 compares the thresholds for the HCM's pedestrian LOS with the thresholds for our LPPD which takes 2-meter physical distancing into account. The much higher thresholds required for our LPPD show the justification for using the proposed LPPD rather than HCM's PLOS when trying to maintain physical distancing during a pandemic. In addition, the HCM PLOS does not take exposure time into account whereas our proposed input indicator II (compliance rate) is designed to capture exposure time."

Table 5-2: Thresholds in HCM's pedestrian LOS and our proposed LPPD

HCM without considering physical distance		**Proposed LPPD considering physical distance**	
PLOS HCM	**Pedestrian space (PS) (m^2/p) HCM thresholds**	**Proposed LPPD**	**Pedestrian space (PS) (m^2/p) Our study thresholds**
A	$5.57 < PS$	A	$50.24 < PS$
B	$3.71 < PS \leq 5.57$	B	$28.26 < PS \leq 50.24$
B	$2.22 < PS \leq 3.71$	C	$12.56 < PS \leq 28.26$
D	$1.39 < PS \leq 2.22$	D	$3.14 < PS \leq 12.56$
E	$0.74 < PS \leq 1.39$	E	$0.78 < PS \leq 3.14$
F	$PS \leq 0.74$	F	$PS \leq 0.78$

5.5.2. Implications of Risk Analysis Algorithm

Figure 5-7 uses a 3D format to show the estimated change in the relative value of composite risk indicator by changes in Indicators I and II. The location of the three scenarios is shown on the risk surface. The figure clearly shows that a decrease in Indicator I and/or Indicator II, increases the estimated value of composite risk indicator. A small decrease in Indicator I leads to a relatively large increase in the estimated composite risk indicator. Changes in Indicator II are less sensitive and do not lead to such dramatic changes in the estimated composite risk indicator. This analysis may imply that reducing density by improving capacity is the most effective way to reduce the relative risk of possible viral infection (although such approach is expected to be more expensive).

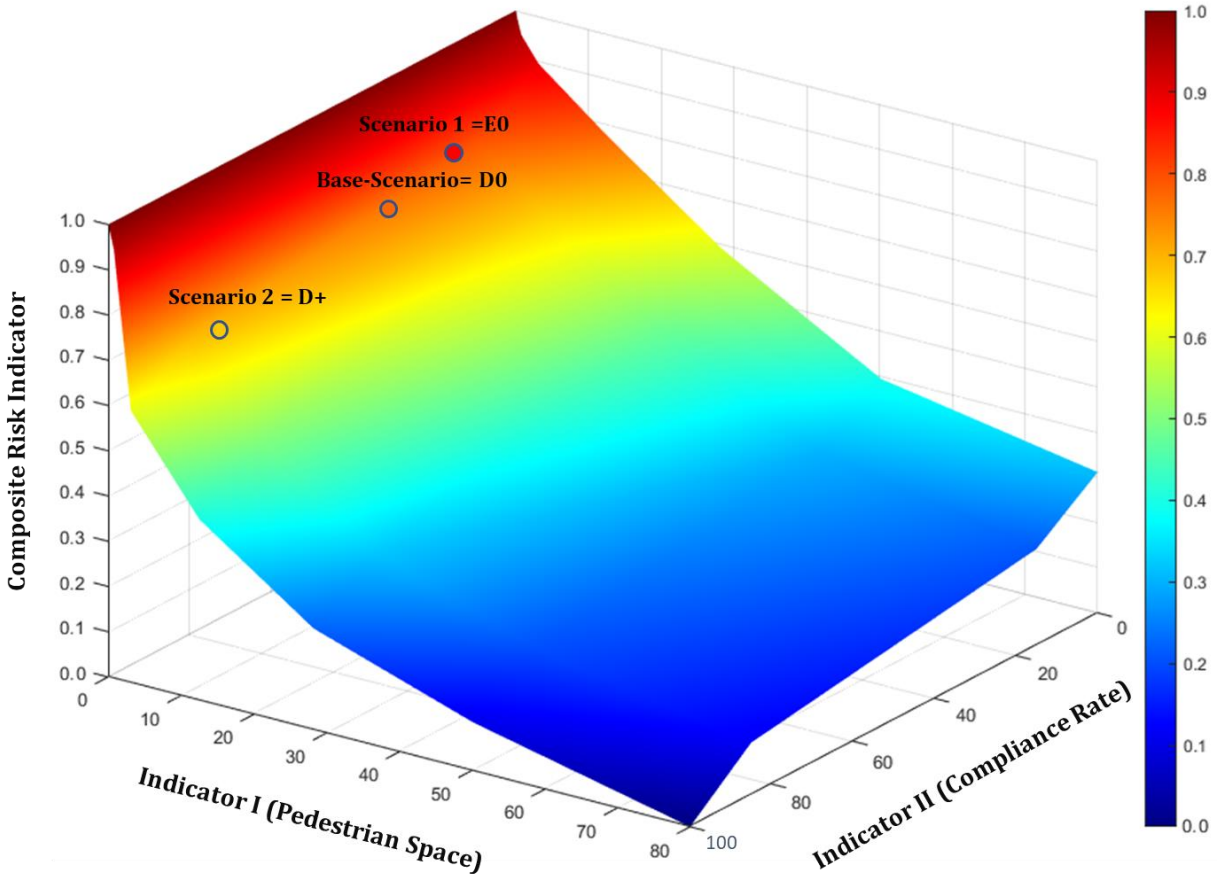


Figure 5-7: 3D representation of the estimated change in the relative value of composite risk indicator

The violation of physical distancing rules may increase the chance of getting COVID-19, but the composite risk indicator in this study does not represent the risk of getting COVID-19. The composite risk indicator refers only to the risk of violating 2 meters of physical distancing between pedestrians.

5.5.3 Recommendations for Future Work

Our study considered an outdoor space where pedestrians are not required to wear masks. As wearing masks may reduce the physical distancing threshold to less than 2 meters, a study of indoor space would require different calibration. Future research could investigate the effect of mask wearing on physical distancing by recalibrating the approach developed in this study. The study would require a scientifically acceptable estimate of appropriate physical distancing (e.g., 0.5 meter) for pedestrians wearing masks, but this information is not yet clearly known. The results of a such study would provide estimates of the presumably reduced risk of viral infection. This study focused on pedestrian sidewalks on urban streets. Future work could consider developing a similar LPPD and/or risk analysis algorithm targeting for indoor places (e.g., commercial plazas, major shopping centers, factories, etc.) where people are often close together. Future work also needs to automate the process of data collection and data process to determine the LPPD and the relative risk associated with the two proposed indicators. Various vision technologies based, for example, on CCTV can detect the second by second movement of pedestrians and are already used in various areas in transportation engineering (Teknomo et al., 2016; Brunetti et al., 2018).

We conducted this study to combat the current spread of COVID-19, but the proposed approach is also relevant to the countless types of influenza and the possibility that another pandemic will occur. We clearly need to know as much as possible about the potential role of mobility interventions and other strategies designed to limit the future pandemic.

5.6. Chapter 5 References

American Association of State Highway and Transportation Officials (AASHTO), (2018) "A Policy on Geometric Design of Highway and Streets".

Barua, A., Mudunuri, L. S., and Kosheleva, O. (2013). Why trapezoidal and triangular membership functions work so well: Towards a theoretical explanation. Journal of Uncertain Systems, Vol.8, No.3, pp.164-168.

