The Dissertation Committee for Yang Xu Certifies that this is the approved version of the following Dissertation:

DATA-DRIVEN METHODOLOGIES FOR SUPPORTING DECISION-MAKING IN ROADWAY SAFETY AND PAVEMENT MANAGEMENT

Committee:

Amit Bhasin, Supervisor

Carlos H. Caldas

Stephen D. Boyles

Jenny Li

DATA-DRIVEN METHODOLOGIES FOR SUPPORTING DECISION-MAKING IN ROADWAY SAFETY AND PAVEMENT MANAGEMENT

by

Yang Xu

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2023

Dedication

This dissertation is dedicated to the cherished memory of my grandfather, Xingyuan Xu, and to my parents, Wenjun Yang and Weiqiang Xu, and my husband, Hao Wu, for their unconditional love, trust, and support. I also extend my dedication to the remarkable individuals who have supported me throughout my academic journey.

“博学之，审问之，慎思之，明辨之，笃行之。”

——《礼记·中庸》

Acknowledgments

I would like to express my heartfelt appreciation to my esteemed advisor, Dr. Zhanmin Zhang, for his invaluable guidance, patience, and support throughout my doctoral journey. His mentorship played a pivotal role in shaping my research and academic growth. I am grateful for his open-mindedness, which allowed me to explore topics aligned with my interests and inspired me to achieve my research goals.

I extend my deepest gratitude to my current advisor, Dr. Amit Bhasin, for stepping in at a crucial time and serving as my mentor during the final phase of my research. His expertise, valuable insights, and steadfast support have been instrumental in completing this dissertation within a relatively short period.

I am immensely grateful to my dissertation committee members, Dr. Carlos H. Caldas, Dr. Stephen D. Boyles, and Dr. Jenny Li, for their expert guidance, invaluable suggestions, and constructive feedback. Their contributions have significantly enhanced the quality and depth of this dissertation.

I would like to acknowledge all the members of the CRISC lab, who have created a pleasant academic environment and provided valuable support and collaboration. Their expertise and friendship have enriched my research experience. I also want to express my sincere appreciation to Dr. Michael R. Murphy and Dr. Zhe Han for their support and assistance during my research.

I am deeply grateful to all my friends and colleagues who have provided encouragement and support during this challenging journey. Your presence and camaraderie made this experience more enjoyable and memorable. In particular, I want to thank Dr. Joseph Koo for sharing his knowledge and wisdom with my husband and me.

Finally, I am profoundly grateful to my family for always being my best cheerleaders. My mother's courageous triumph over cancer serves as a powerful inspiration, demonstrating the resilience and fortitude required to overcome obstacles in life. I extend my immense gratitude to my father for his exceptional care and commitment to my mother's well-being during the most

difficult times, which allowed me to continue my Ph.D. research. Their enduring love, sacrifices, and belief in my potential have been a constant source of motivation. I am also deeply grateful for the presence and support of my husband and my beloved corgi boy, Yishow (a.k.a. “一休”), who have been my pillars of strength and loyal companions throughout this demanding and rewarding journey.

Abstract

Data-Driven Methodologies for Supporting Decision-Making in Roadway Safety and Pavement Management

Yang Xu, Ph.D.

The University of Texas at Austin, 2023

Supervisor: Amit Bhasin

(Initiated by: Zhanmin Zhang)

There has been a significant rise in the utilization of data-driven methods within the contemporary realm of transportation engineering. This trend is primarily attributed to the limitations associated with experience-based methods, such as subjectivity and non-reproducibility. In contrast, data-driven methods have proven to offer a more objective and effective approach to problem analysis, thereby providing decision-makers with a reliable basis for informed decision-making. This present research focuses on two types of data-driven methodologies: geostatistical analyses utilizing geographic information systems (GIS) and cutting-edge algorithms associated with artificial intelligence (AI). In numerical analysis, data provides a means to gain valuable insights into a problem of interest. While AI-oriented methods have been shown in many studies to be more effective than traditional approaches, the accuracy of the analysis still heavily depends on the quality of the data. This dissertation endeavors to shed light on the pivotal role that data plays in both roadway safety analysis and pavement management. To accomplish this, four distinct studies are

proposed that examine different aspects of data-driven methods. The studies encompass an evaluation of data consistency in motor vehicle crash databases, the identification of crash hot spots within a road network, a synthesis of advancements in the application of AI algorithms to various activities of pavement management, and an exploration of the relationship between pavement conditions and roadway safety using AI-oriented methods. The knowledge acquired from these studies serves as a foundation for future research, advancements, and the adoption of innovative approaches to enhance the efficiency of safety analysis and pavement management. This research ultimately facilitates informed decision-making, effective resource allocation, and the implementation of cost-effective interventions to enhance roadway safety and optimize pavement management practices.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Research Background

Transportation infrastructure has consistently served as the cornerstone of a nation's economy, facilitating the movement of goods and individuals between locations. This encompasses a wide range of physical structures such as roads, railroads, airports, ports, bridges, tunnels, pipelines, public transit facilities, and others. The impact of transportation infrastructure on economic activities and social interactions is substantial. According to the United States Transportation and Infrastructure Committee (2023), "there are more than 4,000,000 miles of public roads, 19,700 civil airports, and over 138,000 miles of freight rail throughout the United States." The extensive network of transportation infrastructure has been instrumental in connecting people from geographically disparate locations, allowing them to collaborate toward a shared vision of building a better world.

As the most frequently used transportation infrastructure, roads play a significant role in shaping the way we live, work, and move around our communities. Without roads, our normal activities such as commuting to work, shopping for groceries, visiting doctors, and traveling to other places would become incredibly challenging if not impossible. Roads provide essential access to businesses, services, and social interactions that are vital to our well-being and economic prosperity. Figure 1.1 presents the total length of U.S. public roads from 1980 to 2020.

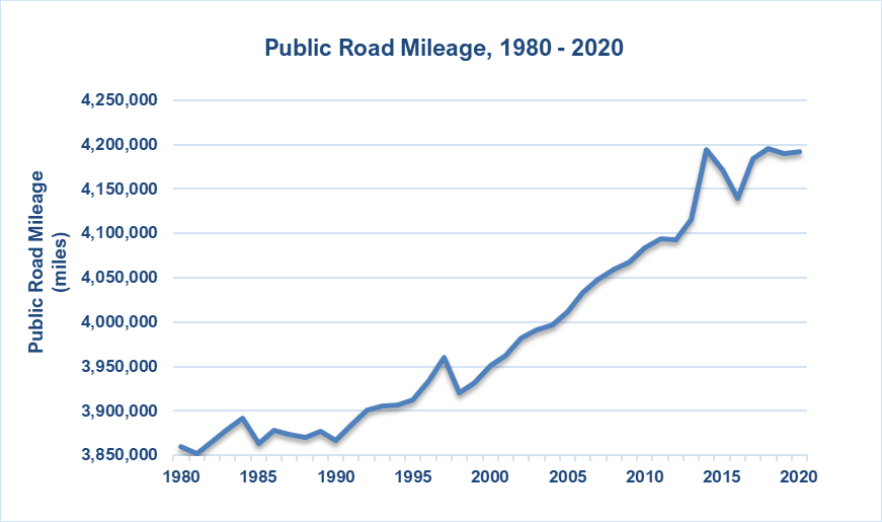


Figure 1.1 U.S. Public Road Mileage: 1980-2020[1]

One of the crucial functions of our transportation infrastructure systems is to ensure the safety and mobility of the traveling public. According to the World Health Organization (WHO), road traffic crashes result in more than 1.3 million fatalities and over 20 million non-fatal injuries every year around the world. Road traffic injuries have been ranked as a leading cause of death for children and young adults aged 5-29 years (WHO 2022). The economic losses caused by road traffic crashes – for most countries – are around 3 percent of the national gross domestic product (WHO 2022). In total, a study estimated that traffic injuries will cost the world economy $1.8 trillion (in 2010 USD) from 2015 to 2030; an equivalent to an annual tax of 0.12% on the global gross domestic product (GDP) (Chen et al. 2019).

[1]Data Source: FHWA, Highway Statistics 2020.

1.1.1 Roadway Safety

Motor vehicle travel has provided an unparalleled degree of mobility in the United States. The American Society of Civil Engineers (ASCE) reported that U.S. highways and roads facilitate the transportation of 72% of the country's goods, valued at approximately $17 trillion. However, according to the Centers for Disease Control and Prevention (CDC), the U.S. had the highest population-based motor vehicle crash death rate, with 11.1 fatalities per 100,000 population, which was 2.3 times greater than the average rate of 4.8 per 100,000 population among 29 high-income nations (Yellman 2022).

Over the past decade, transportation incidents in the United States have resulted in a total death toll exceeding 370,000. Among these incidents, roadway fatalities constitute the largest component (USDOT 2022). An estimate shows that nearly 95 percent of transportation deaths occurred on roadways (USDOT 2022). The CDC identifies motor vehicle crashes as one of the leading causes of death; more than 100 people are killed every day on U.S. roadways (CDC 2020). A study conducted by the National Highway Traffic Safety Administration (NHTSA) indicates that the total economic and societal costs attributed to motor vehicle crashes were 836 billion dollars in 2010 (Blincoe et al. 2015).

The trends in fatalities and fatality rates from 1994 to 2020 are depicted in Figure 1.2. A steady fluctuation in the number of deaths is observed from 1994 to 2006, followed by a decline starting in 2007 and reaching its lowest point in 2011, with some minor fluctuations from 2012 to 2014. There was then an increase in 2015, followed by a significant rise in 2020. The fatality rate steadily decreased from 1994 to 2010, remained stable with fluctuations from 2011 to 2019, and experienced a significant increase in 2020. Specifically, in 2020, there were more than 5,250,000 policed-reported traffic crashes including 35,766 fatal crashes that occurred across roadways in the U.S., which results in 38,824 fatalities and more than 2,282,000 injured (Stewart 2022, IIHS 2022). More

recently, a NHTSA estimate shows that 42,915 people died in traffic crashes in 2021, a 10 percent increase from 2020. This is the highest number of fatalities since 2005 and the largest increase in the annual percentage of traffic fatalities since 1975 (NHTSA 2022a).

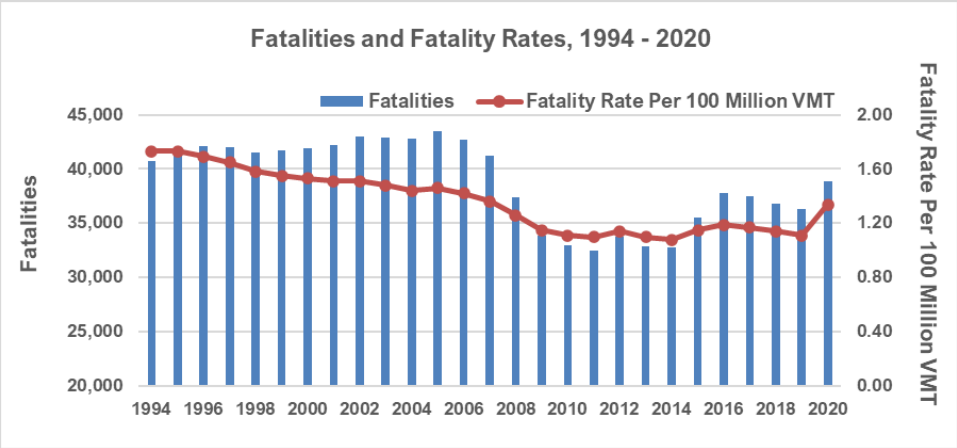


Figure 1.2 Fatalities and Fatality Rate Per 100 Million VMT, 1994-2020[2]

Contributing Factors of Motor Vehicle Crashes

Motor vehicle crashes are complicated events that are usually influenced by various factors. These contributing factors can be broadly classified into five categories: vehicle condition, human behavior, environmental conditions, roadway geometric characteristics, and miscellaneous factors. Specifically, lack of proper maintenance and/or defective/malfunction of key components (i.e., tire, brake, steering, or vehicle lighting failure) are instances of common crash contributors that are related to the condition of vehicles (Murphy et al. 2018). Examples of human behavioral factors that play vital roles in traffic crash occurrence or severity are driver distraction, drug/alcohol intoxication, not wearing seatbelts, and speeding (USDOT 2022). Alcohol-impaired driving is one of the major contributors to motor vehicle crashes on U.S. roadways. Based on the estimates from

[2]Data Source: Fatality Analysis Reporting System (FARS) – NHTSA.

a NHTSA fact sheet, 11,654 people died in alcohol-impaired crashes in 2020 (NHTSA 2022b). Data from the National Safety Council (NSC) shows that the percentage of drivers distracted by hand-held electronic devices has increased by 127% over the past decade (NSC 2022). Environmental conditions (e.g., weather, lighting, and road surface conditions) have a significant impact on the occurrence of crashes. According to USDOT, approximately 21% of traffic crashes across the U.S. are weather-related, resulting in around 5,000 people killed and over 418,000 people injured each year (FHWA 2022). These crashes are either directly caused by adverse weather (e.g., rain, sleet, snow, fog, glare, or severe crosswinds) or indirectly caused by slick road surfaces (e.g., wet, snowy/slushy, icy, or muddy pavement) developed under inclement weather. Roadway geometric characteristics (e.g., horizontal curves, number of lanes, shoulder type and width, median width, superelevation, etc.) are also critical elements affecting crash risk. Among these factors, horizontal curves have been identified as a key contributor to fatal crashes by many previous studies (Albin et al. 2016, Bauer and Harwood 2013, Souleyrette 2011, Hummer et al. 2010). According to Federal Highway Administration (FHWA), about 27 percent of all fatal crashes occurred on or close to horizontal curves (FHWA 2021). Miscellaneous factors include all other elements that can influence the risk of motor vehicle crashes. Examples of such factors include major holidays (e.g., New Year's Day, Independence Day, Thanksgiving, and Christmas), obstructed views (due to buildings or vegetation especially at intersections) or driver blind spots (trucks and cars), road debris, etc. In particular, compared to non-holiday periods, holiday periods typically have more motor vehicle crash fatalities and serious injuries due to increased traffic volume, long-distance travel, more alcohol consumption, and excessive speeding, among others (NHTSA 2019, Liu et al. 2005a, Subramanian et al. 2004, Liu et al. 2005b).

Safety Impacts of Horizontal Curves

As an integral component of roadways, horizontal curves play a significant role in improving the safety and comfort of drivers and passengers by preventing a sharp turn from one direction to another. Over the past decades, several efforts have been made to explore the safety impacts of horizontal curves on crash frequency and severity. In an early study, Glennon et al. (1985) found that highway curves were more likely to be associated with severe, wet-icy, and single-vehicle run-off-road (ROR) crashes when compared with tangent segments. The average crash rate of highway curves was about three times greater than that of tangent segments on the same road; similarly, the average single-vehicle ROR crash rate of highway curves was approximately four times higher than that of highway straight segments. Moreover, the severity of roadway departure crashes on curved segments was significantly higher when compared to tangent segments (Glennon et al., 1985). Torbic et al. (2004) proposed preventive strategies and countermeasures to reduce crashes on U.S. highway curves and concluded that about one-fourth of highway fatalities in 2002 were located at horizontal curves. Another study showed that nearly 40 percent of speeding-related fatalities in the U.S. occurred on curved segments from 1983 to 2002 (Liu et al. 2005c). Hummer et al. (2010) pointed out that horizontal curves in rural areas tended to increase the risk of crashes as compared to all other roads in North Carolina. In terms of crash severity, the study found that crashes on two-lane curved segments had both higher fatality and severe injury rates than crashes on other road segments. Souleyrette (2011) found that the degree of curvature and length of curve had significant impacts on crash rates; specifically, shorter curves tended to experience more crashes compared to longer curves. Bauer and Harwood (2013) identified that crash frequency increased with decreasing horizontal curve length and with decreasing horizontal curve radius. Also, short, sharp horizontal curves, short horizontal curves at sharp crest vertical curves, and short

horizontal curves at sharp sag vertical curves tended to experience higher crash frequencies than other road sections. According to a FHWA report, more than half of U.S. motor-vehicle fatalities in 2013 were caused by roadway departure crashes; further, horizontal curves typically had higher risks of roadway departure crashes than tangent road segments (Albin et al. 2016).

Motor Vehicle Crash Databases

With the aim of promoting traffic safety and protecting the lives of the traveling public, transportation agencies – both at the national and state levels – develop and maintain crash databases to effectively manage reportable motor vehicle crashes. As a national operating administration for highway safety, the mission of NHTSA is to "save lives, prevent injuries, and reduce economic costs" resulting from motor vehicle crashes. To accomplish this mission, NHTSA collected crash data across the country. These databases include Fatality Analysis Reporting System (FARS), Crash Report Sampling System (CRSS), Crash Investigation Sampling System (CISS), Special Crash Investigations (SCI), Non-Traffic Surveillance (NTS), Crash Injury Research & Engineering Network (CIREN), etc. Among these data, FARS is one of the most widely used databases for conducting nationwide safety analysis. It contains a census of fatal crashes on public roadways within the 50 States, the District of Columbia, and Puerto Rico since 1975. Fatal crashes are defined as crashes that result in the death of at least one person within 30 days of the crash. FARS was developed to evaluate highway safety performance, identify traffic safety problems, and assess the effectiveness of highway safety standards and programs. Data sources of FARS include police crash reports, state vehicle registration files, state driver licensing files, state highway department data, vital statistics data, death certificates, and emergency medical service reports, among others (NHTSA 2022c). In

addition to the nationwide crash databases, state transportation agencies collect and maintain statewide crash databases. Examples of such databases are Crash Records Information System (CRIS) in Texas, Crash Analysis Reporting System (CAR Online) in Florida, Georgia Electronic Accident Reporting System (GEARS), and Statewide Integrated Traffic Records System (SWITRS) in California.

The significance of motor vehicle crash data cannot be overstated, as it serves as the foundation for identifying locations with high crash risk and developing effective countermeasures to prevent similar accidents in the future. Additionally, the utilization of historical crash records in crash prediction models further underscores the importance of ensuring the quality and reliability of crash data. Thus, transportation agencies must make it a top priority to maintain the validity of crash data.

1.1.2 Infrastructure Asset Management

Typically, transportation infrastructure refers to physical structures that support the movement of people, products, and resources. These structures are key components of transportation assets which also include transportation equipment such as motor vehicles, ships, and aircraft. Transportation infrastructure plays a crucial role in the economic framework of nations by facilitating trade, commerce, and travel. It helps to connect people and communities, leading to increased social and economic opportunities. Effective management of transportation infrastructure can significantly help to lower costs, improve efficiency, and expand market accessibility, which are all essential factors in promoting economic growth and development.

Transportation infrastructure not only exerts a significant impact on economies and societies but it is also considered a valuable investment – perhaps the most valuable asset – to many countries. According to the U.S. Bureau of Economic Analysis (BEA), the

estimated net value of the publicly owned roadway network, including highways and streets, was $4.54 trillion in 2021 (BEA 2022), as demonstrated in Figure 1.3. Both the public and private sectors invest substantial funds into the construction of new roadways and the preservation of existing infrastructure, with the annual investment in U.S. highways and streets surpassing $100 billion since 2018, as reported by the BEA (BEA 2022).

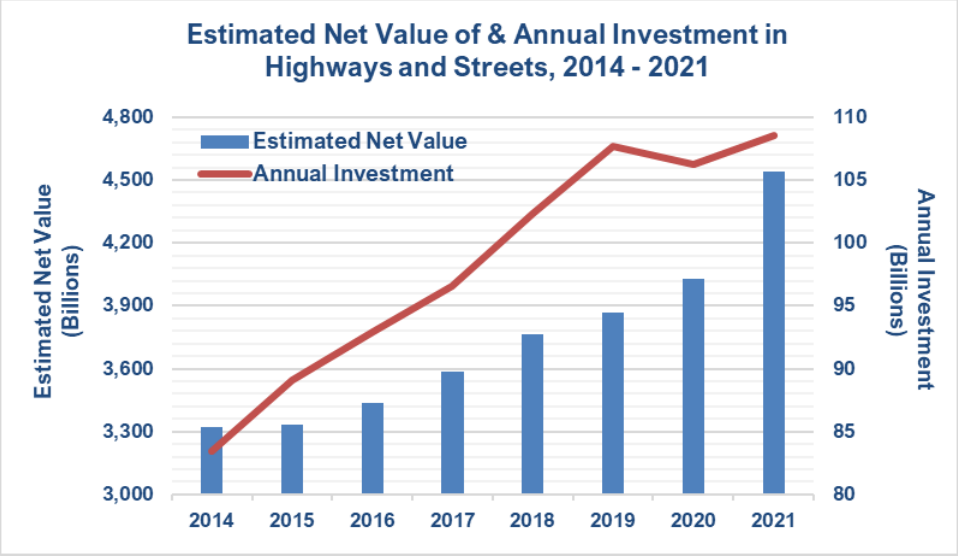


Figure 1.3 Estimated Net Value of Highways and Streets, 2014-2020[3]

Given the substantial impacts and the significant asset value of transportation infrastructure, it is imperative for transportation agencies at all levels to implement a systematic approach for the effective management of infrastructure assets. The objective of infrastructure asset management is to maintain, upgrade, and operate infrastructure systems in a cost-effective manner (Gao and Zhang 2010). By effectively managing infrastructure assets, agencies can enhance system performance, optimize resource

[3]Data Source: BEA, Fixed Asset Tables 2022.

allocation, advance program planning, improve the serviceability of infrastructure, and provide better service for users.

A sizable amount of the expenditures on road infrastructure is dedicated to pavements which is one of the most important components of transportation infrastructure. Pavements are subject to degradation over time due to factors such as traffic loads, environmental impacts, and material degradation. Consequently, periodical maintenance and rehabilitation (M&R) actions are required to maintain the serviceability of pavements. The significance of this investment places a formidable responsibility on local agencies to identify cost-effective strategies and tools for making informed decisions regarding the selection of pavement treatment alternatives.

Pavement management is a well-advanced subfield of infrastructure management. Pavement management includes a wide range of activities such as data collection, condition assessment, performance prediction, maintenance and rehabilitation (M&R) programming, and budget planning and allocation (Uddin et al. 2013). Proper implementation of pavement management activities allows transportation agencies to (Steudle et al. 2012):

- Understand pavement conditions and maintenance needs
- Improve the accuracy of performance predictions
- Compare the outcomes from different M&R strategies
- Predict appropriate budgetary needs
- Be more effective in making use of available resources
- Make more informed decisions on budget management

1.1.3 Data-Driven Methods

In this study, data-driven methods refer to those approaches that emphasize the collection, analysis, and interpretation of data to make informed decisions and solve

problems. Such methods include statistical methods, machine learning, deep learning, artificial intelligence algorithms, and data visualization techniques, to name a few. These methods usually encompass some common activities such as investigating relevant data sources, acquiring available data, preparing data for the proposed study, analyzing the data to extract meaningful insights, and ultimately applying these insights to guide actions and decisions. Due to the effectiveness in uncovering patterns, relationships, and trends in the data, data-driven methods have been commonly used in a variety of fields, including scientific research, healthcare, business, finance, and marketing.

There has been a significant rise in the utilization of data-driven methods within the contemporary realm of transportation engineering. This trend is primarily attributed to the limitations associated with experience-based methods, such as subjectivity and non-reproducibility. In contrast, data-driven methods are based on objective and empirical evidence rather than intuition or experience. By using data to guide decision-making, one can make more accurate, efficient, and effective choices, leading to improved outcomes and productivity.

This present research focuses on two types of data-driven methodologies: geostatistical analyses utilizing geographic information systems (GIS) and cutting-edge algorithms associated with Artificial Intelligence (AI).

GIS-Based Spatial Analysis

Geographic Information Systems (GIS) are computer-based tools that facilitate the storage, management, visualization, analysis, and mapping of data with geographic information, linking it to specific spatial locations on a map. These tools provide a powerful approach to exploring spatial patterns and relationships within a geographic context, making them a popular choice for road safety analysis. In this field, GIS-based methods

have been used to investigate the impact of neighboring locations on motor vehicle crashes (Mohaymany et al. 2013), the impacts of spatially varying parameters on road safety (Quddus 2013), and the identification of high-risk areas for fatal and severe crashes (Satria and Castro 2016, Shahzad 2020).

Crashes are commonly characterized as random events (HSM 2010), however, the probability of crashes as well as the potential severity of these events are not uniformly distributed throughout transportation networks (Xie and Yan 2008). Instead, crash locations typically exhibit specific spatial patterns such as clustering. By utilizing GIS-based spatial analysis, many studies have investigated the spatial patterns of traffic crashes (Hazaymeh et al. 2022, Schneider et al. 2021, Tola et al. 2021, Thakali et al. 2015, Truong and Somenahalli 2011). These studies have applied various geostatistical techniques from GIS, including kernel density estimation (KDE), kriging, K-means, nearest neighbors, Moran's I index, and Getis Ord Gi* statistics, to map and identify crash clusters.

AI-Oriented Algorithms

AI is a multidisciplinary field of study that aims to create intelligent machines capable of autonomous problem-solving and decision-making that mimic human behavior. It encompasses a range of disciplines including computer science, mathematics, information theory, neurobiology, cybernetics, linguistics, psychology, philosophy, and others to achieve its goals (Russel and Norvig 2013).

Over the past decade, many efforts have been made to utilize AI-oriented methods to facilitate decision-making in pavement management. The improvement in automation and sensor technology largely has facilitated the wide application of automated data acquisition in pavement condition surveys (Coenen and Golroo 2017, Ragnoli et al. 2018). Computer image processing techniques have been widely implemented in pavement data

processing, especially in distress identification and classification (Zakeri et al. 2017, Gopalakrishnan 2018, Koch et al. 2015, Mathavan et al. 2015). As one of the most popular tools in engineering, artificial neural networks (ANN) algorithm has been integrated into computational models to solve problems in the general domain of pavement engineering (Adeli 2001, Ishak and Trifiro 2007, Ceylan et al. 2014).

1.2 RESEARCH OBJECTIVES

In numerical analysis, data serves as the foundation, regardless of whether it is based on GIS tools, AI methods, or other statistical techniques. It is the input data that provides a means to gain valuable insights into a problem of interest. While AI-oriented methods have been shown in many studies to be more effective than traditional approaches, the accuracy of the analysis still heavily depends on the quality of the data. Most AI-based techniques, such as supervised and unsupervised learning algorithms, require historical data and variables as input for training the model. Without accurate data inputs, the model cannot learn the relationships between inputs and outputs, resulting in an inability to produce reliable outcomes.

This dissertation endeavors to shed light on the pivotal role that data plays in both roadway safety analysis and pavement management. It accomplishes this objective through four distinct studies, each with its own specific goals:

- The first study focuses on creating a methodological procedure that enables the assessment of data consistency in motor vehicle crash databases. This procedure serves as a valuable tool and a template for ensuring the reliability and accuracy of crash data, laying the foundation for safety analysis.
- The second study develops a methodological framework capable of identifying locations that have the greatest potential for contributing to safety improvements.

By implementing this method, policymakers and practitioners can prioritize their efforts and allocate resources effectively to enhance roadway safety performance.

- The third study conducts a comprehensive synthesis of state-of-the-art advancements in applying AI algorithms across various stages of pavement management. By summarizing key findings, highlighting achievements and limitations, and identifying potential research gaps, this review enables a better understanding of applying advanced AI algorithms to address real-world problems in pavement management.
- Lastly, the fourth study employs a deep learning approach to investigate the impacts of pavement surface conditions on crash prediction. Leveraging a deep neural network, this research explores the intricate relationship between pavement conditions and crash frequency, resulting in a better understanding of the correlation between pavement conditions and roadway safety.

The findings from these studies provide valuable insights into the potential of data-driven approaches to drive advancements in safety analysis and pavement management. By leveraging data to its fullest extent, practitioners and policymakers can make more informed decisions, allocate resources efficiently, and implement cost-effective interventions to improve roadway safety and optimize pavement management practices. In addition, the knowledge gained from these studies also paves the way for future research, advancements, and the adoption of innovative approaches to further elevate the cost-effectiveness and efficiency of safety analysis and pavement management.

1.3 Organization of the Dissertation

This dissertation consists of six chapters. Chapter 1 provides a detailed introduction, covering the background, objectives, and potential contributions of the

research. Chapter 2 outlines a methodological procedure for evaluating data consistency in motor vehicle crash databases. Moving on to Chapter 3, a methodological framework is proposed, which focuses on developing a GIS-based network screening method for evaluating roadway safety performance. This framework incorporates two novel safety performance metrics to enhance accuracy and effectiveness. Chapter 4 provides a comprehensive review that delves into various studies regarding AI applications in pavement management. It includes a summary of achievements and limitations in each reviewed area, along with highlighted perspectives for future research. Chapter 5 introduces a deep learning-oriented method that seeks to explore the influence of pavement conditions on roadway safety. Lastly, Chapter 6 concludes the research by highlighting its contributions and suggesting potential directions for future extensions.

Chapter 2: A Methodological Procedure for Evaluating Data Consistency in Motor Vehicle Crash Databases[4]

2.1 Problem Statement and Objectives

Motor vehicle crashes have been identified as a leading cause of death all over the world. To better promote traffic safety and protect the lives of the traveling public, transportation safety agencies develop and maintain crash databases to effectively manage reportable motor vehicle crashes. Accurate and reliable data on motor vehicle crashes is critical to safety analysis. "Good data about motor vehicle crashes is critical to help explain yearly fluctuations in motor vehicle deaths and injuries and guide policymakers as they consider appropriate investments to reduce those deaths and injuries (NHTSA 2017)."

Law enforcement crash reports serve as a primary source for motor vehicle crash databases. Due to data migrating, interpreting, and transforming, however, the information stored in crash databases may not be identical to the original records on police reports. Examples of such inconsistencies include whether a crash occurred on a horizontal curve, whether it occurred at an intersection, and whether it occurred in a rural area. Although transportation agencies may have taken certain actions to enhance and ensure the consistency and reliability of crash data, it is inevitable that certain biases and errors – at least to some degree – can be introduced to crash databases during data processing. As a case in point, Shipp et al. (2018) developed an online GIS-based curve identification tool using motorcycle-related crashes from Texas' CRIS database. Two types of curve-related

[4]based on two articles: [1] Xu, Y., Han, Z., Zhang, Z. & Murphy, M. (2023). A Methodological Procedure for Evaluating Curve-Related Misclassifications in Motor Vehicle Crash Databases. submitted to *Research in Transportation Economics*. [2] Xu, Y., Han, Z., Murphy, M. and Zhang, Z. 2022. Development of an Automated Methodological Procedure to Improve the Identification of Curve-Related Crashes in the Crash Records Information System (CRIS) (No. FHWA/TX-22/0-7050-1). University of Texas at Austin. Center for Transportation Research. https://rosap.ntl.bts.gov/view/dot/64509

misclassifications were identified in the study. Type A: motorcycle crashes not identified as being on a curve by the GIS tool but identified as such by the CRIS. Type B: The GIS tool identifies crashes as being on a curve, but the CRIS data indicates otherwise. The study reported that 6.5 percent of the analyzed motorcycle crashes from 2010 to 2017 were Type A misclassifications and 22.7 percent were Type B misclassifications.

Aiming at better understanding and evaluating curve-related misclassifications in motor vehicle crash databases, this chapter presents the development of an automated methodological procedure to systematically identify, classify, and quantify such crashes in motor vehicle crash databases. The outcome of the proposed methodology can help improve the overall data consistency of crash databases and enhance the reliability of associated crash analyses.

The method can serve as an effective approach to diagnosing and verifying data quality, forming the foundation of data preparation in various safety analyses. Examples of such analyses include diagnosing crash frequency and severity, identifying locations with the greatest potential for crash reductions through improvements, selecting optimal countermeasures, and conducting cost-benefit analyses.

By implementing this method, transportation agencies can significantly improve the quality of crash data, enabling more effective interventions and investments in reducing motor vehicle-related fatalities and injuries. Ultimately, enhancing data quality through this methodology contributes to the achievement of "Toward Zero Deaths", a national goal on highway safety to reduce serious injuries and deaths in the U.S. transportation system.

2.2 METHODOLOGICAL PROCEDURE

Figure 2.1 illustrates the conceptual framework of the procedure that was developed to improve the identification of curve-related crashes in motor vehicle crash databases,

which is comprised of five major modules: (1) data collection, (2) data review, (3) data examination and cleaning, (4) data integration, and (5) data analysis. Each of these five modules is discussed in detail in the following sections.

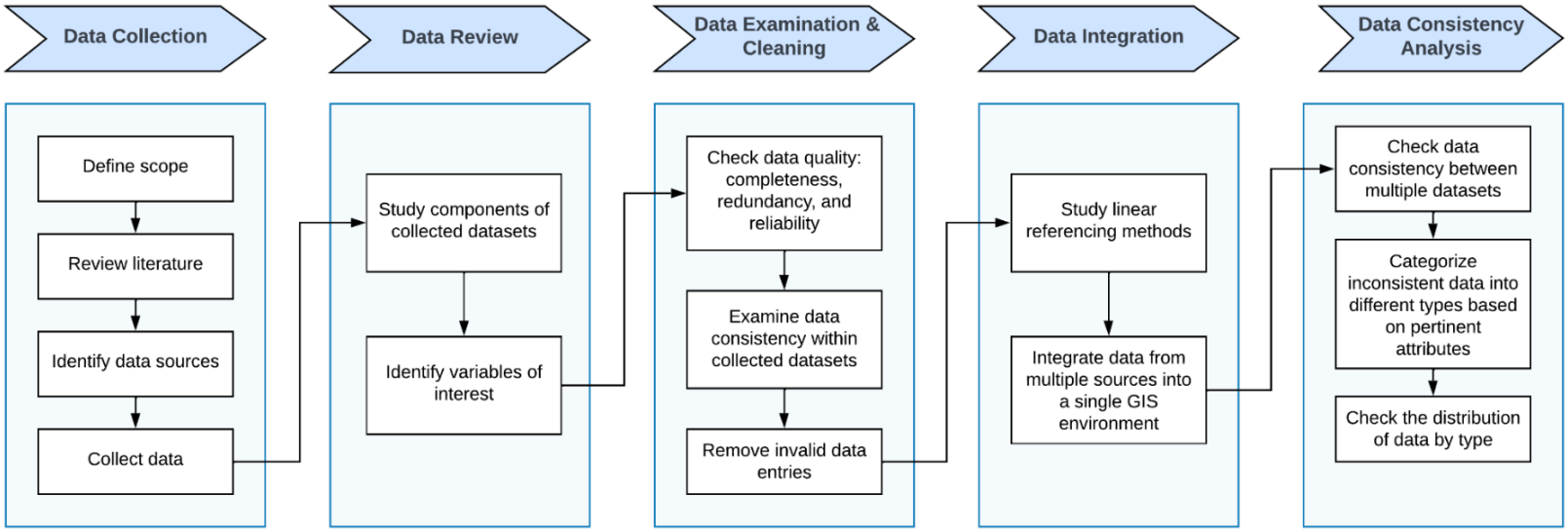


Figure 2.1 Conceptual Framework of the Proposed Methodological Procedure

2.2.1 Data Collection

In this study, the scope of data collection is to acquire crashes, horizontal curves, and roadway inventory datasets. To verify road alignment information in the obtained motor vehicle crash dataset, a dataset of reliable road geometry characteristics with respect to horizontal curves should first be acquired. If such data is not readily available, a horizontal curve dataset will need to be developed. Some efforts have been made to identify horizontal curves (Othman et al. 2012, Findley et al. 2012, Xu and Wei 2016, Bogenreif et al. 2012), and methodologies in these studies can be helpful to build a horizontal curve dataset. Roadway inventory data usually contains a database and a static GIS shapefile which can be used as the master map. Crashes and horizontal curves will be mapped and integrated into the master map using a proper referencing method.

Based on the scope, publicly available data sources can be investigated by either directly searching keywords or reviewing literature that focuses on relevant topics. In general, crash databases usually comprise information in terms of three aspects: crash-related fields, vehicle-related fields, and person-related fields. Crash-related information includes crash identifier, location, date and time, severity, fatality and injury count, road geometrics, weather condition, etc. Vehicle-related data contains basic information about involved vehicles such as type, make, year, model, vehicle damage rating, etc. Person-related information focuses specifically on motor vehicle occupants, for example, age, gender, injury severity, drug test result, belt usage, etc. Considering the scope of this study, crash-related information should be used since it includes all pertinent variables that fulfill the data needs.

Once the optimal data sources have been identified, the analysis period of the study should be determined. It is worth noting that having more years of data does not always

contribute to more accurate analysis results. For example, most safety analyses depend heavily on traffic volume. Due to the fast development of some regions, the traffic volume on the same road can change significantly within a short period. The crash data collected five years ago might not provide meaningful information on safety analysis for future planning. Hence, it is important to collect data within a proper timeframe that can best support the analysis.