Transportation Research Board. (2016). Highway capacity manual: A guide for multimodal mobility analysis. 6th ed. Washington, DC: Transportation Research Board of the National Academies.

Bourouiba, L. (2020). Turbulent gas clouds and respiratory pathogen emissions: potential implications for reducing transmission of COVID-19. Jama, 323(18), 1837-1838.

Brunetti, A., Buongiorno, D., Trotta, G. F., and Bevilacqua, V. (2018). Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing, 300, 17-33.

Centers for Disease Control and Prevention (CDC). Frequently Asked Questions and Answers. [Online] 19 March 2020. Available at: https://www.cdc.gov/coronavirus/2019-ncov/faq.html [Accessed June, 2022].

Chen, S., Yang, J., Yang, W., Wang, C., and Barnighausen, T. (2020). COVID-19 control in China during mass population movements at New Year. The Lancet, 395(10226), 764-766.

Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., and Hajizadeh, A. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The

Lancet. https://doi.org/10.1016/S0140-6736(20)31142-9

City of Toronto (CoT). COVID-19: ActiveTO. 2020. Available at: https://www.toronto.ca/home/covid-19/covid-19-protect-yourself-others/covid-19-reduce-virus-spread/covid-19-activeto/. [Accessed June, 2022].

Das, P., Parida, M., and Katiyar, V. (2014). Review of Simulation techniques for microscopic mobility of pedestrian movement. Trends in Transport Engineering and Applications, 1(1), 27-45.

de Bruin, Y. B., Lequarre, A. S., McCourt, J., Clevestig, P., Pigazzani, F., Jeddi, M. Z., and Goulart, M. (2020). Initial impacts of global risk mitigation measures taken during the combatting of the COVID-19 pandemic. Safety Science, 104773. https://doi.org/10.1016/j.ssci.2020.104773

European Centre for Disease Prevention and Control (ECDC). 2020. Considerations Relating to Social Distancing Measures in Response to COVID-19—Second Update. March 23. Available at: https://www.ecdc.europa.eu/en/publications-data/considerations-relating-social-distancing-measures-response-covid-19-second [Accessed June, 2022].

Feng, Y., Marchal, T., Sperry, T., and Yi, H. (2020). Influence of wind and relative humidity on the social distancing effectiveness to prevent COVID-19 airborne transmission: A numerical study. Journal of Aerosol Science, 105585. https://doi.org/10.1016/j.jaerosci.2020.105585

Finch, E., Iverson, G., and Dye, J. (2010). Pedestrian Comfort Guidance for London. In European Transport Conference, 2010. Glasgow Scotland, United Kingdom Available at:

http://content.tfl.gov.uk/pedestrian-comfort-guidance-technical-guide.pdf [Accessed June, 2022]

Fong, M. W., Gao, H., Wong, J. Y., Xiao, J., Shiu, E. Y., Ryu, S., and Cowling, B. J. (2020). Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings social distancing measures. Emerging Infectious Diseases, 26(5), 976.

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. Physical review E, 51(5), 4282.

Hou, Y., Sun, C., and Edara, P. (2012). Statistical test for 85th and 15th percentile speeds with asymptotic distribution of sample quantiles. Transportation Research Record, 2279(1), 47-53.

Huynh, T. L. D. (2020). Does culture matter social distancing under the COVID-19 pandemic?. Safety Science, 104872. https://doi.org/10.1016/j.ssci.2020.104872

Knoblauch, R. L., Pietrucha, M. T., and Nitzburg, M. (1996). Field studies of pedestrian walking speed and start-up time. Transportation research record, 1538(1), 27-38.

Leung, N. H., Chu, D. K., Shiu, E. Y., Chan, K. H., McDevitt, J. J., Hau, B. J., and Seto, W. H. (2020). Respiratory virus shedding in exhaled breath and efficacy of face masks. Nature Medicine, 26(5), 676-680.

Mathworks. MATLAB & Simulink: Simulink reference r2015b. 2016. [Online]. Available at: http://www.mathworks.com/help/releases/R2015b/pdf_doc/simulink/slref.pdf [Accessed June, 2022].

Mendel, J. M. (1995). Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE, 83(3), 345-377.

Milliken, G. (1998). Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits. Special Report 254.

Mussa, R. N. (2004). Speeds on rural interstate highways relative to posting the 40-mph minimum speed limit. Journal of Transportation and Statistics, 7(2-3), 71.

National Association of City Transportation Officials (NACTO). (2020.) Streets for pandemic response & recovery. Global Designing Cities Initiative. Available at: https://nacto.org/wpcontent/uploads/2020/09/Streets_for_Pandemic_Response_Recovery_Full_20-09-24.pdf

Parshina-Kottas, Y., Saget, B., Patanjali, K., Fleisher, O., and Gianordoli, G. (2020). This 3-D simulation shows why social distancing is so important. The New York Times. Available at: https://www.nytimes.com/interactive/2020/04/14/science/coronavirus-transmission-cough-6-feet-ar-ul.html. [Accessed June, 2022]

PTV Group: PTV Viswalk (2020). Available at: https://usermanual.wiki/Document/Vissim20102020 Manual.1098038624.pdf [Accessed June, 2022].

San Francisco Municipal Transportation Agency (SFMTA). (2020). Slow streets program to help with social distancing. Available at: https://www.sfmta.com/blog/slow-streets-program-help-social-distancing. [Accessed June, 2022]

Stadnytskyi, V., Bax, C. E., Bax, A., and Anfinrud, P. (2020). The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. Proceedings of the National Academy of Sciences, 117(22), 11875-11877

Shukman D. (2020). Coronavirus: Could social distancing of less than two metres work? BBC. June 23, Available at: https://www.bbc.com/news/science-environment-52522460

[Accessed June, 2022]

Teknomo, K., Takeyama, Y., and Inamura, H. (2016). Tracking system to automate data collection of microscopic pedestrian traffic flow. arXiv preprint arXiv:1609.01810.

Transportation Research Board (TRB). Highway capacity manual (2000). Washington, D.C.: Transportation Research Board, National Research Council. Available at: https://sjnavarro.files.wordpress.com/2008/08/highway_capacital_manual.pdf [Accessed June, 2022]

Transportation Research Board (TRB), Highway Capacity Manual (2020). A Guide for Multimodal Mobility Analysis, 6th Edition.

Varotsos, C. A., and Krapivin, V. F. (2020). A new model for the spread of COVID-19 and the improvement of safety. Safety Science, 132, 104962. https://doi.org/10.1016/j.ssci.2020.104962

Vuorinen, V., Aarnio, M., Alava, M., Alopaeus, V., Atanasova, N., Auvinen, M., and Hayward, N. (2020). Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors. Safety Science, 130, 104866.

World Health Organization. 2020. Coronavirus Disease (COVID-19) Advice for the Public. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [Accessed June, 2022].

World Health Organization (WHO) (2022). Coronavirus disease (COVID-2019) situation report of June 29. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [Accessed June, 2022].

Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.

Zhang, Y., Jiang, B., Yuan, J., and Tao, Y. (2020a). The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study. medRxiv. https://doi.org/10.1101/2020.03.04.20031187

Zhang, R., Li, Y., Zhang, A. L., Wang, Y., and Molina, M. J. (2020b). Identifying airborne transmission as the dominant route for the spread of COVID-19. Proceedings of the National Academy of Sciences. 117 (26) 14857-14863.

CHAPTER SIX

6. CONCLUSIONS AND FUTURE RESEARCH

6.1. Study Summary

Despite the high number of intersection collisions, collisions on roadway segments are a major concern. The development of collision avoidance systems (CASs) designed to reducing the risk of segment collisions needs a special attention as segment collisions often result in severe consequences especially if the collision occurs outside urban centers where vehicles may be traveling at high speeds. This study focused on developing advanced CASs based on advanced risk analysis algorithms and testing. The study considered three types of segment collision: 1) animal-to-vehicle collisions, 2) pedestrian-to-vehicle collisions, and 3) pedestrian-to-pedestrian collisions.

The currently available CASs for the three selected segment collisions are: roadside animal detection systems (RADSs) designed to reduce the risk of animal-to-vehicle collisions, pedestrian protection systems designed to reduce the risk of pedestrian-to-vehicle collisions, and pedestrian physical distance monitoring systems designed to reduce the risk of pedestrian-to-pedestrian collisions. These existing systems rely on one input indicator and apply a simple risk analysis approach by applying a fixed threshold to differentiate safe situations from risky situations.

As no single indicator can simultaneously capture the many factors (road user behavior, spatial and temporal interactions between two road users, etc) that contribute to a collision, collision risk assessment requires the use of multiple input indicators and a sophisticated algorithm that can combine the different input indicators into a single composite risk indicator. Simultaneously capturing important factors is a challenging task as we do not have the level of

understanding that would justify assigning certain combinations of input indicators to certain risk levels (e.g., low, medium, high). In addition, data are often limited or inadequate. For example, no data for the x- and y-coordinates and speeds of road users are readily available for the seconds building up to a collision or a near miss collision. We cannot simply access available data and apply statistical algorithms such as regression models to assess the statistical relationships between the multiple inputs in order to produce an output in the form of a composite risk indicator output.

To develop appropriate risk analyses for the three selected CASs, this thesis screened and selected or developed multiple input indicators designed to simultaneously capture the different behaviors and spatial and temporal interactions between road users. The study then applied artificial intelligence (AI) algorithms in the risk analysis. These algorithms combined multiple input indicators to create a composite risk indicator that could estimate the varying levels of risk posed by road users on the roadway using real-time data.

The thesis goal was to explore the development of composite risk indicators for different CASs by conducting various risk analyses using different AI algorithms. This goal was achieved by conducting the four studies discussed in Chapters 2 to 5. The following subsections present the main findings for each study.

6.2. Conclusions for Chapter 2

A RADS is a modern CAS designed to detect the presence of large animals near the road and to warn approaching drivers of their presence by activating dynamic warning lights. The current systems available use the position of the animals as the only input indicator to activate the warning lights. As a result, existing RADSs simply produce a binary output (on/off warning lights) based on a binary input (presence/absence of animals) which results in several limitations (e.g., failure

to capture unpredictable animal behaviors such as sudden jumping, etc., and certain driver behaviors such as aggressiveness).

Firstly, we proposed a next generation roadside animal detection system (NG RADS) to reduce the risk of animal-to-vehicle collisions. The full version of the proposed system has three phases:

Phase I (detection): The system uses radar sensors to detect the characteristics of the animal (e.g., position and speed) and the vehicle on a real-time (second by second) basis;

Phase II (risk analysis): The system uses the seed information from Phase I to develop various input indicators to capture animal behavior, driver behavior, and the spatial and temporal interaction between the animal and the vehicle. The system then estimates various levels of risk using a sophisticated AI algorithm that combines all the information into a single composite risk indicator; and

Phase III (communication): The system uses advanced vehicle-to-infrastructure (V2I) communication technologies to broadcast different levels of warning according to the estimated composite risk indicator. Different levels of warning can be directed to different vehicles on a real-time basis to alert drivers to the need to take appropriate action to avoid a collision.

This study focused on Phase II (risk analysis). We selected/developed four input indicators (distance from animals to roadway, magnitude of speeding, stopping sight distance (SSD), and physical distance) to capture animal behavior, driver behavior, and spatial and temporal interaction between animal and driver. We then developed a fuzzy rule-based algorithm which combined the four input indicators and created a composite risk indicator with a value that varied from zero (lowest risk) to one (highest risk).

The main findings of chapter two included followings:

- Changes in input indicators are properly reflected in the composite risk indicator. The composite risk indicator increases as (1) Indicator A (lateral distance from animal to roadway) decreases, (2) Indicator B (magnitude of speeding) increases, (3) Indicator C (SSD) changes from 0 (safest) to one (riskiest), and (4) Indicator D (physical distance) decreases.

- The risk analysis algorithm enables the NG RADS to capture the unpredictable behavior of the animal in real-time. As the estimated value of the composite risk indicator is updated every second, three of the four proposed indicators (Indicators A, C and D) indirectly reflect the unpredictable behavior of a moving animal. For example, Indicator A (lateral distance from animal to roadway) decreases faster each second for an animal moving at speed towards the roadway than for an animal moving slowly towards the roadway. This kind of factor cannot be captured by conventional RADSs.

- The proposed risk analysis algorithm captures driver behavior (aggressiveness/normal) in the composite risk indicator: Indicator B (magnitude of speeding) results in an increase in the value of the composite risk indicator for an aggressive driver (one who drives faster than the posted speed limit). A conventional RADS cannot take driver behavior into account.

- The proposed risk analysis algorithm can capture the risk difference between the leading/following vehicle by using Indicator D (physical distance). A leading vehicle has a lower value for Indicator D (physical distance) which results in an increase in the composite risk indicator. A conventional RADS does not differentiate risk for leading/following vehicle and produces a similar risk for both.

- Conventional RADSs compare the position of an animal relative to the roadway and apply a threshold (e.g., 5 meters) that assigns a binary risk. This approach results in a major deficiency. Assume two situations: (1) the animal is positioned 5.1 meters to the roadway, and (2) the animal is positioned 4.9 meters to the roadway. A conventional RADS produces an alarm for the second animal, but not for the first although the first animal may suddenly jump forward and be within the alarm threshold. Our proposed risk analysis algorithm is more logical because it allows risk to increase as an animal moves closer to the road.
- A conventional RADS produces a fixed and constant level of risk for the entire roadway segment even after the vehicle passes the animal and there is no risk. The proposed algorithm can generate different levels of risk in real time. For example, the proposed algorithm considers the interactions between an animal and a vehicle and the risk becomes zero after the vehicle passes the animal.

6.3. Conclusions for Chapter 3

Modern CASs include pedestrian protection systems designed to alert a pedestrian (e.g., a police officer) to a possible risk from an approaching risky vehicle and then warn the pedestrian to take proactive evasive action if necessary. Existing systems use only a single input indicator (e.g., time to collision (TTC)) and apply a simple risk analysis approach (comparing the indicator with a fixed threshold) to decide whether to alert the pedestrian or not. The limitations of this approach include failing to capture pedestrian behavior, driver behavior, etc.

Firstly, we proposed a smart protection system designed to reduce the risk of a pedestrian (officer)-to-vehicle collisions. The full protection system envisaged involves three phases:

Phase I (detection). A radar system installed in a police officer's car detects the positions (i.e., x- and y-coordinates), trajectories, and speeds of passing vehicles and police officers in real-time (i.e., second by second);

Phase II (risk analysis). The system uses the seed information from Phase I to develop multiple input indicators to capture police officer behavior, driver behavior, and the spatial and temporal interaction between the police officer and passing vehicle. A risk analysis algorithm using a sophisticated AI algorithm is then developed to combine input indicator and to estimate a composite risk indicator; and

Phase III (warning and communication). The system generates the appropriate level of audio and/or vibration warning according to the estimated level of risk. The warning can be disseminated to the police officer's equipment (e.g., mobile radio) or wearable device (e.g., wrist band) and the police officer can take evasive action.