2.2.2 Data Review

Before manipulating the data, it is important to understand the components of the obtained datasets. This step includes but is not limited to studying the scope and characteristics of each data element, examining the distribution of each variable, and analyzing the relationship between variables. To gain a better understanding of a dataset, it is necessary to obtain its data dictionary. A data dictionary provides detailed descriptions (e.g., definitions, formats, range of values, etc.) of each data element in a database, which can significantly facilitate the interpretation of information stored in the database. By leveraging data dictionaries, variables of interest can be identified efficiently. In addition, a sufficient understanding of data elements is a prerequisite for data examination and cleaning.

2.2.3 Data Examination and Cleaning

The purpose of data examination is to inspect the quality of the obtained data, including completeness, redundancy, and reliability. Data completeness refers to the wholeness of the data; in other words, it is to check if all required data are available in the dataset. This can be completed by obtaining the percentage of missing data for each data item. Data redundancy occurs when a dataset has multiple copies of the same data entry.

An easy approach to eliminating redundancy is to remove identical data entries from the dataset. For crash data, the unique crash identifier number can be used to check data redundancy. In this study, data reliability means whether the value of a variable is within the predefined range which makes it reasonable and meaningful.

Here, data consistency examination focuses exclusively on whether horizontal curve-related information in crash databases is identical to its original record in the law enforcement report. To achieve this, it is fundamental to obtain data from both motor vehicle crash databases and the original records from law enforcement reports. For example, in TxDOT's CRIS database, the original records from law enforcement crash reports (e.g., fatal crash identifier, crash location, roadway alignment, surface condition, etc.) are saved in the database together with other system-generated crash attributes (e.g., latitude, longitude, street name, on-system flag, rural flag, etc.).

After a comprehensive examination of the obtained datasets, data cleaning is performed to remove invalid data from the datasets. Examples of invalid data include those data with missing, duplicated, or incorrect values (i.e., out of the reasonable range). Such data is not able to contribute to the analysis and may even result in incorrect results.

2.2.4 Data Integration

Data integration is a process of consolidating information from multiple data sources into a single working dataset that can adequately support further use and analysis. In this study, there are three data sources: motor vehicle crash data, horizontal curves, and roadway inventory data. These data can be integrated into a single dataset using linear referencing methods (LRMs). An LRM is an approach to identifying spatial locations based on a known point along linear geographic features, which provides an efficient approach for transportation agencies (e.g., DOTs) to integrate information from multiple data sources

into a comprehensive database (AASHTO 2021). LRMs have been widely applied in transportation research for combining and visualizing various data on the same GIS map (Long et al. 2014, Li et al. 2013). According to a FHWA report (Hausman et al. 2014), LRMs can be grouped into three categories:

- **Absolute methods**: measurements from the origin of the route (or segment) to the event (i.e., crashes in the scope of this study).
- **Relative methods**: measurements from a known reference location to the event of interest.
- **Interpolative methods**: Measurements as a fraction of the entire section distance.

In addition, TxDOT has expanded the concept of LRM by adding route coordinates that express locations along a route using latitude and longitude coordinates as another type of LRM. The most popular LRMs adopted by DOTs are Route Mile-point and Reference Point Offset (Vandervalk et al. 2016). Route Mile-point refers to all linear measurements from the origin of the route, while Reference Point Offset refers to linear measurements from nearby reference markers along the route. LRMs not only make it possible to access multiple data simultaneously but also minimize potential data redundancy in databases (AASHTO 2021).

2.2.5 Data Analysis

Using the integrated dataset, the proposed analysis can be performed to obtain useful information from the data. Activities in data analysis encompass inspecting data consistency between different datasets, categorizing data inconsistencies into different types based on the attributes of interests, and examining the distribution of data by type. Data consistency within each dataset is examined in the previous steps. In this module, data consistency focuses mainly on checking the consistency of data retrieved from different

datasets. In this study, this means identifying inconsistencies by verifying horizontal curve information in the motor vehicle crash dataset with a reliable horizontal curve dataset. Subsequently, crash data with inconsistent curve-related characteristics can be categorized into different types. To gain a better understanding of data inconsistency in the integrated dataset developed previously, the percentage of each inconsistency type is computed.

2.3 CASE STUDY

The case study focused exclusively on motor vehicle crashes that occurred on on-system highway routes (i.e., roadways on the state highway system and maintained by TxDOT) in Texas, U.S.

2.3.1 Data Collection

The public extract CRIS 2017-2020 data (TxDOT 2021a), and the Texas Highway Curves Geographic Information System (GIS) Layer (TxDOT 2021b) were obtained to perform the analysis. CRIS is an automated database that collects and tracks statewide traffic crash records in Texas; it contains all the data received from the Texas Peace Officer's Crash Report (form CR-3). Two types of data files are available through the automated crash data extraction method: the standard extract and the public extract (TxDOT 2021c). The former is available only to certain governmental agencies since it contains sensitive, personally identifiable information. After consulting with TxDOT experts, the most recent public extract CRIS data from 2017 to 2020 was selected for this study, in which the crash-specific data files were used to help identify potential curve-related crash misclassifications in CRIS.

The Texas Highway Curves GIS Layer, published by TxDOT Transportation Planning and Programming Division (TPP), contains information on horizontal curves (i.e.,

curve degree and curve class) and roadway referencing attributes (i.e., route name, beginning and ending distance from origin). This information is presented in a data table and a static GIS shapefile, as illustrated in Figure 2.2. The base map of the GIS shapefile was used as the master map for this study. The Texas Highway Curves GIS Layer provides an integrated dataset of horizontal curves with roadway inventory data, which largely facilitates subsequent data analysis for curve-related crash misclassification in CRIS. Therefore, it was used as a reliable roadway geometry source for identifying curve-related crash misclassification in CRIS.

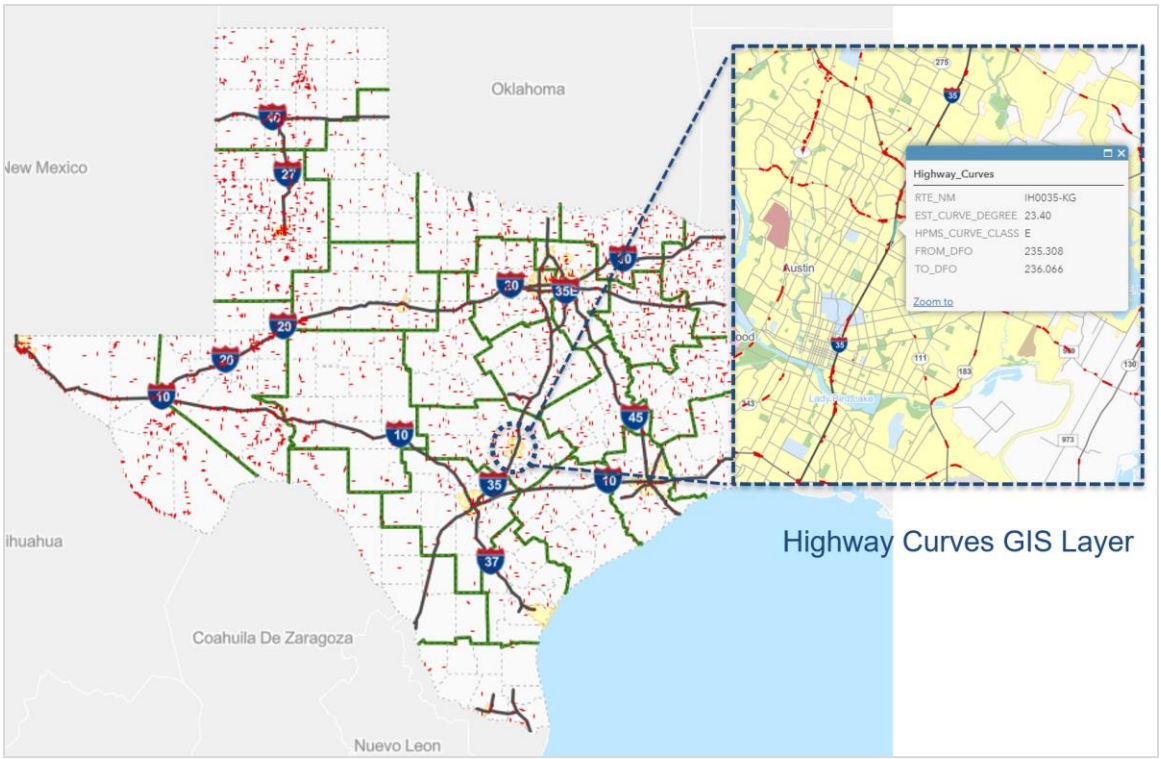


Figure 2.2 An Illustration of Texas Highway Curves GIS Layer

2.3.2 Roadmap for Implementing the Methodological Procedure

After data collection, Python programming language was used to perform data cleaning and generate customized CRIS datasets via Jupyter Notebook (an open-source web application). Then, the customized CRIS datasets were automatically imported into ArcGIS Pro for further visualization and analysis using Python libraries including ArcGIS API for Python and ArcPy. Figure 2.3 presents the framework of the automated methodological procedure, which is comprised of two phases. The first phase focuses on data preparation, and the second phase is data processing.

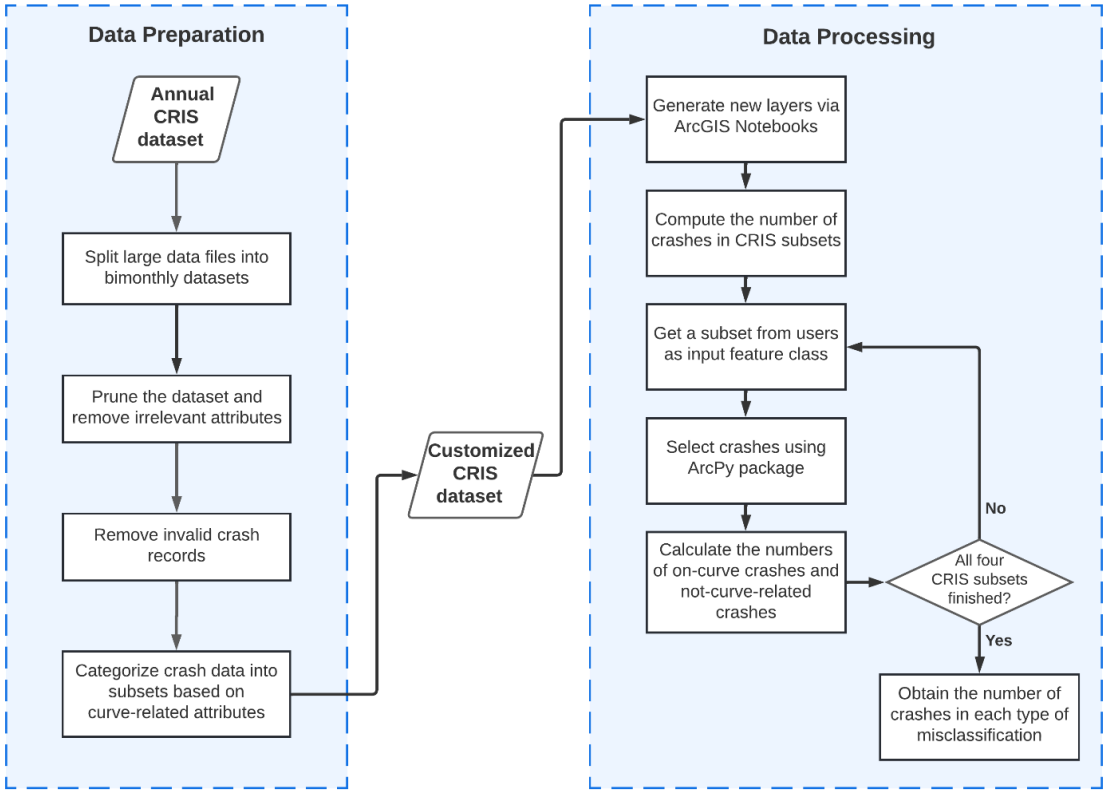


Figure 2.3 A Framework of the Automated Methodological Procedure for Evaluating Curve-Related Crash Misclassifications in CRIS

2.3.3 Data Preparation

To improve the efficiency of data processing, a series of data preparation activities were performed using Jupyter Notebook, which includes reviewing, examining, and cleaning the obtained data. The annual CRIS dataset served as the input, and the outputs were six bimonthly subsets. In total, there were 24 (6 per year * 4 years) bimonthly subsets for the analysis period of 2017 – 2020. The following presents a thorough introduction to data preparation.

Split Large Data Files into a Manageable Size

The data files obtained from the TxDOT CRIS Public Extract are massive because these files contain all crashes that occurred in a specific year across the state. It is challenging and very time-consuming to directly process this magnitude of data. To improve data processing efficiency, the annual crash data is partitioned into several smaller, bimonthly datasets, as shown in Figure 2.4.

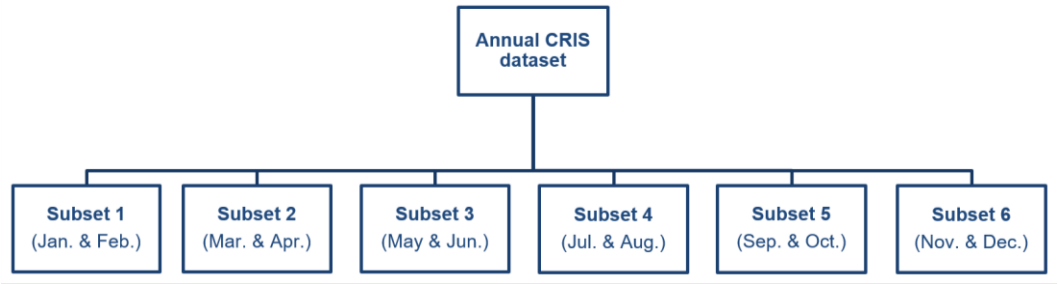


Figure 2.4 Large Annual Crash Data File Split into Six Bimonthly Datasets

Prune the Dataset and Retrieve Useful Attributes

The obtained CRIS dataset contains more than 170 attributes covering roadway identification, geographic information, and crash-related attributes. These attributes can be categorized into the following groups (TxDOT 2021c):

- **CR-3 Reported Data Fields**, e.g., crash ID, fatal crash identifier, school bus crash identifier, railroad crash identifier, crash date, crash time, county name, city name, roadway alignment, and surface condition, among others.
- **Interpreted Fields**, e.g., if bridge related, if intersection related, if object struck, if manner of collision, and if first injury or damage–producing event, among others.
- **System-Generated Fields**, e.g., county ID, city ID, latitude, longitude, highway number, street name, beginning and ending Distance From Origin (DFOs), control section, on-system flag, and rural flag, among others.
- **Appended Roadway Attributes**, e.g., highway design lane ID, median width, base type, number of lanes, width of the right-of-way, roadbed width, surface width and type, curb type, shoulder type and width, curve type, curve length, curve degree, delta left or right identifier, and delta degree, among others.
- **Count Fields**, e.g., suspected serious injury count, non-incapacitating injury count, possible injury count, total injury count, and death count, among others.

After a thorough review of attributes in CRIS, a total of 15 attributes were identified and selected to assist in the identification of curve-related crash misclassification. These attributes included unique identifiers of crashes, locations in the format of different referencing methods, and horizontal curve information. In particular, Road Align ID and Curve Type ID were the indicators representing whether a crash occurred on a curve. Other attributes such as length of curve, curve degree, and curve delta degree were also used to verify the data completeness in terms of curve-related information. The details of these attributes are presented in Table 2.1.

Table 2.1 CRIS Attributes Relevant to Curve-Related Crash Misclassification

No.	Attribute Name	Column Name in CRIS	Description	Field Category
1	Crash ID	Crash_ID	System-generated unique identifying number for a crash	CR-3 Reported
2	Located Flag	Located_Fl	Indicates whether the CRIS locator application was able to locate the crash	System Generated
3	Latitude	Latitude	Latitude map coordinate of the crash	System Generated
4	Longitude	Longitude	Longitude map coordinate of the crash	System Generated
5	Street Name	Street_Name	Name of the road crash occurred on, as determined by the Locator application	System Generated
6	DFO	Dfo	The distance from the origin of the highway to the spot where the crash occurred	System Generated
7	On System Flag	Onsys_Fl	Indicates whether the primary road of the crash was on the TxDOT highway system	System Generated
8	Ref. Marker Nbr	Ref_Mark_Nbr	Reference marker number on the primary highway nearest the crash location	System Generated
9	Ref. Marker Displ.	Ref_Mark_Displ	The distance from the reference marker to the crash location	System Generated
10	Roadway Alignment	Road_Algn_ID	The geometric characteristics of the roadway at the crash site	CR-3 Reported
11	Curve Type ID	Curve_Type_ID	Type of curve, for crashes located on the state highway system	Appended Roadway Attributes
12	Length of Curve	Curve_Lngth	Length of curve, for crashes located on the state highway system	Appended Roadway Attributes
13	Curve degrees	Cd_Degr	Curve degrees (N & S type only), for crashes located on the state highway system	Appended Roadway Attributes
14	Curve delta degrees	Dd_Degr	Curve delta degrees (for crashes located on the state highway system)	Appended Roadway Attributes
15	Delta Left/Right ID	Delta_Left_Right_ID	Identifies whether the curve is right or left (for crashes located on the state highway system)	Appended Roadway Attributes

Remove Invalid Crash Records

In this study, crash records with blank location information were treated as invalid records and removed from the datasets, as these data are not able to be used to identify curve-related crash misclassification. In addition, only on-system crashes were retained for further analysis, and crashes that occurred on off-system roads were removed from the datasets since these data are beyond the scope of this study. Table 2.2 shows the components of CRIS 2017-2020 data and the corresponding percentages of each part.

Table 2.2 Data Components in CRIS 2017 – 2020

Description	Number of Crashes
CRIS 2017 – 2020 total crashes	2,442,781
Number of duplicated crashes	3
Total crashes after removing duplicated crashes	2,442,778
Number of off-system crashes	1,206,090
Total crashes after removing off-system crashes	1,236,688
Number of crashes that are not able to be located	100,137
Total crashes after removing crashes that are not able to be located	1,136,551

Select a Proper LRM for Data Integration

Like many other DOTs, TxDOT takes advantage of well-developed LRMs to manage roadway networks and inventory across the state. LRMs used by TxDOT to locate features along on-system roadways include DFOs, Texas reference markers (TRM), control section mile points (CSM), and route coordinates (latitudes and longitudes) (TxDOT 2018). Three of them are used in CRIS: route coordinates, DFOs, and TRMs. In this study, the route coordinates method was selected to locate CRIS data in ArcGIS Pro because of its accuracy and ability to accommodate complex data.

Categorize Crash Data into Subsets Based on Curve-Related Attributes

In CRIS, the information indicating whether a crash occurred on a curved segment can be derived from two attributes: Curve Type ID and Road Align ID (Xu et al. 2022). Curve Type ID is an attribute generated automatically by the CRIS system to help identify the alignment of the road segment where the crash occurred. According to the data dictionary, the values of Curve Type ID can be 1, 2, 3, or blank. Particularly, 1 represents Normal Curve, 2 represents Point of Intersection (PI) Curve, and 3 represents Spiral Curve. Approximately 80 percent of crash data entries in this field are blank, which was interpreted as the crash did not occur on a curve. Road Align ID, directly derived from CR-3 reports, reflects the judgment of the police investigator on the horizontal alignment of the road segment where a crash occurred. Road Align ID can take numeric values ranging from 1 to 9, and 94. A value of 4, 5, or 6 means that the crash occurred on a curved segment, and other values indicate the crash is not curve-related. Based on guidance from the TxDOT experts, the system-generated attribute – Curve Type ID – is identified as the primary curve indicator of the CRIS dataset in this study.

To improve the efficiency and effectiveness of the methodology, the data in Curve Type ID and Road Align ID fields were regrouped into two categories: 1) on-curve/curve-related or 2) non-curve/not curve-related. For example, if the information retrieved from Curve Type ID was a numeric value (i.e., 1, 2, or 3), then the crash was grouped as curve-related; if the Curve Type ID field was blank, then the crash was grouped as not curve-related. Similarly, when the value of Road Align ID was 4, 5, or 6, the crash was classified as curve-related; otherwise, the crash was classified as not curve-related. The crosstab for the two regrouped attributes is listed in Table 2.3.

Table 2.3 Four Subsets of CRIS Based on Regrouped Curve Type ID and Road Align ID

Curve Indicators in CRIS		**Road Align ID**	
		On curve	***Not on curve***
Curve Type ID	***On curve (Numeric)***	Curve-related crash	Type A conflict
	Not on curve (Blank)	Type B conflict	Non-curve crash

Based on Table 2.3, crashes in CRIS are categorized into four subsets:

- **Curve-related crash**: both Curve Type ID and Road Align ID indicate the crash occurred on a curve segment.
- **Type A data conflict**: Curve Type ID indicates the crash occurred on a horizontal curve, but Road Align ID indicates the crash did not occur on a curve segment.
- **Type B data conflict**: Curve Type ID indicates the crash did not occur on a horizontal curve, but Road Align ID indicates the crash occurred on a curve.
- **Non-curve crash**: both Curve Type ID and Road Align ID indicate the crash did not occur on a curve segment.

It is worth noting that the purpose of this step is to categorize crashes based on two curve-related attributes in CRIS; whether a crash is curve-related or not still needs to be verified with the Highway Curve GIS in subsequent analysis, which is presented in the Data Processing Section. Figure 2.5 shows the decision tree that categorizes CRIS crashes into different subsets.

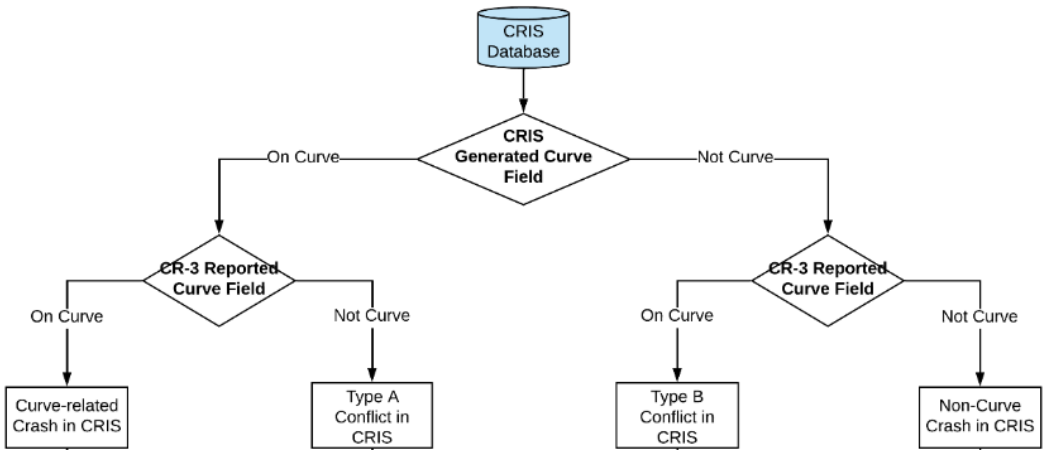


Figure 2.5 Decision Tree for Developing Four CRIS Subsets

After performing all the aforementioned activities, each bimonthly dataset obtained from Figure 2.4 will generate four categorized subsets that formed a customized CRIS dataset, which is used as the input for data analysis in ArcGIS Pro.

2.3.4 Data Processing

Data processing includes integrating and analyzing the customized CRIS dataset obtained from the previous step. The analysis was performed in ArcGIS Notebooks, which are built-in Python notebooks that provide users with a convenient real-time environment to manage (e.g., create, edit, and save) Python codes. By integrating ArcGIS Notebooks, ArcGIS Pro allows users to access GIS map content, conduct real-time data analysis, and obtain instant results that can be visualized in a geographic context (ESRI 2022a). In addition, ArcGIS Notebooks can automatically execute the workflow, which is more efficient by avoiding the significant workload of repetitive operations. Other usages of ArcGIS Notebooks include data cleaning, numerical simulation, statistical modeling, and machine learning, among others (ESRI 2022a).

Classify Curve-Related Crash Misclassifications in CRIS

To identify crash misclassifications in CRIS, the curve-related attributes (i.e., Curve Type ID and Road Align ID) in CRIS were verified using the curve information provided by the Highway Curves GIS layer. If an inconsistency in curve-related information is detected, it will be treated as a curve-related crash misclassification. Accordingly, curve-related crash misclassifications in CRIS can be categorized into the following six types:

- **Type 1 Misclassification**: Both Curve Type ID and Road Align ID indicate that the crash occurred on a curve, but the Highway Curves GIS layer shows the crash was on a non-curve segment.
- **Type 2 Misclassification**: Curve Type ID indicates the crash occurred on a curve, but Road Align ID indicates the crash did not occur on a curve; the Highway Curves GIS layer shows the crash was on a highway curve.
- **Type 3 Misclassification**: Curve Type ID indicates the crash occurred on a curve, but Road Align ID indicates the crash did not occur on a curve; the Highway Curves GIS layer shows the crash was on a non-curve segment.
- **Type 4 Misclassification**: Curve Type ID indicates the crash did not occur on a curve, but Road Align ID indicates the crash occurred on a curve; the Highway Curves GIS layer shows the crash was on a highway curve.
- **Type 5 Misclassification**: Curve Type ID indicates the crash did not occur on a curve, but Road Align ID states the crash occurred on a curve; the Highway Curves GIS layer shows the crash was on a non-curve segment.
- **Type 6 Misclassification**: Both Curve Type ID and Road Align ID indicate the crash did not occur on a curve, but the Highway Curves GIS layer shows the crash was on a highway curve.

As shown in Table 2.4, the identification of curve-related crash misclassification is based on the comparison among information presented by Curve Type ID, Road Align ID, and the Highway Curves GIS layer. Any inconsistency among these data sources would result in a certain type of curve-related crash misclassification in CRIS.

Table 2.4 Six Types of Curve-related Crash Misclassification in CRIS

Type of Misclassifications	**Curve-Related Attributes in CRIS**				**Highway Curves GIS Layer**	
	Curve Type ID		***Road Align ID***			
	On Curve	Not Curve	On Curve	Not Curve	On Curve	Not Curve
Type 1	✓		✓			✓
Type 2	✓			✓	✓	
Type 3	✓			✓		✓
Type 4		✓	✓		✓	
Type 5		✓	✓			✓
Type 6		✓		✓	✓	

To illustrate the procedure for evaluating curve-related crash misclassification in the case study, a decision tree was developed, as shown in Figure 2.6. The eight boxes at Level 3 were highlighted with three colors:

- **Two green boxes** represent crash data with consistent curve-related information between the Highway Curves GIS layer and the CRIS dataset.
- **Four orange boxes** represent four types (i.e., Type 1, Type 3, Type 4, and Type 6) of curve-related crash misclassifications. These four types of misclassifications are highlighted using the same color because they all reflect a conflict between the Highway Curves GIS layer and the Curve Type ID attribute, which is the primary curve indicator of the CRIS dataset.

- Type 1 and Type 6 are caused by inconsistency between the Highway Curves GIS layer and the CRIS dataset, but Curve Type ID and Road Align ID are consistent.
- Type 3 and Type 4 are caused by inconsistency between the Highway Curves GIS layer and the Curve Type ID attribute from the CRIS dataset.

- **Two pink boxes** present two types (i.e., Type 2 and Type 5) of misclassifications caused by inconsistency between the Highway Curves GIS layer and the Road Align ID attribute from the CRIS dataset.

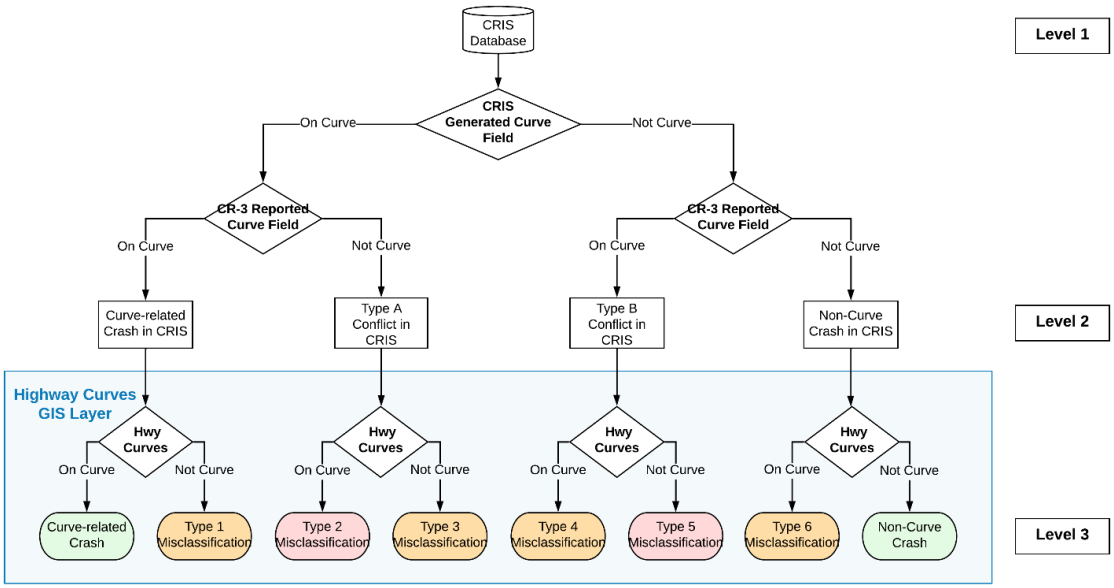


Figure 2.6 Decision Tree for Classifying Six Types of Curve-Related Crash Misclassification in CRIS

Automate the Proposed Procedure

Leveraging Python programming language and ArcGIS Python libraries, a comprehensive data processing and analysis was conducted to automatically locate the customized CRIS data in ArcGIS Pro and evaluate misclassifications using the Highway

Curves GIS layer as a reference. The step-by-step workflow of the automated procedure has been illustrated in Figure 2.3 and is summarized in detail as follows:

- **Step 1** – A Python program was developed to automatically generate a new feature layer for each of the four CRIS subsets presented in Figure 2.5, which was completed via ArcGIS Notebooks using the Highway Curves GIS map as a base map.
- **Step 2** – The number of crashes that belonged to each of the four subsets was automatically computed using a developed Python program.
- **Step 3** – Python program was applied to obtain one CRIS subset as the input feature class. In this step, users are expected to interact with the program by selecting one subset from the four CRIS subsets.
- **Step 4** – Crashes belonging to the subset selected in Step 3 were identified using the Layers and Table Views toolset provided by the *ArcPy* package. Parameters defined in this step include input features, selecting features, relationship between selected features, search distance (set as 1 foot in this study), selection type, etc.
- **Step 5** – The number of on-curve and non-curve crashes were calculated, respectively.
- **Step 6** – Finally, the number of each type of six misclassifications and two correct classifications – based on Table 2.4 and Figure 2.6 – were obtained.

2.3.5 Results and Discussion

By implementing the automated methodological procedure, systematic data analysis was performed to identify the patterns and characteristics of curve-related crash misclassification in the CRIS database. This section presents the key findings from the analysis.

Data Consistency in CRIS

To explore potential data conflicts in curve-related crash classification in CRIS, a comprehensive investigation was conducted to check data consistency in the CRIS database. As illustrated in Figure 2.5, by comparing the values in Curve Type ID with the information in Road Align ID using four-year CRIS data (2017-2020), it was found that approximately 77.3 percent of crash records have consistent curve-related information, whereas 22.7 percent of crash records contain internally inconsistent curve attributes. As illustrated in Figure 2.7, both Curve Type ID and Road Align ID agree that about 74.1 percent of crashes did not occur on a curve segment while around 3.2 percent of crashes occurred on a curve segment. However, for 17.2 percent of crashes, the Curve Type ID attribute indicates that they occurred on a curve segment, but the Road Align ID indicates they are not curve-related (i.e., Type A data conflict). In contrast, about 5.5 percent of crashes did not occur on a curved segment based on the Curve Type ID, but the Road Align ID shows these crashes occurred on curved sections (i.e., Type B data conflict).

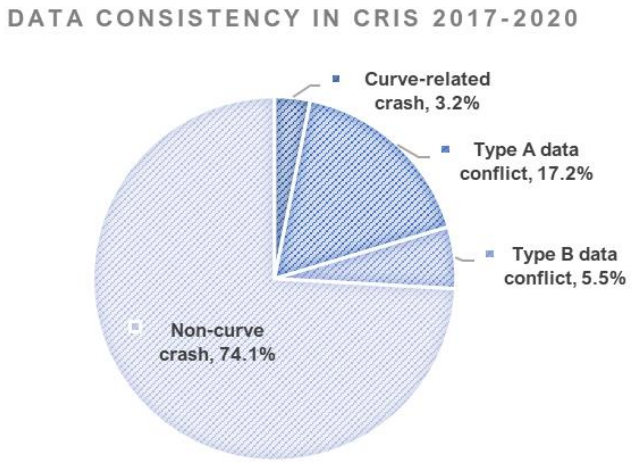


Figure 2.7 Data Consistency between Curve Type ID and Road Align ID in CRIS 2017-2020 (Total crashes: 1,136,551)

Curve-Related Crash Misclassification in CRIS

The study further examined the accuracy of curve information in CRIS using the Highway Curves GIS layer as a reference. Based on the decision tree chart illustrated in Figure 2.6, the percentage of crashes that fell into each of the six misclassification categories was calculated and summarized in Table 2.5 and Figure 2.8, respectively.

Table 2.5 Percentage of Crashes by Year and Type

Crash Type	2017	2018	2019	2020	4-yr average
Total on-system crashes	291,555	292,998	292,577	259,421	284,138
Curve-related (correctly classified)	0.4%	0.4%	0.4%	0.5%	0.4%
Non-curve (correctly classified)	72.3%	72.6%	72.9%	72.2%	72.5%
Type 1 Misclassification	2.8%	2.7%	2.6%	2.9%	2.8%
Type 2 Misclassification	1.5%	1.4%	1.4%	1.5%	1.5%
Type 3 Misclassification	16.0%	15.8%	15.8%	15.6%	15.8%
Type 4 Misclassification	0.4%	0.4%	0.3%	0.4%	0.4%
Type 5 Misclassification	5.0%	5.1%	4.9%	5.4%	5.1%
Type 6 Misclassification	1.6%	1.6%	1.6%	1.6%	1.6%

Note: Percentages may not total 100% due to rounding.

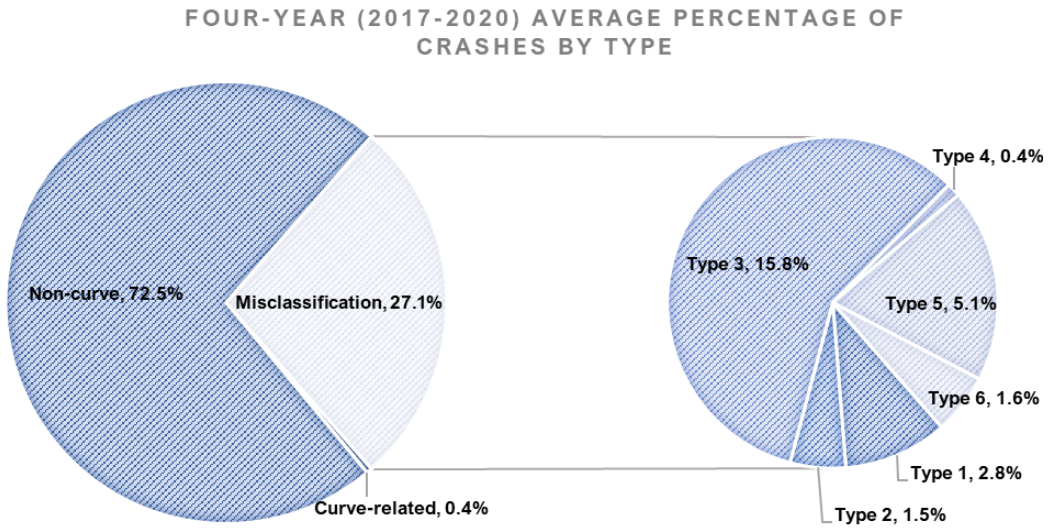


Figure 2.8 Average Percentage of Crashes by Curve Relationship and Misclassifications Type from 2017 to 2020

As demonstrated in Figure 2.8, approximately 72.9 percent of crashes in the CRIS dataset had consistent curve-related information with the Highway Curves GIS layer: about 72.5 percent of crashes are non-curve while around 0.4 percent occur on a curved segment. In contrast, about 27.1 percent of crash records are inconsistent with the information in the Highway Curves GIS layer, which is classified into six curve-related crash misclassification types as illustrated in Table 2.4 and Figure 2.6. Of these, Type 3 Misclassification is the most common type, accounting for about 15.8 percent of all crash records. The next is Type 5, which includes approximately 5.1 percent of all crashes, followed by Type 1 (about 2.8 percent), Type 6 (about 1.6 percent), and Type 2 (about 1.5 percent). Type 4 Misclassification is the most infrequent, accounting for only about 0.4 percent of crashes.