This study focused on Phase II (risk analysis). We developed a fuzzy rule-based algorithm to combine four developed input indicators (lateral distance from police officer to traveled lane, magnitude of speeding, SSD, and direct distance) into a single composite risk indicator to estimate of level of risk of a collision. The value of the composite risk indicator in this study varied from zero (low risk) to one (high risk).

The main findings of chapter three included the followings:

- The 3-D surfaces developed in the study clearly show that when Indicator A (lateral distance from police officer to traveled lane) decreases, the composite risk indicator increases. When Indicator B (magnitude of speeding) increases, the composite risk indicator increases. We see a sudden increase in the composite risk value when the binary Indicator C (SSD) changes from zero (safer) to one (riskier). When Indicator D (direct

distance) decreases, the composite risk value increases. All these results are as we would expect.

- The proposed risk analysis algorithm appears to capture police officer behavior successfully. For example, if a police officer moves toward the traveled lane, the composite risk indicator increases. If a police officer moves away from the traveled lane, the composite risk indicator decreases. If police officer does not move at all, the composite risk indicator does not change (assuming that the other indicators are constant). All these results are as we would expect.

- The proposed risk analysis algorithm appears to capture driver behavior successfully. Indicator B (magnitude of speeding) directly reflects driver behavior as the value of Indicator B is higher for an aggressive driver (a driver traveling above the posted speed limit) and the higher value results in a higher composite risk indicator. Our algorithm also captures driver behavior indirectly. If a driver rapidly increases his speed, indicator D (direct distance) decreases rapidly and results in increased risk. These results are as we would expect.

- The proposed risk analysis algorithm appears to be successful in using indicator C (SSD) and indicator D (direct distance) to capture the spatial and temporal interaction between officer and vehicle. As a vehicle approaches a police officer, the composite risk indicator increases (indicator C changes from zero to one, and indicator D decreases) and when the vehicle passes safely, the composite risk indicator changes to zero. These results are as we would expect.

- The study used near miss collisions identified by driver braking behavior to validate the results of the proposed fuzzy rule-based algorithm. The proposed algorithm was 78 % successful in detecting near miss collisions, a promising result.
- The proposed risk analysis algorithm successfully reflected slow and rapid changes in officer behavior, driver behavior, and the spatial and temporal interaction between officer and vehicle and successfully reflected the changes in the second by second values of the composite risk indicator.

6.4. Conclusions for Chapter 4

Most existing smart protection systems for pedestrians working on roadways (e.g., work zone workers and police officers) use only one input indicator (e.g., intrusion of work zone/safety area by a passing vehicle) and apply a simple risk analysis approach (compare with a fixed threshold) to identify approaching vehicles that create a risk for roadway based workers and then provide workers with an early warning. A few studies have developed sophisticated machine learning algorithms (a popular type of AI algorithm) to combine multiple input indicators into a composite risk indicator that estimates the risk associated with collisions, but these studies focused on vehicle-to-vehicle collisions rather than pedestrian-to-vehicle collisions (Wang et al., 2015; Xiong et al., 2019). No research has yet investigated the possible development of machine learning algorithms designed to conduct the risk analysis required for a smart protection system based on a combination of multiple input indicators.

Firstly, we proposed a smart protection system designed to reduce the risk of pedestrian-to-vehicle collisions (police officer in our case). The proposed system involves three phases:

Phase I (detection). A radar system installed in a police officer's car detects the positions (i.e., x- and y-coordinates), trajectories, and speeds of passing vehicles and police officers in real-time (i.e., second by second);

Phase II (risk analysis): The system uses the seed information from Phase I to develop multiple input indicators to capture police behavior, driver behavior, and the spatial and temporal interaction between the police officer and passing vehicle. A risk analysis algorithm using a sophisticated AI algorithm is then developed to combine the input indicators and estimate a composite risk indicator; and

Phase III (warning and communication). The system provides early warning of high-risk situations to police officers working on foot on the roadway.

This study focused on Phase II (risk analysis). We developed a semi-supervised machine learning algorithm to combine four input indicators (lateral distance from police officer to traveled lane, magnitude of speeding, stopping sight distance, and direct distance) into a single composite risk indicator to estimate the risk of a collision. The composite risk indicator in this study had three labels: low-risk, medium-risk, and high-risk. We also applied an unsupervised machine learning algorithm to compare with our semi-supervised machine learning algorithm.

The main findings of chapter four included the followings:

- No datasets exist for vehicle collisions to pedestrian roadway workers which provides trajectories of pedestrian and vehicle several seconds prior to a collision. Furthermore, our advanced protection system is not yet implemented to collect such data. This study collected a large amount of data from a radar where vehicle drivers were walking along the shoulder lane to look into possible mechanical issues in their vehicles. This study found an

approach (driver braking behavior) to identify several near miss collisions. We considered this data as labeled data. The rest of our data is unlabeled data where the risk label is unknown.

- The labeled data were insufficient to design an algorithm. It was necessary to develop semi-supervised ANN algorithms (which used the proposed fuzzy rule-based algorithm discussed in Chapter 3) to give labels to the unlabeled data. We should state that earlier research used a simple risk analysis, i.e., they compared an input indicator with a fixed threshold to differentiate a safe event from a risky event. In the earlier research, it was not possible to combine more than one input indicator to capture the diverse possibilities involved in collision risk.

- Our study developed three semi-supervised ANN algorithms and three unsupervised Fuzzy C-Means algorithms. The results showed that the semi-supervised ANN algorithm could combine the four input indicators, classify the risk of a collision into low, medium and high, and successfully capture police officer behavior, driver behavior, and the spatial and temporal interaction between pedestrians and vehicles. This algorithm showed a high level of performance: 94% accuracy when estimating the three risk labels, and 88% success when identifying near miss collisions.

- The Fuzzy C-Means algorithm showed a low level of performance: 33% accuracy when estimating the three risk labels. The Fuzzy C-Means algorithm tended to estimate many datapoints as high-risk labels. Numerous alerts would result in many unnecessary alarms and police officers coming to ignore the warnings. Compared to the Fuzzy C-Means algorithm, the semi-supervised ANN assigned fewer datapoints as high-risk. As a result, the semi-supervised ANN only alerted police officers to high-risk events.

- The study found that the proposed semi-supervised algorithm could generate a warning three to four seconds before a near miss collision. This warning period is in line with the early warning period suggested in the literature and provides an acceptable amount of time for a police officer to react to a potential collision risk.

6.5. Conclusions for Chapter 5

Real time monitoring systems for physical distancing detection on segments such as urban sidewalks are a new CAS during COVID 19. The existing systems use only one input indicator (i.e., distance between two pedestrians) and apply a simple risk analysis approach (compare with a fixed threshold; for example, 2m) to identify a safe and risky situation and evaluate how well physical distance is being achieved on segments (level of pedestrian physical distancing (LPPD)). However, the existing studies neglect two other important factors which are density and exposure time. This study proposed a pedestrian physical distance monitoring system to reduce the risk of pedestrian-to-pedestrian collision. The study involved two phases:

Phase I (detection). The monitoring system (e.g., a camera) detects the positions (i.e., the x- and y-coordinates), trajectories, and speeds of passing pedestrians in real-time (second by second); and

Phase II (risk analysis). The system uses the seed information from Phase I to develop a risk analysis and develop a sophisticated AI algorithm that evaluates and monitors a segment (sidewalks in our case) in terms of level of physical distancing. The evaluation can be used to select the most appropriate mobility interventions.