Findings from Texas Peace Officer's Crash Report (CR-3)

To better understand potential reasons behind these curve-related crash misclassifications in CRIS, the authors conducted a preliminary investigation using CR-3 reports which are legal documents used to collect motor vehicle crash data by law enforcement agencies in Texas. These reports include field drawings and narrative notes documented by the police officers who conducted the on-site investigation. This information is expected to provide meaningful insights into potential reasons why the crash was misclassified.

In Texas, the state transportation agency – TxDOT – collects and maintains data for motor vehicle crashes that occurred on public roadways and the state highway system for the current year and the previous ten years (TxDOT 2022). CR-3 crash reports are the sole data source for these crashes. When the damage caused by a crash exceeds $1,000, or when a crash results in injury or death, the CR-3 crash report is obligatory to be submitted to TxDOT (TxDOT 2022). The TxDOT Traffic Safety Division (TRF) division creates new data entry in CRIS by overlaying a GIS layer of crashes that are directly retrieved from CR-3 reports with a spatial GIS layer (provided by the TPP division) that contains roadway inventory information including horizontal curves. During this process, crashes are projected to the location where they occurred.

In the preliminary investigation, the authors reviewed 60 randomly selected CR-3 reports (i.e., 10 crashes for each type of misclassification) for detailed crash analysis. It was extremely challenging to pinpoint the exact cause(s) of these misclassifications based solely on this preliminary investigation; however, rational inferences on probable reasons were made using the available information. By inference, Type 1 and Type 6 misclassifications are most likely caused by inaccurate GPS coordinates that fail to precisely reflect the actual location of the crash. Type 2 and Type 5 misclassifications are

most likely caused by either incorrect curve classification derived from CR-3 reported data fields or inaccurate GPS coordinates. Type 3 and Type 4 misclassifications are primarily caused by incorrect curve information generated by the CRIS system or inaccurate GPS coordinates.

These findings imply that inaccurate GPS coordinates could be the primary contributor to curve-related crash misclassifications in CRIS. The case study used GPS coordinates from the CRIS database to locate crashes on the Highway Curves GIS layer and then verified the alignment of the road segment at that location; if the CRIS coordinates of a crash are unable to reflect the actual location where it occurred, the outcome of the verification would be invalid. Therefore, to better uncover the causes of curve-related crash misclassifications, more efforts are needed to ensure the accuracy of GPS/route coordinates in CRIS.

2.4 SUMMARY

Motor vehicle crash data is usually collected from crash reports provided by law enforcement authorities. However, the migration and rearrangement of data could introduce problems of data consistency in crash databases. One such data inconsistency that has been identified is regarding road alignment (i.e., horizontal curves), which would significantly impede any kind of safety analysis using such data. To solve this issue, an automated methodological procedure for evaluating (i.e., identifying, classifying, and quantifying) curve-related crash misclassifications in motor vehicle crash databases was developed in this part of the dissertation research.

A case study was presented to demonstrate the applicability and effectiveness of the proposed methodological procedure. A total of 1,136,551 on-system crashes in TxDOT CRIS were analyzed and categorized into six curve-related misclassification types. The

percentage for each type of misclassification was automatically computed. Results indicated that about 27.1 percent of crash data has curve-related misclassification issues, with Type 3 as the most common type (15.8 percent), followed by Type 5 (about 5.1 percent), Type 1 (about 2.8 percent), Type 6 (about 1.6 percent), Type 2 (about 1.5 percent), and Type 4 (about 0.4 percent). The case study proved that the automated methodological procedure could help evaluate curve-related crashes both effectively and efficiently.

In summary, the main contribution of this study is the development of a method to effectively improve the identification of curve misclassifications in crash databases, thereby enhancing the accuracy and reliability of safety analysis based on the data (e.g., crash prediction models). The proposed method is flexible and can be customized by transportation agencies depending on their data availability, attributes of interest, preferred data examination and integration techniques, as well as specific needs. Transportation agencies can utilize this procedure to improve the quality of crash data, leading to more effective interventions and investments in reducing motor vehicle-related fatalities and injuries.

Chapter 3: A Methodological Framework for Identifying Crash Hot Spots and Prioritizing Unsafe Horizontal Curves in Roadway Network: Case Studies of Interstate Highways in Texas

3.1 Problem Statement and Objectives

Providing safety and mobility for the traveling public is the first priority of our transportation infrastructure systems. To tremendously reduce roadway fatalities and serious injuries, transportation agencies at all levels of government have initiated various proactive programs such as Vision Zero, Road to Zero, and Toward Zero Deaths. In practice, however, it will be a long-term endeavor to achieve the goal of "zero fatalities" given that the total length of the U.S. roadway network is more than four million miles. To be more practical, the major focus – at least for the current stage – is to mitigate traffic crashes that result in deaths and serious injuries. To achieve this goal, many proactive actions have been taken to determine areas that can contribute to the most significant improvement of roadway safety performance. One such area is to identify crash hot spots (a.k.a., black spots) which are defined as locations with relatively higher crash rates or with more fatal and serious crashes. Identifying crash hot spots is important because it can help better understand the causes of motor vehicle crashes and significantly support the development of effective countermeasures in the most appropriate locations, thus contributing to a tremendous reduction in crash frequency or severity.

During the past decade, many studies have illustrated the effectiveness of using global Moran's I and Getis Ord Gi* methods to evaluate the spatiotemporal distribution of traffic crashes (Hazaymeh et al. 2022, Soltani and Askari 2017, Truong and Somenahalli 2011, Le et al. 2022, Tola et al. 2021). These crash hot spot studies can significantly help identify sites with high crash risk in a network, thus providing meaningful insights into ranking candidate locations that are most likely to contribute to a safety improvement.

However, very few efforts have been made to assess the safety performance of those locations with special geometric characteristics (i.e., horizontal curves) which are known to be extremely susceptible to fatal and severe crashes. Statistics show that the change in road alignment can introduce uncertainty that results in a dramatic increase in the potential of crash occurrence and severity (Hummer et al. 2010, Torbic et al. 2004, Glennon et al. 1983, Donnell et al. 2019). The Federal Highway Administration (FHWA) reports that nearly one-fourth of fatal crashes are associated with horizontal curves which account for merely 5 percent of the road network across the U.S. (FHWA 2021, FHWA 2016). Also, it is estimated that over half of traffic fatalities are caused by roadway departure crashes which are more likely to occur on horizontal curves (Donnell et al. 2019). There are more than 10 million horizontal curves along two-lane U.S. highways (Xin et al. 2017, Torbic et al. 2004). The significant impacts of horizontal curves on roadway safety have been thoroughly assessed by previous studies (Albin et al. 2016, Xin et al. 2017, Donnell et al. 2019, Findley et al. 2012, Khan et al. 2013, Schneider IV et al. 2010). Given the large number and crucial impacts of horizontal curves, the prioritization of hazardous horizontal curves can make a significant contribution to the reduction of fatal and serious crashes, which can dramatically facilitate the achievement of the "zero fatalities" goal.

To fill this gap, this part of the research aimed to develop a GIS-based network screening method for evaluating roadway safety performance and identifying crash hot spots where improvements have the most potential to yield a reduction in crash frequency or severity. Further, the study proposed a composite safety performance measure that can quantify the safety performance of horizontal curves across the network. Using this measure, horizontal curves that pose the most safety risk in the network can be effectively prioritized. The hazardous locations identified by the proposed method can serve as candidates for further detailed investigation. It is anticipated that this method will be useful

to local transportation agencies for screening roadway networks and identifying locations with the most safety need.

3.2 METHODOLOGY

To accomplish the goal of the study, various methods have been employed and integrated to develop the proposed method for crash hotspot identification and horizontal curve prioritization. This section provides general background information for these supportive methods. Subsequently, the conceptual framework for implementing the proposed study is introduced. Most importantly, two safety performance measures are proposed to assess the safety performance of road segments and horizontal curves.

3.2.1 Methodological Background

This section presents a brief overview of fundamental concepts and methods that are used to support the development of the methodological framework proposed in this study.

Crash Severity Classification

The most commonly used method for classifying motor vehicle crash injuries is the KABCO crash injury scale which was introduced by the National Safety Council (NSC). The scale and definitions of KABCO may vary slightly from one state DOT to another, but most DOTs employ the KABCO system for crash severity classification. According to NHTSA, the scale and definitions of injury severity level for the KABCO system with unit cost are presented in Table 3.1 (NHTSA 2017, Harmon et al. 2018).

Table 3.1 KABCO Crash Severity Scale and Unit Cost

KABCO	Crash Severity Level	Description	Crash Unit Cost
K	Fatal Crash	Any motor vehicle crash that causes fatality within 30 days after the crash occurred.	$4,288,422
A	Suspected Serious Injury Crash	For example, unconsciousness, significant loss of blood, broken arms or legs, crush injuries, significant burns, etc.	$781,094
B	Suspected Minor Injury Crash	Examples are minor bruises, lumps, abrasions, shallow cuts, etc.	$174,335
C	Possible Injury Crash	Reported injuries but without apparent wounds such as claims of injury, limping, complaints of pain, etc.	$98,188
O	Property Damage Only (PDO)	Neither physical evidence of injury to any person involved in the crash nor reported injuries.	$10,582

It is worth noting that a crash may result in multiple people suffering from different levels of injuries. The severity of a motor vehicle crash is determined by the most severe injury to all persons involved. Moreover, various methods have been developed to quantify the economic loss caused by motor vehicle crashes. The unit cost of a specific type of crash may vary depending on the specific method used to calculate the number. As a pragmatic approach, this study selects the average crash unit cost obtained from 33 states to perform the analysis (Harmon et al. 2018).

Roadway Safety Performance Measures

Performance measures are usually numerical indicators that are used to quantify the risk of crashes (e.g., number of crashes and crash severity level) at a particular location. Various measures have been developed to perform this task. Some commonly used safety performance measures are crash frequency, crash rate, and equivalent property damage only (EPDO). Table 3.2 lists the definition and equation used to calculate these measures.

Table 3.2 Examples of Commonly Used Safety Performance Measures

Performance Measures	**Definition**	**Equation**
Crash Frequency	The number of crashes occurring within a specific roadway segment over the study period (FHWA 2011).	N N: Total number of crashes over the analysis period.
Crash Averaging	The average of crashes over the analysis period (FHWA 2011).	$\frac{N}{t}$ t: Number of years analyzed
Crash Rate	Crash rate normalizes the frequency of crashes with the exposure, measured by traffic volume (HSM 2010).	$R = \frac{N}{\left(\frac{365 * t * AADT * L}{100{,}000{,}000}\right)}$ R: Crash rate for the road segment expressed as crashes per 100 million vehicle-miles of travel (VMT) $AADT$: Annual Average Daily Traffic L: Length of the roadway segment in miles
Equivalent Property Damage Only (EPDO)	An EPDO is a weighted score that evaluates the frequency and severity of crashes by assigning weighting factors related to societal costs by crash severity (HSM 2010).	$\sum_{i=1}^{3} \beta_i * N_i$ $\begin{cases} i = 1, fatal \\ i = 2, injury \\ i = 3, PDO \end{cases}$ β_i: EPDO weighting factor for level i crashes N_i: number of crashes in level i
EPDO Rate	A normalization of EPDO with exposure (e.g., VMT)	$\frac{EPDO}{exposure}$

Note: All the definitions and equations only apply to roadway segments. Equations for intersections could be different.

Among the measures listed above, crash frequency is the most simple and intuitive method. It is easy to implement, however, roadway segments may not always have the identical length in practice. Hence, this method may not yield accurate results when the length of segments varies significantly from one to another. More importantly, the number of crashes that occurred within a roadway segment is usually assumed to be positively related to traffic volumes. When other variables are identical, a segment with more traffic would expect more crashes compared to one with less traffic volume. Due to this reason, crash rate is more accurate than crash frequency. As discussed previously, crash severity

also plays a vital role in safety analysis, for example, the safety impacts of fatal and serious injuries could be much heavier than PDO injuries. EPDO makes it possible to take crash severity into consideration. Similar to crash rate, EPDO rate can be computed by dividing VMT from EPDO, which normalizes the frequency of EPDO with exposure.

Spatial Autocorrelation (Moran's I)

Spatial autocorrelation analysis is a popular method for assessing the spatial structure of quantifiable attributes associated with a set of spatial units (Songchitruksa and Zeng 2010). As discussed in the previous chapter, it has been widely used in many studies of the spatial distribution of traffic crashes. Spatial autocorrelation can be determined by conducting statistical tests. In this study, the global Moran's I index is employed to measure spatial autocorrelation based on locations and values of features simultaneously. The index is a value ranging from -1 to 1. A positive value usually indicates a cluster of traffic crashes; a negative score shows a dispersion of crashes; and an approximate zero represents a randomly distributed pattern of crashes. Moran's I index is calculated from the following mathematical Equation (1) (Soltani and Askari 2017, Truong and Somenahalli 2011).

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{1}$$

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \tag{2}$$

where:

n: total number of spatial units (i.e., road segments).

w_{ij}: spatial weight of location i and j with $w_{ii} = 0$.

x_i: CSI value at location i.

$\bar{x}$: global mean of CSI values.

S_0: aggregate of all spatial weights, computed from Equation (2).

Based on the value calculated above, the z-score for Moran's I statistic can be computed to evaluate the significance of the Moran's index using Equation (3) listed below:

$$Z[I] = \frac{I - E[I]}{\sqrt{Var[I]}} \tag{3}$$

where:

$$E[I] = -\frac{1}{n-1}$$

$$\sqrt{Var[I]} = E[I^2] - E[I]^2$$

Getis-Ord Gi* Statistic

In addition to the global measure, a local measure – Getis-Ord Gi* – is used to evaluate the spatial autocorrelation associated with a particular spatial unit (Getis 2010). At the local level, a statistically significant hot spot is a feature that has a high value for the attribute of interest, and its nearby features also with high values (ESRI 2022b). ArcGIS Pro provides a hot spot analysis tool that can be used to compute the Getis-Ord Gi* statistic. To identify potential hot spots, the tool first compares the local sum of a feature and its nearby neighbors to that of all features in the dataset. For each feature in the dataset, the ArcGIS tool will generate a Gi* statistic (i.e., z-score). If the local sum is significantly distinct from the expected value, the tool will generate a statistically significant z-score; the bigger the z-score, the more intense high values clustered (ESRI 2022b). The output of

the Gi* statistic is a hot spot map that illustrates the locations of spatial clusters (Soltani and Askari 2017). The Getis-Ord Gi* statistic is computed from Equations (7-11) (Getis 2010, Getis and Ord 1992, Hazaymeh et al. 2022):

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} x_j} \tag{4}$$

$$Z(G_i^*) = \frac{\sum_{j=1}^{n} w_{ij} x_j - \bar{X} \sum_{j=1}^{n} w_{ij}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij} \right)^2 \right]}{n-1}}} \tag{5}$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{6}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \bar{X}^2} \tag{7}$$

where:

x_j: attribute value for feature j.

w_{ij}: spatial weight between feature i and j.

3.2.2 Methodological Framework

This study utilized spatial autocorrelation analysis to gain a better understanding of the spatial distribution of motor vehicle crashes that occurred on interstate highways in Texas. To quantify safety performance, each route in the network was broken down into small roadway segments with known length and traffic volume (i.e., AADT). Hence, the spatial unit used in the analysis is road segments. Instead of applying a regular screen window method (HSM 2010), the study took full advantage of the ready-to-use

segmentation developed for pavement condition survey data. Traffic crashes were aggregated for each roadway segment based on their crash severity level. Then, the proposed CSI values were computed using Equation (2); the coefficients associated with the equation were obtained using the EPDO weighting factors. Based on the CSIs, the spatial patterns of crash data were examined using Moran's I index, and a crash hotspot map was developed using the Getis-Ord Gi* method. In addition to the hotspot map, the safety performance of horizontal curves was evaluated using the proposed measure. This measure provides an effective approach to prioritize unsafe horizontal curves across the network. The proposed methodological framework can be employed as a preliminary network screening approach for ranking and determining candidate segments that have a higher number of crashes compared to other sites. By implementing the method, locations with a higher potential for safety issues can be identified effectively. With the identified candidate sites, local agencies can conduct field tests and analyses to determine future needs, prioritize projects, and allocate funding.

Figure 3.1 illustrates the methodological framework of the proposed study, and the major activities conducted in this study are listed below:

- Collect data sources to perform the analysis
- Determine the scope of spatial analysis and the reference population in the roadway network
- Integrate multiple data sources into a single GIS environment using a proper linear referencing method (LRM)
- Divide roadways into small segments
- Aggregate crash data for each road segment based on the crash severity level
- Compute the CSI for each road segment

- Perform Data Scaling for the CSIs computed in the previous step to make all CSIs range from 0-100
- Examine the Spatial Patterns of crashes based on CSI using Moran's I statistics
- Construct a Crash Hotspot Map based on Hot Spot Analysis (Getis-Ord Gi*)
- Identify road segments associated with horizontal curves
- Compute SI for all curves along IHs
- Prioritize unsafe horizontal curves based on the SI ranking

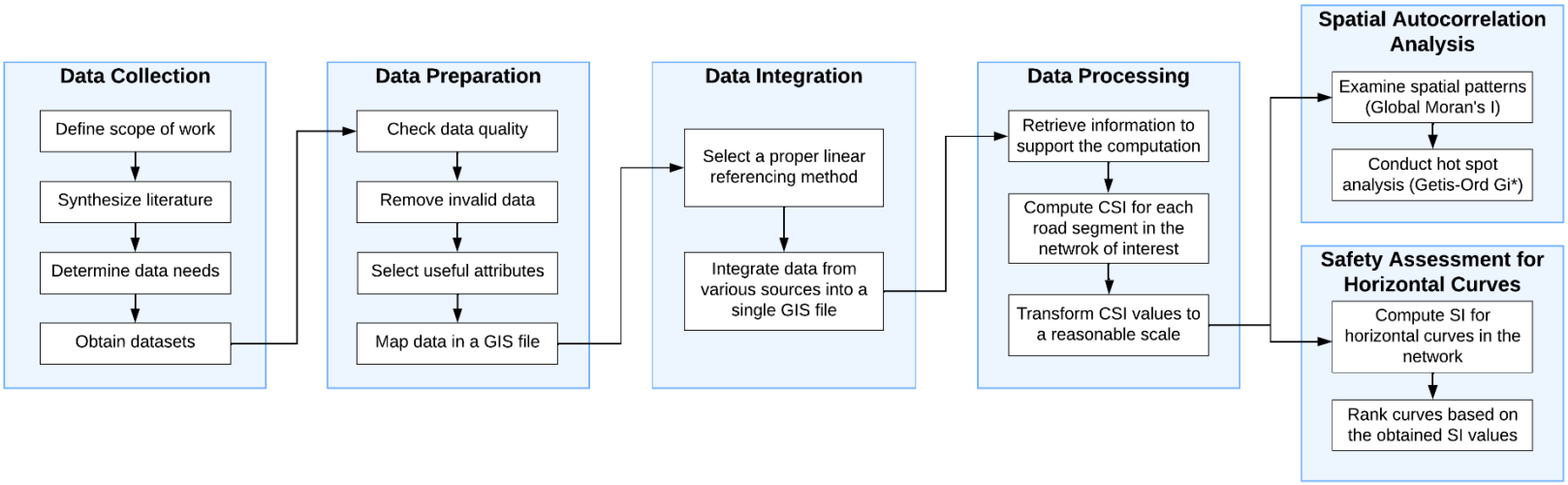


Figure 3.1 Methodological Framework for the Proposed Study

3.2.3 Proposed Safety Performance Metrics

To efficiently assess roadway safety performance and identify crash hot spots for further improvements, this study proposed two composite safety performance measures that can quantify the safety performance of road segments and horizontal curves, respectively. Using these measures, sites that pose the most safety risk in the network can be effectively prioritized.