This study focused on Phase II (risk analysis). The risk analysis involved three-steps. In the first step, Indicator A (pedestrian space) was developed to capture and quantify pedestrian

density, and Indicator B (compliance rate) was developed to capture and quantify pedestrian exposure time. In the second step, the indicators were combined to create LPPD includes six levels of density (ranging from "A" to "F") each level having three sub-levels (+, 0, -) to capture exposure time. The goal of LPPD is to help decision makers to select and implement the most appropriate mobility interventions include slow streets program (completely prohibiting non-essential vehicular traffic on designated urban streets to provide significantly more space for active transportation), education program (e.g., social media advertisements and outdoor signage) and enforcement program (e.g., tickets and fines for physical distancing violators). Finally, in the third step, an artificial intelligence (AI) algorithm was developed to combine the two input indicators and estimate a composite risk indicator that could estimate the relative risk of viral transmission on the basis of the different physical distances between pedestrians. The value of the composite risk indicator in this study ranged from zero (low risk) to one (high risk).

The study used PTV Viswalk simulation software to develop a virtual pedestrian walkway with a base (default) scenario and two test scenarios (Scenarios 1 and 2).

In the base scenario, pedestrians could walk randomly without attempting to maintain the required physical distancing. The flow rate was set at 25p/min/m.

Scenario 1 was created to demonstrate the impact of change in Indicator A (pedestrian space). The flow rate was (35p/ min/m) which allowed all pedestrians to walk randomly without attempting to maintain physical distancing.

Scenario 2 was created to demonstrate the impact of change in Indicator B (compliance rate). The flow rate was 25p/min/ m (the same as the base scenario). All pedestrians in this scenario were forced to maintain 2 m of physical distancing.

The main findings of chapter four included followings:

- In the case of input indicator A (pedestrian space), comparing the base scenario with scenario 1 showed that the increase in the pedestrian flow rate from 25p/min/m in the base scenario to 35p/min/m in scenario 1 decreased Indicator A (pedestrian space) from 3.23 m^2/p in the base scenario to 2.36 m^2/p in scenario 1. The LPPD for the base scenario was D0, but the decrease in Indicator A (pedestrian space) in scenario 1 reduced the LPPD to E0. We concluded that the LPPD properly reflected the changes we made in indicator A (pedestrian space).

- In the case of input indicator B (compliance rate), comparing base scenario with scenario 2 showed that forcing pedestrians to maintain 2 m of physical distancing significantly increased indicator B (pedestrian space) (from the 39th percentile in the base scenario to the 87^{th} percentile in scenario 2). The LPPD for the base scenario was D0, but the increase in Indicator A (pedestrian space) changed the LPPD to D+ in scenario 2. We concluded that the LPPD properly reflected the changes we made in indicator B (pedestrian space).

- We concluded that the LPPD could be used to evaluate and select the most appropriate mobility interventions. To improve the LPPD of scenario 1 from E0 to the base scenario's D0, we can modify the surface infrastructure (by introducing, for example, a slow street program) to increase sidewalk capacity, i.e., to increase the space available for pedestrians. If, however, we wish to improve the LPPD of the base scenario (D0) to scenario 2′s D+, increasing sidewalk capacity may not be the best approach. In this case, we need to focus on mobility interventions that influence pedestrian walking behavior, for example, educational programs or enforcement programs.

- We developed a 3-D risk surface that showed that a decrease in Indicator A and/or Indicator B increased the composite risk indicator. A small decrease in Indicator I led to a relatively large increase in the estimated risk. Changes in Indicator II were less sensitive and did not lead to such dramatic changes in the estimated risk. We concluded that reducing density by improving capacity was the most effective way to reduce the relative risk of possible viral infection (although such an approach is expected to be more expensive).

6.6. Thesis Contributions

This thesis makes a major research contribution to the improvement of existing risk analysis algorithms in collisions avoidance systems through its investigation of four objectives:

1) Develop a fuzzy rule-based algorithm for a next generation roadside animal detection system (Chapter 2);

2) Develop a fuzzy rule-based algorithm for a smart protection system to reduce the number of collisions with police officers on duty on the roadway (Chapter 3);

3) Develop a semi-supervised machine learning algorithm for a smart protection system to reduce the number of collisions with police officers on duty on the roadway (Chapter 4); and

4) Develop a risk analysis approach to evaluate physical distancing on urban sidewalks (Chapter 5).

The discussion that follows focuses on the differences in the research contributions made by each objective.

Chapter 2: Develop a fuzzy rule-based algorithm for a next generation roadside animal detection system

- The animal detection system study developed a rule-based AI risk analysis algorithm (a fuzzy rule-based algorithm) to assist drivers by providing a driver with warning of a potential animal-to-vehicle collision.
- The study collected real-world microscopic trajectory data (i.e., x and y coordinates and vehicle speed) of 35 events recorded by a RADS located on a two-lane rural highway (Trans-Canada HWY 1) in Glacier National Park, British Columbia, Canada. Each event included the movement of one or more animals and the movement of one or more vehicles and each event occurred during the same time window of at least 10 seconds.
- The study relied on past studies and/or expert opinion to build and train a fuzzy rule-based algorithm that could be applied to a limited amount of data.
- The study was designed to disseminate a continuous real-time warning using different levels of warning (e.g., low, medium, and high) as appropriate depending on the temporal and spatial interactions between animal and driver while the driver was driving through the monitored area.
- The study generated zero risk (a safe situation) where a stopping criterion met by fuzzy rule-based algorithm. The stopping criterion is the moment when a vehicle passes the nearest point of conflict with an animal (see Figure 2-2) and the animal cannot follow the vehicle even at the animal's maximum running speed. The stopping criterion eliminates unnecessary warnings and ensures that the driver receives only necessary warnings.

- The study used sensitivity analysis to show the adequacy of the proposed risk analysis algorithm by providing 3D surfaces representation of the estimated change in the relative value of risk by changes in input indicators.

Chapter 3: Develop a fuzzy rule-based algorithm for a smart protection system to reduce the number of collisions with police officers on duty on the roadway

- The study developed a rule-based AI risk analysis algorithm (a fuzzy rule-based algorithm) to warn a police officer to take evasive action to avoid a potential police officer-vehicle collision.
- The study used data from a section of Highway 416 near Kemptville, Ontario to obtain real-world microscopic trajectory data between pedestrians and vehicles for 51 events. Each event occurred during the same time window of at least 10 seconds.
- The study relied on past studies and/or expert opinion to build and train a fuzzy rule-based algorithm. The calibration of the algorithm was different from the calibration of the algorithm used for the animal detection system study in Chapter 2.
- The study did not aim to disseminate different levels of warning to the police officer. The level of warning might change based on temporal and spatial interactions between police officer; however, this study disseminated only the high-risk level event as a warning to police officer which reduces the level of distractions for police officers who are doing different activities on the road (e.g., ticketing, speed enforcement).
- The study generated zero risk (a safe situation) where a stopping criterion met by fuzzy rule-based algorithm. The stopping criterion was the moment when a vehicle passed by the from police officer.

- The study validated the proposed risk analysis algorithm by using driver hard braking behaviour (indicated by a severe deceleration rate) to identify near miss collisions. This approach identified 14 near miss collisions between a pedestrian (police officer) and vehicle. The fuzzy rule-based algorithm was 78% successful in identifying these near miss collisions.

Chapter 4: Develop a semi-supervised machine learning algorithm for a smart protection system to reduce the number of collisions with police officers on duty on the roadway

- The study developed a learning-based AI risk analysis algorithm (i.e., a semi-supervised machine learning algorithm) to warn a police officer to take evasive action to avoid a potential police officer-vehicle collision.
- The study used data from a section of Highway 416 near Kemptville, Ontario to obtain real-world microscopic trajectory data between pedestrians and vehicles for 70 events. Each event occurred during the same time window of at least 10 seconds.
- The study relied on input data and near miss collisions data to train and build the semi-supervised machine learning algorithm.
- The study also applied an unsupervised machine learning algorithm that classified data into different risk levels according to similarities in the data. The results were compared with the results obtained from the semi-supervised machine learning algorithm.
- The system aimed to disseminate only the high-risk level event as a warning to police officer.
- The study clearly showed that the semi-supervised machine learning algorithm can provide warning of a potential near miss collision 3 to 4 seconds in advance, an adequate time for a police officer to take evasive action.

- The study used accuracy as a performance measure for validating the proposed machine learning algorithms. The unsupervised machine learning algorithm that classified data into different risk levels according to similarities in the data was only 33% accurate when estimating the risk levels. The semi-supervised machine learning algorithm had an accuracy of the 94% when estimating different risk levels and 88% when identifying near miss collisions. This was a 10% improvement compared to the animal detection system discussed under Chapter 2.

Chapter 5: Develop a risk analysis approach to evaluate physical distancing on urban sidewalks

- The study developed a new level of pedestrian physical distancing (LPPD) measure to evaluate how well physical distance is maintained on urban sidewalks.
- The study developed a virtual walking environment in PTV Viswalk and collected microscopic trajectory data between pedestrians.
- The study took exposure time into account by proposing a new indicator (compliance rate) to capture the amount of time that pedestrians spent not maintaining physical distancing.
- The study used a 3D surface representation of the estimated change in the relative value of risk by changes in pedestrian space and compliance rate.
- The study provided decision makers with a basis for evaluating the effects of different mobility interventions (such as educational programs and temporarily or permanently replacing curbside parking spaces or curbside lanes with widened pedestrian walkways and such educational programs using outdoor signage and social media) on physical distancing and then selecting the most appropriate mobility intervention(s).

6.7. Limitations of this Study and Recommendations for Future Research

This section discusses limitations of this study and opportunities for future research to improve and extend the usefulness of risk analysis algorithms.

Studies designed to develop risk analysis algorithms to reduce the risk of a collision face the major challenge of limited data. The problem arises because, statistically, actual collisions and near miss collisions are rare and random events. As a result, this study selected and developed certain types of AI algorithm for the risk analysis largely on the basis of data availability and uncertainties.

This study was the first attempt to develop risk analysis for the three selected CASs. We selected and developed AI algorithms mainly due to lack of large number of datasets. We strongly encourage future studies to collect more data and conduct more research on developing other AI algorithms including approaches such as Bayesian Inference Network, Random Forest, Support Vector Machine, etc. to compare the performance of the different algorithms in generating a more accurate composite risk indicator.

This study did not include the impact of environmental factors (weather/road conditions) since our data did not include such information. We recommend future studies to consider environmental factors. One way to collect such information is using online sources, for example, OpenWeatherMap, but these sources provide real-time data making it difficult to access historical information.

As this study used equal weights for all input indicators, future studies may consider an improved approach (e.g., a multi-criteria decision-making approach) and produce different weights for different indicators.

The composite risk indicator in this thesis was generated second by second because of limitations in existing CASs. We believe that this interval it too coarse as the position of road users (e.g., vehicles and pedestrians) can change considerably within a second. We recommend future studies to collect data using fractions of a second (e.g., 0.1 second), perform the risk analysis and compare the results. Data provided by the fraction of a second could significantly improve the performance of the risk analysis algorithms.

We faced the shortage of real-world dataset, and the real-world dataset was the data for vehicle drivers who were walking along the road or shoulder lane. As traffic may behave different around police officers, we think that virtual reality (VR) technology can be used to generate a large number of datasets from a real-time virtual reality environment. VR uses computer technology to create a simulated environment where road users wear VR glasses to explore and react to the environment in 360 degrees. For example, in the study on pedestrian-to pedestrian collisions (Chapter 5), we could create a scenario in which a police officer is conducting roadside activities (e.g., ticketing) and a vehicle is approaching. Both police officer and driver need to wear VR glasses. Now, we could define many different scenarios and collect data some of which could potentially include collisions/near miss collisions. An interesting feature of VR technology is the 1:1 ratio between the virtual environment and the real-world, i.e., the object scales in VR are the same as in the real-world (e.g., one meter in VR is exactly one meter in real-world environment). This feature provides a realistic experience for the police officer and driver and produces precise data for researchers. The use of VR is also safer, cheaper and less time consuming than the use of field tests.

Future studies need to analyze the impact of other input indicators (e.g., weather conditions and road conditions) on the risk of a collision. We suggest that future studies might create

simulation scenarios in virtual reality to create a wide range of scenarios (e.g., different weather conditions) and a large number of datasets that include near miss and collisions data. With a large number of datasets, different techniques (e.g., regression) could be used to analyze the impact of other important indicators that might be incorporated into the algorithm

Defining the warning threshold for a smart protection system for police officers is very important and involves two challenges: i) defining an appropriate threshold for differentiating high-risk from medium/low-risk, and ii) understanding the issues involved in the perception and reaction time of a police officer.

i) Appropriate threshold for differentiating high-risk from medium/low-risk: In this study, we evenly dividing the estimated risk into three levels (low, medium, and high-risk). This assumption resulted in a 78 % success rate when identifying near miss collisions using the fuzzy-rule based algorithm. We then applied a semi-supervised machine learning algorithm to further improve the performance. The semi-supervised machine learning algorithm had an 88% success rate in identifying near miss collisions. Collecting and adding additional near miss collisions data to the machine learning algorithm is expected to improve the performance even further. The auto industry is rapidly developing connected autonomous vehicles (CAVs) which could deliver even greater benefits in the proposed smart protection systems by using advanced wireless communication to automatically change the travel speed and trajectory of approaching vehicles and avoid possible collisions.

ii) Perception and reaction time of police officer: The smart protection system should disseminate an alarm only when a high-risk circumstance is expected, and must take into account the perception and reaction time of police officers. A few studies have evaluated the perception and reaction time required for work zone workers or other pedestrians to take evasive action to prevent a collision

(Hines et al. 2015; Lin et al. 2016). We suggest that future studies should be undertaken to determine police officer perception and reaction times and should consider a range of factors (e.g., day/night time).

Defining the warning threshold for a roadside animal detection system is not important since the goal of the system is to improve the situational awareness of drivers in the entire of the corridor. The system provides a driver with real-time feedback based on the temporal and spatial interactions between vehicles and animals.

Due to possibilities of large uncertainties in the estimated risk value, we suggest that future studies generate a composite risk indicator that shows a confidence interval instead of a single value. In this regard, the confidence interval approach can be interpreted as the lower and upper risk bounds (Moore, R. E. 1962; Yue, Z. 2011). Appendix B provides an example of demonstrate how the interval approach works in our problem.