Crash Severity Index (CSI)

For the standard EPDO method (HSM 2010), all three types of injuries (i.e., A, B, and C) are grouped into one category. It is worth noting that not all state departments of transportation (DOTs) use the same crash severity levels as recommended by KABCO. For example, some DOTs (e.g., Texas, Pennsylvania, and Wyoming) use an additional "Unknown (U)" category to group those crashes that have no information to determine whether injuries exist. Similarly, when calculating EPDO weighting factors, DOTs use different schemes to assign EPDO weights for KABCO severity levels. For example, Arizona only considers the economic loss caused by K and A crashes while ignoring other types; Illinois counts K, A, and B crashes; and Texas weighs K and A crashes with the same cost higher than that of B crashes (Harmon et al. 2018).

For this reason, this study proposed a numeric indicator – noted as CSI – which considers all scales of crash severities. As shown in Equation (2), CSI is constructed as a function of crash rate, severity, and exposure. The upper part is a full-scale version of the EPDO method which is used to quantify the economic impacts of crashes. As discussed earlier, exposure is unevenly distributed through the network so high EPDO scores could just be a result of a large exposure; assuming other conditions are identical, a road segment

with more crashes will be more likely to have more traffic. To address this issue, the lower part of Equation (8) represents exposure in the form of 100 million Vehicle Miles Traveled (VMT) which is computed from Equation (9). It is expected that locations with higher CSI values typically cause more economic losses due to their higher crash frequencies or susceptibility to more severe crashes.

$$CSI = \frac{\beta_K N_K + \beta_A N_A + \beta_B N_B + \beta_C N_C + \beta_O N_O}{VMT} \tag{8}$$

$$VMT = \frac{AADT_{mean} \times L_{seg} \times 365 \times n_{yr}}{100{,}000{,}000} \tag{9}$$

where:

$AADT_{mean}$: average AADT in the n year analysis period.

L_{seg}: length of roadway segment in miles.

n_{yr}: analysis period in years.

VMT: 100 million vehicle miles of travel which is used as a measure of exposure.

β_K: coefficient of fatal crashes.

N_K: number of fatal crashes in the n-year analysis period.

β_A: coefficient of incapacitating injury crashes.

N_A: number of incapacitating injury crashes in the n-year analysis period.

β_B: coefficient of non-incapacitating injury crashes.

N_B: number of non-incapacitating injury crashes in the n-year analysis period.

β_C: coefficient of possible injury crashes.

N_C: number of possible injury crashes in the n-year analysis period.

β_O: coefficient of no injury crashes.

N_O: number of no-injury crashes in the n year analysis period.

Due to the wide range of VMT distribution, CSI values computed directly from Equation (9) may also have a few very large numbers (i.e., outliers). Outliers can significantly impede the accuracy of subsequent analyses. To handle this issue, the upper fence derived from a box-and-whisker plot (Figure 3.2) is used to regroup CSI outliers to cut down the heavy tail, as shown in Equation (10).

$$Upper\ Fence = Q_3 + 1.5 * IQR \qquad (10)$$

where:

Q_1: lower quartile, 25% of observations are smaller than this value when they are arranged in ascending order.

Q_3: upper quartile, 75% of observations are smaller than this value when they are arranged in ascending order.

IQR: interquartile range ($IQR = Q_3 - Q_1$), the difference between the upper and lower quartiles.

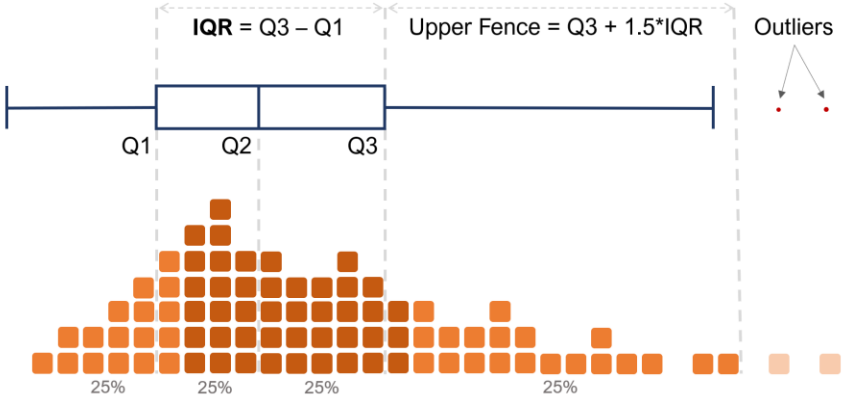


Figure 3.2 An Example of a Box-and-Whisker Plot[5]

[5]Adapted from *A Complete Guide to Box Plots* by Chartio: https://chartio.com/learn/charts/box-plot-complete-guide/#:~:text=The%20third%20quartile%20(Q3)%20is,locations%20of%20these%20three%20quartiles

After regrouping CSIs, the min-max scaling method is applied to transform the data to a scale ranging from 0 to 100, as illustrated in Equation (11). This step is intended to help a better interpretation of the results.

$$x' = \left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) * 100 \tag{11}$$

Safety Index (SI) for Horizontal Curves

To effectively quantify the safety performance of horizontal curves in a roadway network, this study proposed a novel indicator (i.e., SI) – exclusively for horizontal curves – which can be calculated using Equation (12) as below:

$$Safety\ Index\ (SI) = w_{overlap} * \frac{\sum_{i=1}^{n} CSI_i}{n} + w_{nearby_{frt}} * \frac{\sum_{i=1}^{p} CSI_i}{p} + w_{nearby_{bck}} * \frac{\sum_{i=1}^{q} CSI_i}{q} \tag{12}$$

where:

$w_{overlap} + w_{nearby_{frt}} + w_{nearby_{bck}} = 1$

$w_{overlap} > w_{nearby_{frt}}$

$w_{overlap} > w_{nearby_{bck}}$

$w_{overlap}$= weights for road segments overlapped with the curve.

$w_{nearby_{frt}}$= weights for nearby road segments before the curve.

$w_{nearby_{bck}}$= weights for nearby road segments behind the curve.

3.3 DATA

Taking advantage of the most used data maintained by transportation agencies, this study proposed a methodological framework for preliminary network screening of roadway

segments that are most likely to contribute to a significant improvement in safety performance. Various data sources that are developed and maintained by TxDOT were used to perform the crash hotspot analysis, including Pavement Management Information System (PMIS), Crash Records Information System (CRIS), Roadway Inventory Database, and Highway Curves GIS Layer.

3.3.1 Pavement Management Information System (PMIS)

The PMIS is an advanced information management system implemented by TxDOT for storing, retrieving, and analyzing pavement condition data. It is designed to provide reliable information for monitoring pavement conditions, estimating maintenance and rehabilitation (M&R) needs, comparing the effects of multiple M&R alternatives, allocating available funding, etc. According to TxDOT (2015), PMIS contains more than 195,000 data collection segments covering the entire highway network in Texas. For most of these PMIS segments, the length is 0.5 miles; although the length of PMIS segments may vary slightly from one to another, the average length is about 0.5 miles.

3.3.2 Crash Records Information System (CRIS)

The CRIS is an automated database employed by TxDOT for collecting and managing statewide motor vehicle traffic crashes. The data source of CRIS is the Texas Peace Officer's Crash Report (form CR-3). In CRIS, the KABCO system is used to classify crash severity.

3.3.3 Roadway Inventory Database

The TxDOT's Roadway Inventory is a statewide dataset that contains various data attributes associated with GIS linework and roadway characteristics. These attributes are associated with multiple categories including identification/referencing, geographic,

administrative, operational, physical/cross-section, traffic, and common statistics (TxDOT 2020). The most recent ten-year annual roadway inventory data can be accessed via TxDOT Open Data Portal. The Roadway Inventory geodatabase contains predefined linear measurements (M-values) of the distance from the origin of a route (DFOs) which make it possible to locate events (e.g., traffic crash, pavement condition, construction project, etc.) along routes within the network. Hence, the Roadway Inventory data was used as a master map for integrating PMIS data with crashes and curves in a single GIS file.

3.3.4 Highway Curves Geographic Information System (GIS) Layer

The Highway Curves GIS Layer identifies and visualizes horizontal curves along state-maintained roadways across Texas. The attribute table includes degree of curve, curve classification, and roadway referencing fields. Curves are presented on a base map for visualization in a GIS environment.

3.4 CASE STUDY

A case study was conducted to implement the proposed methodology for analyzing spatial autocorrelation and assessing the safety performance of road segments to facilitate the identification of high-priority locations for potential improvement. This numerical study focused exclusively on road segments of interstate highways (IH) in Texas. Leveraging Python programming, data filters were applied to remove non-interstate roadways from all obtained datasets (i.e., PMIS, CRIS, Roadway Inventory, and Highway Curve). Also, crashes that were associated with intersections along IH routes were removed from the CRIS data since such crashes were out of the scope of the case study. Finally, data from 225,888 crashes that occurred on 25 IH routes (excluding intersections) throughout Texas from 2016 to 2018 were analyzed, as shown in Figure 3.3.

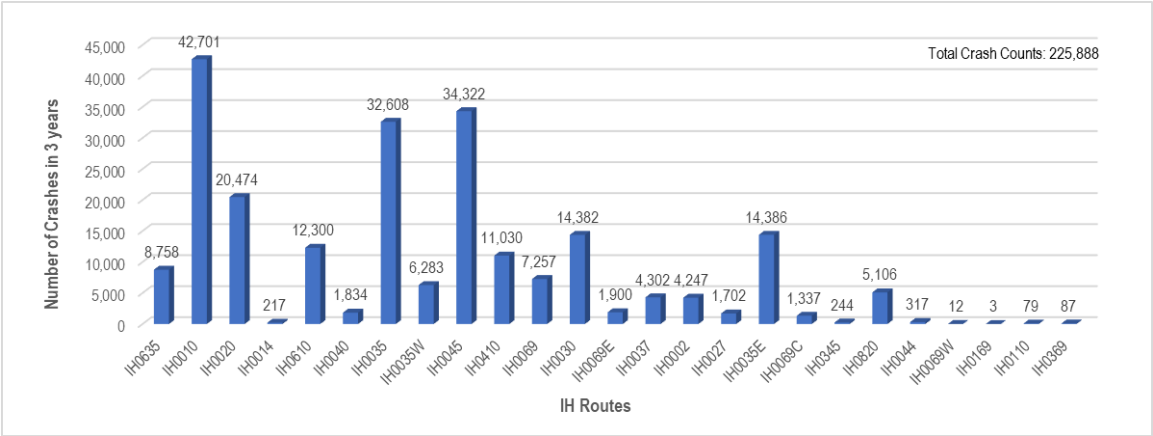


Figure 3.3 Distribution of Obtained Crashes in Texas IH Routes from 2016 to 2018

The case study consisted of two phases. In the first phase, the proposed CSI was employed to conduct a spatial autocorrelation analysis for all road segments within the network, leading to the generation of a crash hot spot map. The framework of the spatial analysis performed in the first phase is demonstrated in Figure 3.4.

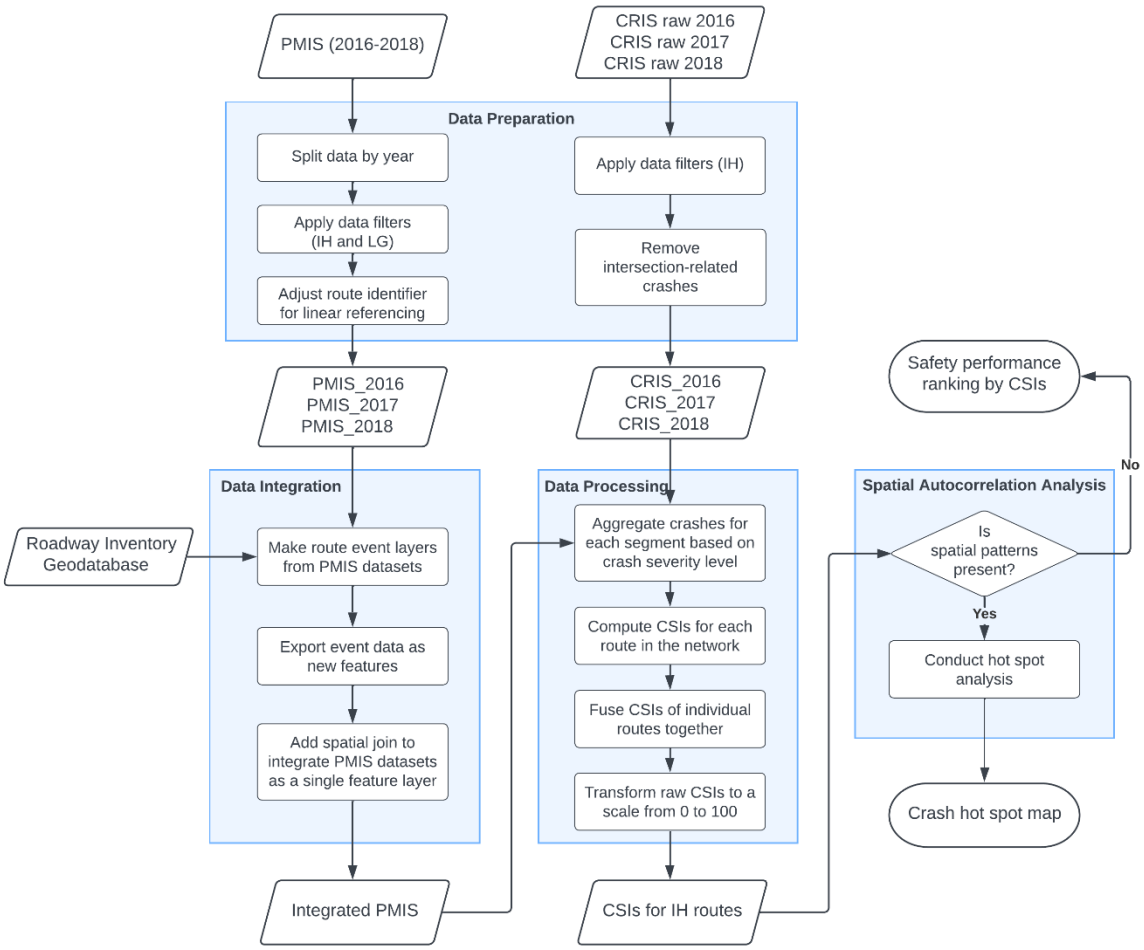


Figure 3.4 Framework of Developing a Crash Hot Spot Map Based on CSIs

In the second phase of the case study, the objective was to identify horizontal curves with the highest potential for safety improvement, utilizing the proposed performance metric SI. The framework for prioritizing horizontal curves is depicted in Figure 3.5.

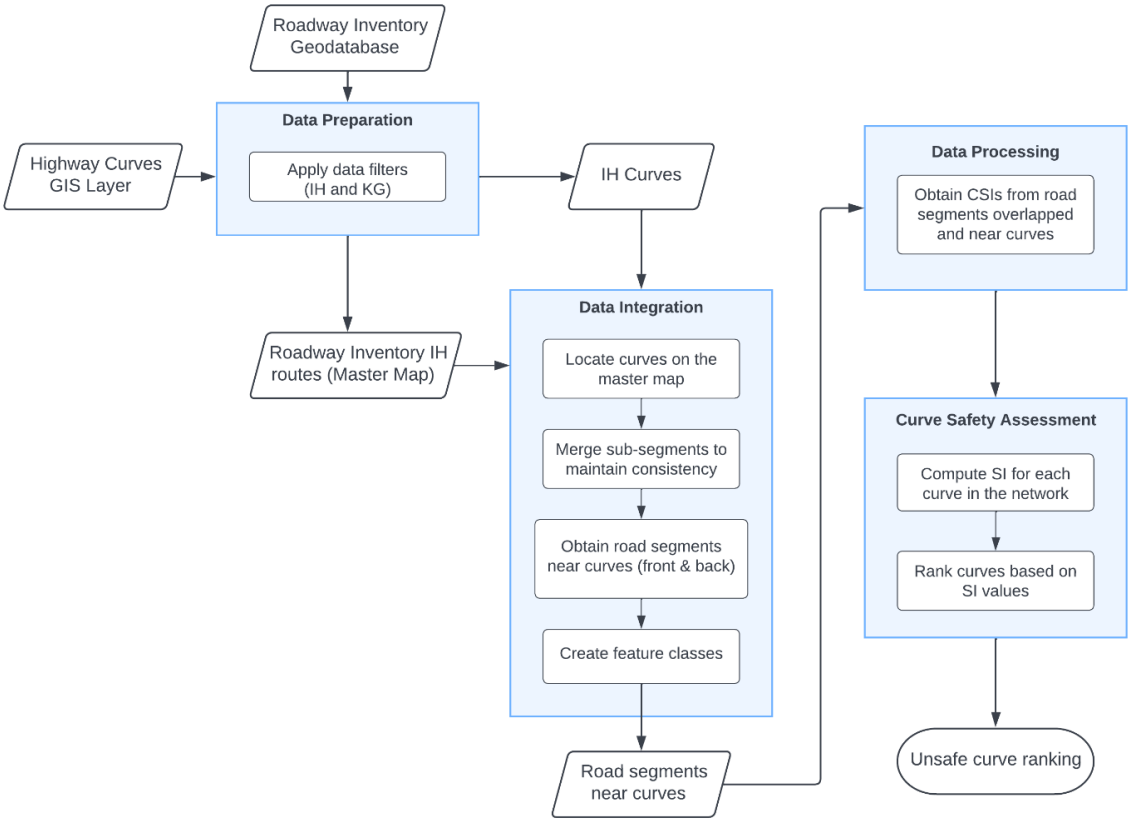


Figure 3.5 Framework of Prioritizing Horizontal Curves Based on SI Values

3.4.1 Data Preparation

To perform the spatial autocorrelation analysis, the authors first obtained PMIS 2016 – 2019 data but, upon examination, it was found that road segments in PMIS 2019 are different from that in PMIS 2016 – 2018. Specifically, the beginning and ending points of road segments are completely different in 2019 data from other years' data. This might be due to the implementation of automatic survey methods adopted since 2019, which consequently introduced new segmentation approaches for pavement condition data. Hence, PMIS 2019 was removed from the study. Correspondingly, CRIS 2016 – 2018 was acquired through the publicly automated crash data extraction approach. Since highway

curve data for the analysis period is not available, the Highway Curves GIS Layer used for the analysis was the one published in September 2022. As a matter of fact, roadway centerlines change frequently over time. To provide a better data visualization, this study used the most recent Roadway Inventory 2021 data which can provide a consistent roadway network as the one used for the base map of the Highway Curves GIS Layer.