Previous studies designed to estimate collision risk using machine learning algorithms allocated an equal weight to different levels of risk. For example, Wang et al. (2015) proposed a K-Means method to classify vehicle-to-vehicle interactions into three different risk levels (low, medium, high), and used equal weights for each risk level. Xiong et al. (2019) used a spectral clustering and K-Means method to classify vehicle-to-vehicle interactions into three risk levels (low, medium, high). They also used equal weights for estimating the accuracy of the machine learning algorithm. Similar to previous studies, we used equal weights for each risk level. As this study used equal weights for all risk levels (low, medium, high), future studies may consider an improved approach (e.g., a multi-criteria decision-making approach) and produce different weights for different risk levels.

This thesis did not consider the consequences (severity) of a collision (e.g., fatal, injury, property damage only). Potential indicators for capturing the consequences (severity) of a collision include speed differences, mass differences, relative angle, and fragility of the road users involved in the situation (Johnsson et al., 2018). Future studies could test these indicators to explore the consequences of collisions.

The results of this thesis could be helpful and beneficial in other transportation safety areas including CASs for commercial Unmanned Aerial Vehicles (UAV), road maintenance during the winter season, the determination of dynamic posted speed limits in smart cities, and the determination of pedestrian level of service (PLOS):

- CASs in commercial Unmanned Aerial Vehicles (UAV). Commercial (UAV), known as drones, have many applications including traffic and road monitoring, road construction, environmental monitoring, law enforcement, etc. UAVs can be controlled by a pilot on the ground or by computer systems in the case of a fully autonomous UAV (Barmpounakis et al., 2016). One of the most important challenges for UAVs is the risk of mid-air collisions (Skowron et al., 2019). Mid-air collisions can be with static obstacles such as buildings, trees, and mountains or moving obstacles such as other UAVs, tower cranes, birds, and even with airplanes. CASs have been developed for UAVs (Aswini et al., 2018). CASs installed in a UAV are equipped with sensors (radar or Lidar) to detect obstacles or objects. The systems use various approaches (probability flow, analytical, numerical, and Monte Carlo methods) in the risk analysis designed to predict the future path (and potential conflict) between the UAV and an object (Skowron et al., 2019). Predicting the future conflict between UAV and another object might be one the many contributing factors which needs to be considered. Some contributing factors (the type of object, the behavior

of the object, etc.) may require more than a prediction of the future path. For example, in the case of mid-air collisions with birds, UAV can not simply predict the future path of birds as the path is unpredictable. In the case of a UAV which has pilot, predicting the future path of a UAV is not easy. UAV research, like the research presented in this study, faces uncertainties and limited data availability especially drone technology is still new. We believe that the research presented in this thesis can help UAV researchers to develop risk analysis algorithms that combine multiple contributing factors.

- Road maintenance during the winter season. Road maintenance during the winter season is a safety critical and resource demanding operation. One of the key activities is determining the condition of the road surface in order to prioritize roads and allocate cleaning efforts such as plowing or salting. Two main approaches are currently used: (1) patrolling the roads to perform on-site inspections (old approach), and (2) visual examination of roadside camera images by trained personnel (advanced approach) (Carrillo et al., 2020). In the second approach (visual examination), trained personnel observe camera images and consider other factors including air temperature, relative humidity, pressure, and wind speed. Instead of using human (trained personnel) intelligence, we suggest designing an artificial intelligence algorithm which could combine different contributing factors and create a composite indicator to evaluate the condition of the road surface. This composite indicator could then be used to prioritize roads and allocate cleaning efforts.

- Determination of dynamic posted speed limits in smart cities. In smart cities where vehicles are fully autonomous, it would be possible for posted speed limits to be dynamic, i.e., to change with the situation. Several factors (input indicators) affect posted speed limit. They

include the number of lanes, traffic condition, road surface, existence of sidewalks, etc. As some factors may change in real-time, a dynamic posted speed limit may have value. Combining the factors (input indicators) and creating a composite indicator (a posted speed limit indicator) could be a promising approach.

- Pedestrian level of service (PLOS). Practitioners use (PLOS) to describe the operational conditions of pedestrian flows on urban sidewalks (TRB, 2016). The US Highway Capacity Manual (HCM) provides PLOS guidelines for practitioners. The guidelines use various separate indicators (e.g., pedestrian space, pedestrian flow rate, average speed, volume/capacity ratio). For example, a practitioner might start by using pedestrian flow rate to establish the HCM's PLOS for a sidewalk. Separately, in the second attempt, a practitioner might use average speed to establish the HCM's PLOS for the same sidewalk, but might find that the two indicators deliver different results. Future PLOS research could combine different input indicators (e.g., pedestrian space, pedestrian flow rate, average speed, volume/capacity ratio, etc) to create a composite indicator and may find the research presented in this thesis of value.

6.8. Chapter 6 References

Aswini, N., Krishna Kumar, E., and Uma, S. V. (2018). UAV and obstacle sensing techniques–a perspective. International Journal of Intelligent Unmanned Systems, 6(1), 32-46.

Carrillo, J., and Crowley, M. (2020). Integration of roadside camera images and weather data for monitoring winter road surface conditions. arXiv preprint arXiv:2009.12165.

Hines, K., Lages, W., Somasundaram, N., & Martin, T. (2015). Protecting workers with smart e-vest. In Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive

and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (pp. 101-104).

Lin, C. H., Chen, Y. T., Chen, J. J., Shih, W. C., & Chen, W. T. (2016). Psafety: A collision prevention system for pedestrians using smartphone. In 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall) (pp. 1-5). IEEE.

Skowron, M., Chmielowiec, W., Glowacka, K., Krupa, M., and Srebro, A. (2019). Sense and avoid for small, unmanned aircraft systems: Research on methods and best practices. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 233(16), 6044-6062.

Wang, J., Zheng, Y., Li, X., Yu, C., Kodaka, K., & Li, K. (2015). Driving risk assessment using near-crash database through data mining of tree-based model. Accident Analysis & Prevention, 84, 54-64.

Xiong, X., Wang, M., Cai, Y., Chen, L., Farah, H., & Hagenzieker, M. (2019). A forward collision avoidance algorithm based on driver braking behavior. Accident Analysis & Prevention, 129, 30-43.

Zheng, L., Sayed, T., & Mannering, F. (2021). Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. Analytic Methods in Accident Research, 29, 100142.

APPENDIX A

Table A-1 shows five indicators commonly used in the literature: time-to-collision (TTC), post-encroachment time (PET), headway, gap acceptance, and time to accident.

TTC is one of the most widely used indicators (Kuang et al., 2015). TTC is "*the time that remains until a collision between two road users (e.g., two vehicles) would have occurred if the collision course and speed difference are maintained (Hydén, 1996).*" TTC is based on the assumption that two moving objects (e.g., vehicles) are on the path leading to a collision. If this assumption is met, TTC could capture collisions including rear-end and turning collisions. Notice that most studies that used TTC were evaluating safety at intersections.

PET is defined as "*the time between the moment when the first road user leaves the path of the second and the moment when the second reaches the path of the first Allen et al. (1978).*" PET has also been used to evaluate safety at intersections, largely for right angle or left angle.

Headway is defined as "*the time between the front of the lead vehicle passing a point on the roadway and the front of the following vehicle passing the same point (Vogel, 2003).*" Headway has been used mainly to evaluate rear-end collisions between two vehicles at or near intersections, for example, when the lead vehicle suddenly stops due to the red light displayed on traffic signals.