In PMIS, condition data is collected for each roadbed. In other words, for divided highways that have two roadbeds, condition data from both directions will be collected simultaneously (Note: frontage roads are not of interest in this case study). Based on reviewing the obtained PMIS datasets, it was found that all the condition data associated with IH routes were collected for both directions: left main lane and right main lane. However, the location referencing information of a road segment in two directions was found to be slightly different. Due to this reason, PMIS data based on left main lanes in IH routes were used in the case study for implementing the proposed methodology. The total number of road segments derived from left main lanes in IH routes in PMIS datasets were 6,907 (2018), 6,821 (2017), and 6,869 (2016).

3.4.2 Data Integration

Originally, the obtained PMIS dataset was stored in a comma-separated values (CSV) file. To boost the efficiency of data preparation, this file was broken down into three individual datasets by year (i.e., 2016, 2017, and 2018). After examining and cleaning individual PMIS datasets, the plain text PMIS files were transformed into feature layers using the "*Make Route Event Layer*" function from linear referencing tools in ArcGIS Pro. Later, taking advantage of the linear referencing method stored in the Roadway Inventory geodatabase, the study integrated individual PMIS features into a single feature layer based on their spatial relationship. After data integration, a total of 6,357 road segments were

generated to support subsequent analysis. Useful attributes retrieved from PMIS road segments included location referencing (e.g., route name, beginning and ending DFOs), length of segment, and AADTs in three years. Figure 3.6 lists the distribution of road segments analyzed in the study by routes.

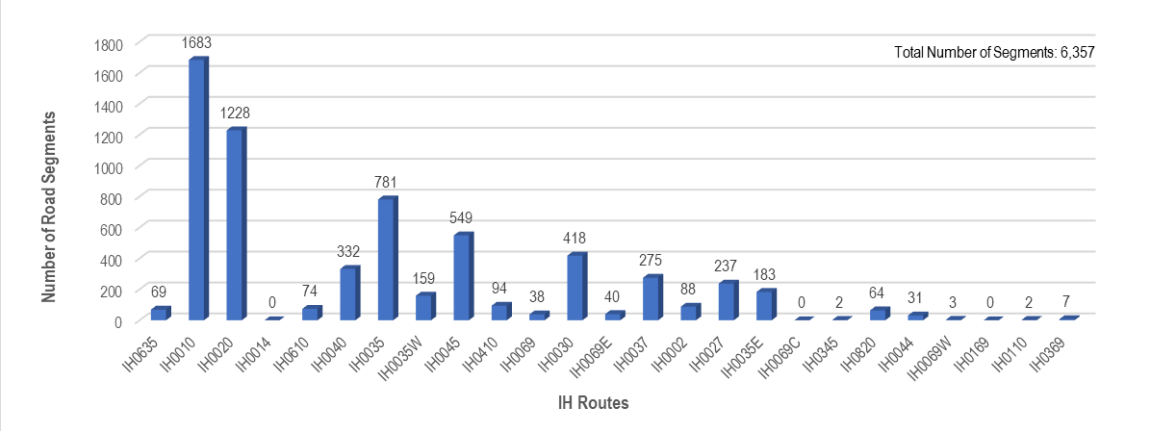


Figure 3.6 Distribution of Analyzed Road Segments in Texas IH Routes

3.4.3 Data Processing

With the divided road segments from the PMIS dataset, the three-year crashes obtained from CRIS 2016 to 2018 were aggregated based on the KABCO severity level. It is worth noting that crashes in CRIS were all mapped into the centerline of roadways by TxDOT. To maintain the consistency of the calculation, the study used half of the number of crashes for each level of severity. For example, if the total counts of crashes by severity for a road segment are: 2 (K), 4 (A), 5 (B), 10 (C), and 19 (O), the actual crash counts used for calculating CSI will be: 1 (K), 2 (A), 2.5 (B), 5 (C), and 9.5 (O).

In the analysis, the computation of CSIs was performed route by route and stored as separate datasets. After that, all the obtained CSI files were fused to construct a single dataset of raw CSIs. The distribution of the raw CSI dataset was highly skewed with a

heavy tail due to the presence of outliers. To address this issue, the 90th percentile of raw CSI values (i.e., 2,282.4) was used as an upper fence to regroup the data; any CSI value that was greater than the 90th percentile was replaced by the 90th percentile value of the raw CSIs. This regrouped CSI dataset was mapped into an interpretable scale ranging from 0 to 100. The histogram plots of CSIs in three stages of data transformation were shown in Figure 3.7.

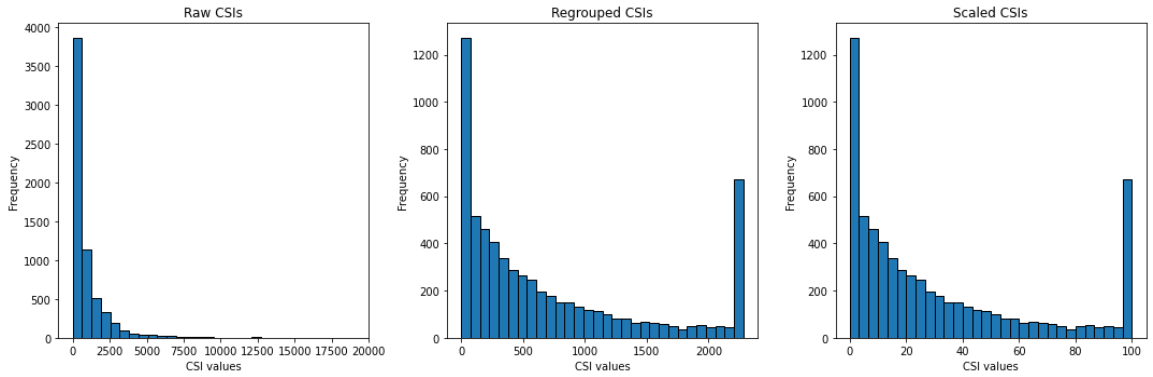


Figure 3.7 Histogram Plots of CSIs (a) Raw CSIs (b) Regrouped CSIs (c) Scaled CSIs

3.4.4 Spatial Autocorrelation Analysis

The scaled CSI data was used as the attribute to perform the spatial autocorrelation analysis. The inverse distance was selected as the conceptual model of spatial relationships to determine the neighbor relationships for the global spatial autocorrelation analysis based on Moran's I index. The inverse distance method has been proven to be appropriate and effective with models that nearby neighboring features are more likely to influence a target feature than features that are far away (ESRI 2022c). However, a problem with this method is that each feature is potentially a neighbor of other features. For a large dataset, it could be challenging to implement the computation. Hence, a distance band was specified as a

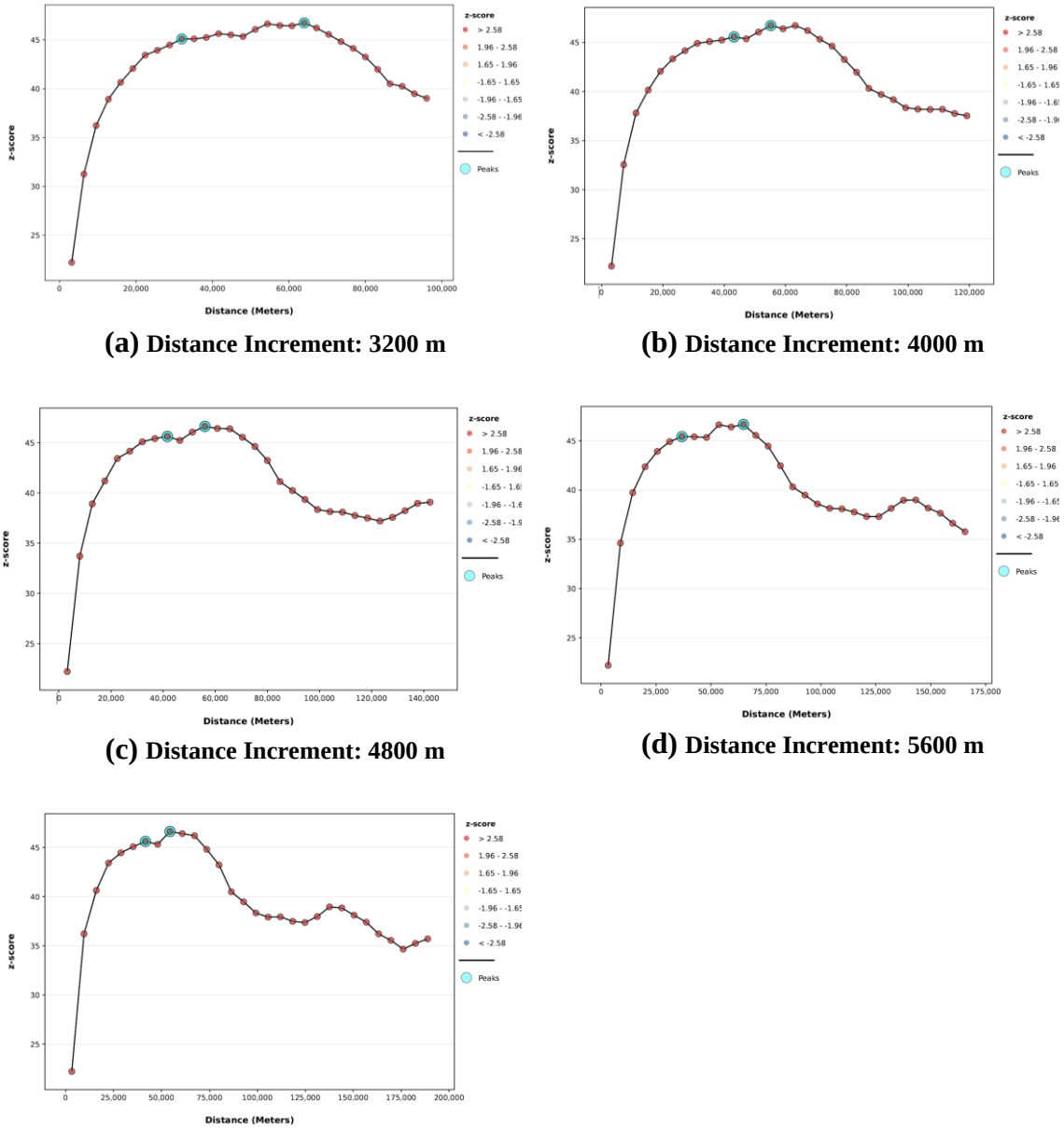
cutoff distance to reduce the burden of computation. Only features within the cutoff distance will be used to perform the analysis.

To determine the distance band, the study first used the tool "*Calculate Distance Band from Neighbor Count*" to get an initial distance. Due to the skewed distribution of the CSI data, the number of neighbors was set as a recommended value – eight – to yield reliable results (Grekousis 2020, ESRI 2022d). The identified average distance that can ensure every feature with eight neighbors was approximately 3,316 meters (2.06 miles).

After that, the "*Incremental Spatial Autocorrelation*" tool was employed to search for an optimal distance using the output obtained from the last step. This tool can examine the spatial autocorrelation for a series of distances based on the input parameters. The distances with statistically significant peak z-scores are where spatial clustering is most pronounced. Hence, the distance with the highest z-score was used as the distance band in this study. Three inputs required for the tool are beginning distance, distance increment, and number of distance bands. In this study, the beginning distance was set as 3,200 meters (near the identified average distance of 3,316 meters). To identify the peak value, multiple values of distance increment were tested as shown in Table 3.3 and Figure 3.8. The number of distance bands was set as 30, which is the maximum value accepted by the tool. Finally, the maximum peak identified by the tool was 64,000 meters (about 39.77 miles) which was then used as the distance band in the subsequent analysis.

Table 3.3 Peak z-score by Distance Based on Different Distance Increments

Distance Increments (m)	**First Peak**		**Max Peak**	
	Distance	z-score	Distance	z-score
3200	32000	45.101	**64000**	**46.735**
4000	43200	45.571	55200	46.730
4800	41600	45.634	56000	46.632
5600	36800	45.422	64800	46.661
6400	41600	45.634	54400	46.639



(a) Distance Increment: 3200 m

(b) Distance Increment: 4000 m

(c) Distance Increment: 4800 m

(d) Distance Increment: 5600 m

(e) Distance Increment: 6400 m

Figure 3.8 Spatial Autocorrelation by Distance Based on Different Distance Increments

The null hypothesis of the global Moran's I test states that the attribute of interest is randomly distributed among all features throughout the study area. In this study, the

attribute of interest was the CSI values and the features analyzed were road segments. The outcomes of the global spatial autocorrelation analysis are presented in Figure 3.9. As shown in the figure, the obtained Moran's Index was 0.104, z-score was about 35.911, and the p-value was 0.000. The likelihood of this clustered pattern of CSI values being a result of random chance was less than 1%. Consequently, the null hypothesis should be rejected based on the positive Moran's I index with a high z-score and a statistically significant p-value. This means that the variable analyzed (i.e., CSI) in this study was not randomly distributed among features (i.e., road segments) in the roadway network. In other words, high CSI values were more likely to be spatially clustered thus resulting in crash hot spots.

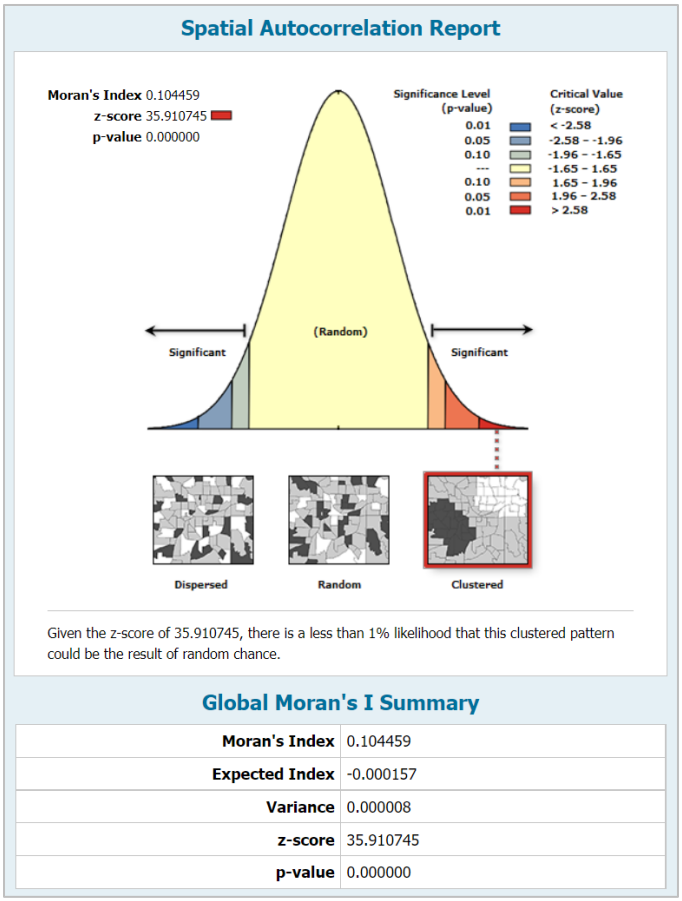


Figure 3.9 Results of the Spatial Autocorrelation

Leveraging the *Hot Spot Analysis (Getis-Ord Gi*)* spatial statistics tool provided by ArcGIS Pro, the statewide crash hot spot map was generated, as illustrated in Figure 3.10. Detailed zoom-in views are provided for some metropolitan areas (San Antonio, Houston, Dallas) where crash hot spots are clustered, as shown in Figure 3.11, Figure 3.12, and Figure 3.13. In these maps, segments marked as red are the identified crash hot spots with 95% confidence while segments marked as orange are hot spots with 90% confidence. Table 3.4 shows the critical p-values and z-scores for different confidence levels.