Gap acceptance is defined as *"a safe opportunity or "gap" in the traffic which a driver looks to before deciding to enter the intersection (Robbins et al., 2018)."* Gap acceptance has been used mainly to evaluate and right angle or left angle collisions at intersections.

Time to accident is defined as *"time remaining to a collision when an evasive action is taken by the relevant road user (Hydén, 1987)".* Time to accident has been mainly used for turning collisions at intersection.

Table A-1. Safety indicators in the literature

#	Safety Indicators	Road configuration	References	Collision type
1	Time to collision (TTC)	▪ T and 4-leg intersection	(Sayed, 1992; Sayed et al., 1994; Dijkstra et al., 2010; Essa and Sayed., 2015)	Mainly for turning, rear-end
		▪ Intersection	(Eisele and Frawley, 2004; Eisele and Toycen, 2005)	
		▪ T-junction	(Archer and Young, 2010b)	
		▪ Link with ramp	(Vanderschuren, 2008)	
2	Post encroachment time (PET)	▪ T-junction	(Archer and Young, 2010a)	Mainly for right angle or left angle collisions
		▪ 3 and 4-leg intersections	(Caliendo and Guida, 2012)	
3	Headway	▪ 4-leg intersection	(Xin et al., 2008)	Rear-end
		▪ Signalized intersection and level crossing	(Kim, 2014)	
4	Gap acceptance	▪ T-junction	(Cooper and Ferguson, 1976; Robbins et al., 2018)	Turning and right angle or left angle collisions
5	Time to accident	▪ T-junction	(Archer, 2005)	Turning

In this study, we were aware that animals or pedestrians on roadway segments may be moving along the roadside and may be involved in various types of collision (e.g., a sideswipe collision). When we designed our risk analysis algorithm, we considered different types of collision that occur on road segments and selected appropriate input indicators.

Appendix A References

Archer, J. (2005). Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections Doctoral dissertation, Royal Institute of Technology Stockholm, Department of infrastructure division for transportation and logistics, Centre for Transportation Research, CTR

Archer, J., and Young, W. (2010a). Traffic Microsimulation Approach to Estimate Safety at Unsignalized Intersections. Transportation Research Board 89th Annual Meeting, Washington DC, United States

Archer, J., and Young, W. (2010b). Signal treatments to reduce the likelihood of heavy vehicle crashes at intersections: microsimulation modeling approach. Journal of Transportation Engineering, 136(7), 632-639.

Caliendo, C., and Guida, M. (2012). Microsimulation approach for predicting crashes at unsignalized intersections using traffic conflicts. Journal of Transportation Engineering, 138(12), 1453-1467.

Cooper, D. F., and Ferguson, N. (1976). Traffic studies at T-Junctions. 2. A conflict simulation Record. Traffic Engineering & Control, 17(7), 306-309.

Dijkstra, A., Marchesini, P., Bijleveld, F., Kars, V., Drolenga, H., and Van Maarseveen, M. (2010). Do calculated conflicts in microsimulation model predict number of crashes? Transportation Research Record, 2147(1), 105-112.

Eisele, W. L., and Toycen, C. M. (2005). Identifying and quantifying operational and safety performance measures for access management: Micro-simulation results. Southwest Region University Transportation Center, Texas Transportation Institute. Report SWUTC/05/167725-1.

Eisele, W. L., and Frawley, W. E. (2004). Estimating the impacts of access management with micro-simulation: Lessons learned. In 6th National Conference on Access Management, Kansas City, Missouri, August (Vol. 29).

Essa, M., and Sayed, T. (2015). Simulated traffic conflicts: do they accurately represent field-measured conflicts? Transportation Research Record, 2514(1), 48-57.

Hydén, C. (1987). The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique. Bulletin Lund Institute of Technology, Department, (70).

Kim, I., (2014) Evaluation of railway level crossing safety using ITS interventions, Doctoral dissertation, School of Civil Engineering, The University of Queensland, Australia.

Kuang, Y., Qu, X., and Wang, S. (2015). A tree-structured crash surrogate measure for freeways. Accident Analysis & Prevention, 77, 137-148.

Robbins, C. J., Allen, H. A., and Chapman, P. (2018). Comparing drivers' gap acceptance for cars and motorcycles at junctions using an adaptive staircase methodology. Transportation Research Part F: Traffic Psychology and Behaviour, 58, 944-954.

Sayed, T. A. (1992). A simulation model of road user behaviour and traffic conflicts at unsignalized intersections (Doctoral dissertation, University of British Columbia).

Sayed, T., Brown, G., and Navin, F. (1994). Simulation of traffic conflicts at unsignalized intersections with TSC-Sim. Accident Analysis & Prevention, 26(5), 593-607.

Vanderschuren, M. (2008). Safety improvements through Intelligent Transport Systems: A South African case study based on microscopic simulation modelling. Accident Analysis & Prevention, 40(2), 807-817.

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. Accident Analysis & Prevention, 35(3), 427-433.

Xin, W., Hourdos, J., Michalopoulos, P., and Davis, G. (2008). The less-than-perfect driver: a model of collision-inclusive car-following behavior. Transportation Research Record, 2088(1), 126-137.

APPENDIX B

If we use A, B, C, and D to represent indicator A, B, C, and D, respectively, we have four definitions of the interval:

Definition 1. The interval number for indicator A is defined as an ordered pair $\tilde{A} = [A^-, A^+] = \{A | A^- \leq A \leq A^+, A^-, A^+ \in R\}$, where A^- is the lower bound value and A^+ is the upper bound value. If $A^- = A^+$, $\tilde{A}$ is obviously equal to a precise value.

Definition 2. The interval number for indicator B is defined as an ordered pair $\tilde{B} = [B^-, B^+] = \{B | B^- \leq B \leq B, B^-, B^+ \in R\}$, where B^- is the lower bound value and B^+ is the upper bound value. If $B^- = B^+$, $\tilde{B}$ is obviously equal to a precise value

Definition 3. The interval number for indicator C is defined as an ordered pair $\tilde{C} = [C^-, C^+] = \{C | C^- \leq C \leq C^+, C^-, C^+ \in R\}$, where C^- is the lower bound value and C^+ is the upper bound value. If $C^- = C^+$, $\tilde{C}$ is obviously equal to a precise value

Definition 4. The interval number for indicator D is defined as an ordered pair $\widetilde{D} = [D^-, D^+] = \{D | D^- \leq D \leq D^+, D^-, D^+ \in R\}$, where D^- is the lower bound value and D^+ is the upper bound value. If $D^- = D^+$, D is obviously equal to a precise value

Now consider matrix (B-1) as a dataset which includes a precise value for each of the four inputs.

$$[A \quad B \quad C \quad D] \tag{B-1}$$

By considering the confidence interval approach , matrix (B-1) will change into matrix (B-2).

$$[[A^-, A^+] \quad [B^-, B^+] \quad [C^-, C^+] \quad [D^-, D^+]] \tag{B-2}$$

Risk value would be the interval value with a lower bound and upper bound generated by matrix (B-2)."

Appendix B References

Moore, R. E. (1962). Interval arithmetic and automatic error analysis in digital computing Doctoral dissertation, Stanford University, Stanford, Calif, USA.

Yue, Z. (2011). An extended TOPSIS for determining weights of decision makers with interval numbers. Knowledge-based systems, 24(1), 146-153.