Table 3.4 Confidence Levels for Identified Crash Hot Spots

Confidence Level	**z-score**	**p-value**	**Color**
90%	< -1.65 or > +1.65	< 0.10	Orange
95%	< -1.96 or > +1.96	< 0.05	Red

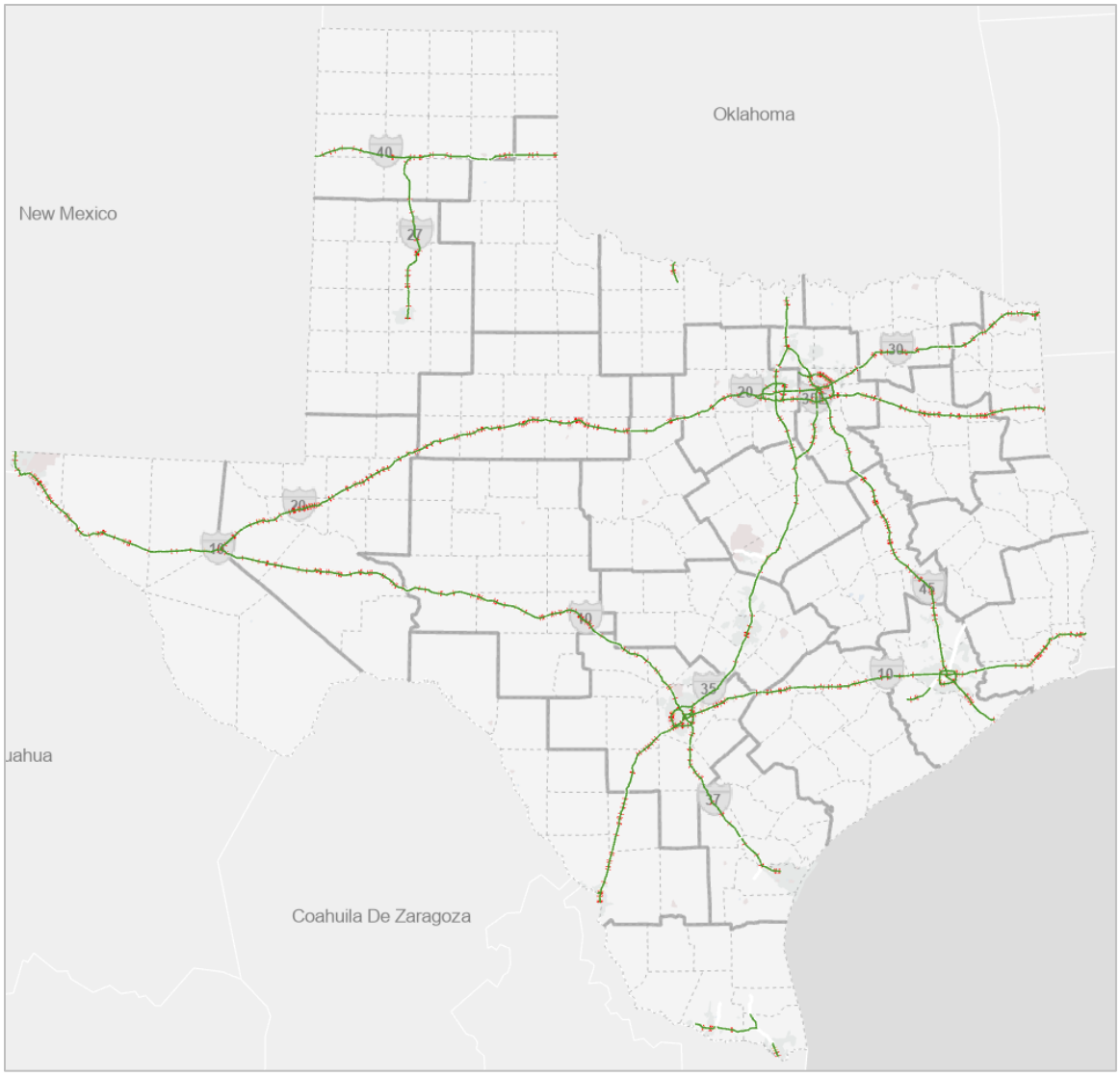


Figure 3.10 An Illustration of the Crash Hot Spot Map in Texas Based on the Proposed Measure (CSI)

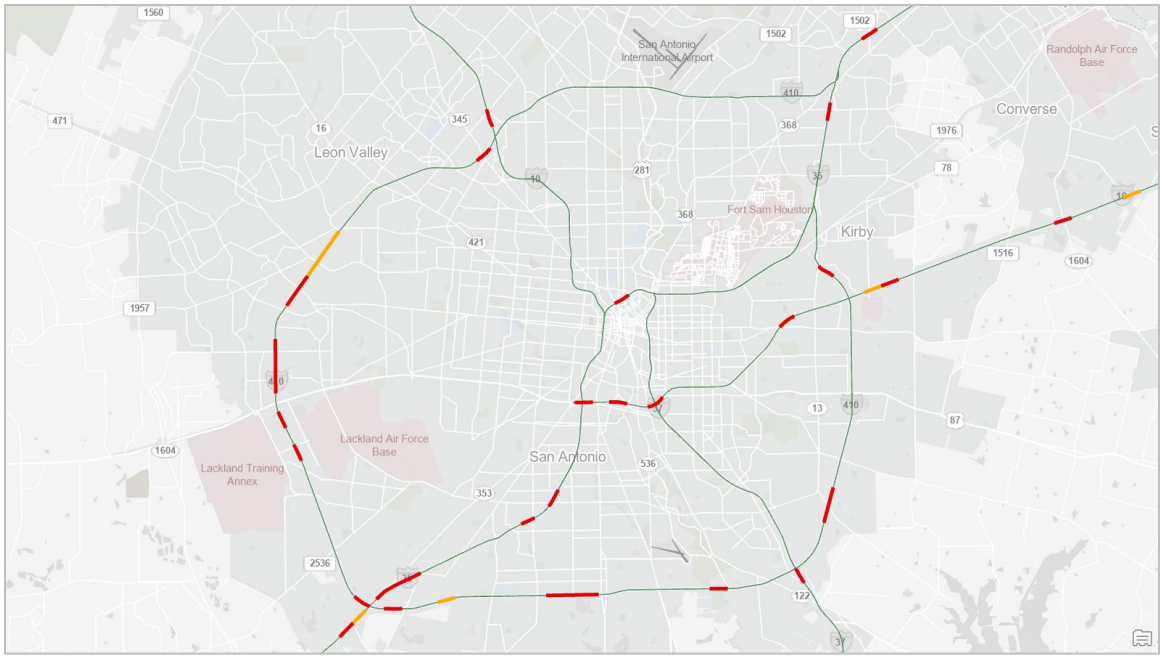


Figure 3.11 Identified Crash Hot Spots in San Antonio Based on CSI

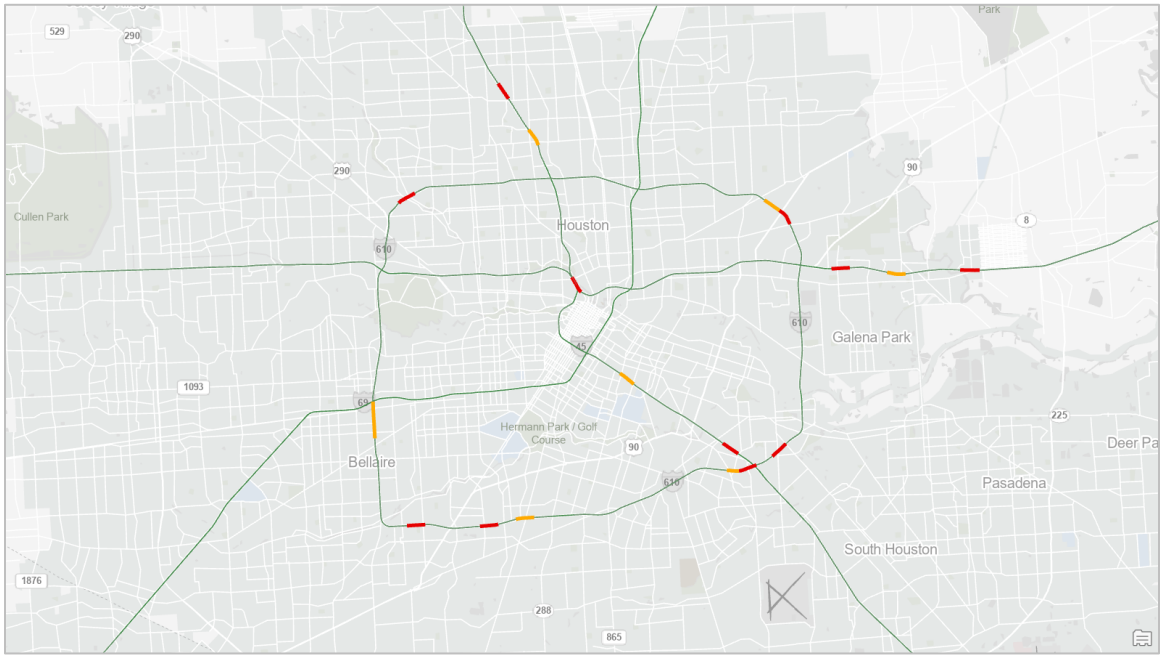


Figure 3.12 Identified Crash Hot Spots in Houston Based on CSI

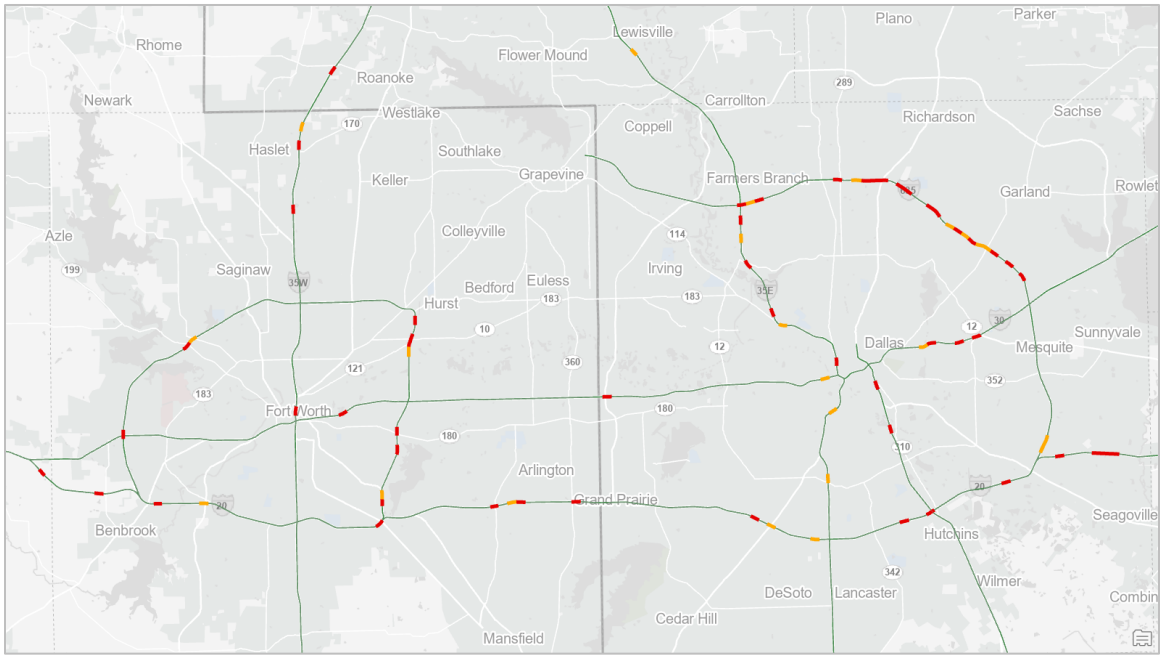


Figure 3.13 Identified Crash Hot Spots in Dallas Based on CSI

3.4.5 Safety Assessment of Horizontal Curves

The second phase of the case study aims to assess the safety performance of horizontal curves using the proposed measure SI. The step-by-step workflow of the automated procedure has been illustrated in Figure 3.5 and is summarized in detail as follows:

- **Step 1** – Locate horizontal curves using LRM from the master map in ArcGIS Pro. This step was necessary as DFOs retrieved from the Highway Curves GIS layer do not match the LRM used in the Roadway Inventory database.
- **Step 2** – Merge sub-segments to maintain data consistency. Due to the operation performed in the previous step, the number of curves with recalculated DFOs was greater than the number of curves obtained from the Highway Curves GIS layer. To be precise, the former has 1,955 data entries while the latter only has 1,791

curves. During the previous operation, a curve could be divided into several segments and each segment was associated with a new DFO. Hence, a Python program was developed to combine road segments within the same curve, assuring a curve only related to one set of DFOs (beginning and ending points).

- **Step 3** – Identify road segments near curves. A Python algorithm was developed to generate road segments that are nearby (before and behind) curves, as illustrated in Figure 3.14 and Figure 3.15. In this study, 0.1 miles was selected to be the distance used to define a nearby segment.
- **Step 4** – Obtain CSI values from identified road segments. The outputs from Step 3 were spatially joined with the regrouped CSI data to support the calculation in the next step.
- **Step 5** – Compute SI values for horizontal curves. A Python program was developed to implement the formula defined in Equation (12).
- **Step 6** – Rank high-risk horizontal curves based on the obtained SI values.

Figure 3.14 Components of Horizontal Curves

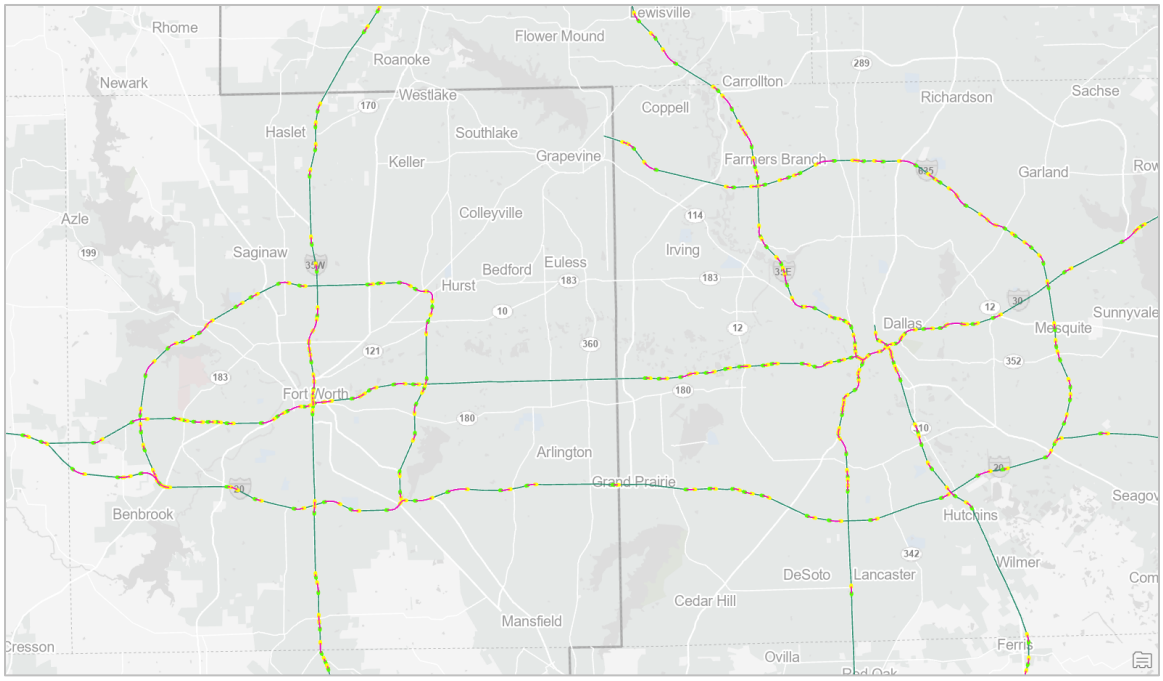


Figure 3.15 An Illustration of Identified Horizontal Curves in Dallas

The assumption was made that the speed limit is 70 miles per hour (approximately 31.3 meters per second). Considering that the nearby segment occurs approximately 5 seconds before the point of curvature, the distance was estimated to be 156.5 meters (equivalent to approximately 0.1 miles). Thus, for this study, a distance of 0.1 miles was chosen to define a nearby segment.

According to a NCHRP report (Campbell 2012), it is widely acknowledged that road segments preceding horizontal curves generally have a higher risk of crashes compared to segments following the curves. Based on this understanding, the weights in Equation (12) were assigned as follows: $\boldsymbol{w_{overlap} = 0.6}$, $\boldsymbol{w_{nearby_{frt}} = 0.3}$, and $\boldsymbol{w_{nearby_{bck}} = 0.1}$. Consequently, the SI values were computed based on these configurations. The top 10 horizontal curves with the highest SI values are presented in Table 3.5.

Table 3.5 Top 10 Unsafe Horizontal Curves Based on SI

RteName	Frm_DFO	To_DFO	LenOfSec	avgAADT	Total	K	A	B	C	O
IH0635	12.125	12.617	0.492	83441	112	1	7.5	28	22	53.5
IH0020	407.341	407.841	0.500	19704	46	1	0.5	0.5	6.5	37.5
IH0045	218.55	219.046	0.496	17713	22	0.5	0.5	1.5	0.5	19
IH0035	5.683	6.183	0.500	21565	19.5	0.5	0.5	1	1.5	16
IH0030	94.396	94.884	0.488	19181	19	0.5	0.5	1	2	15
IH0020	254.947	255.447	0.500	12220	16	0.5	1	1.5	3	10
IH0035	211.536	212.036	0.500	52390	13	1.5	1	1	0.5	9
IH0045	97.539	98.24	0.701	23433	10	1	0	3	1.5	4.5
IH0027	49.643	50.136	0.493	5407	6	0.5	0.5	0	1	4
IH0020	180.628	181.095	0.467	11128	6	1	0	1.5	0	3.5

All ten curves have the highest SI values which were computed based on the CSI values of their nearby road segments. Upon examining these CSI values, it was observed that all three components of the nearby segments associated with these curves exhibited the highest CSI values. Therefore, Table 3.5 is ranked according to the total number of crashes that occurred in each segment. The result demonstrates that the proposed method not only identifies horizontal curves with a high crash frequency but also highlights curves associated with severe crashes.

For example, examining the first row of Table 3.5 reveals that the identified curve had a comparably high AADT and experienced a total of 112 crashes during the analysis period. In contrast, the curve listed in the last row exhibited only 6 crashes. However, despite the lower AADT, this curve received a high SI score due to its significantly lower traffic volume but a substantial proportion of fatal and injury crashes, accounting for approximately 42% of the total crashes.

3.5 SUMMARY

In this part of the research, a GIS-based methodological framework was developed that utilizes the proposed safety performance measures to identify crash hot spots and rank high-risk horizontal curves within a network. The study introduces a composite performance measure (i.e., CSI), based on the EPDO approach, to assess roadway safety performance. Using this measure, the study aims to identify sites where improvements have the potential to significantly reduce crash frequency or severity. Additionally, the study introduces a novel safety measure (i.e., SI) which is designed as a function of CSI. The application of these measures is demonstrated through a case study. In the case study, the CSI measure is employed to generate a crash hot spot map using spatial autocorrelation analysis. Later, the SI measure is computed from CSI values of road segments associated with curves and then used to quantify and prioritize the safety performance of horizontal curves.

To sum up, the proposed method effectively identifies hazardous locations which can serve as candidates for further detailed investigation. The findings from the study indicate that the proposed methodological framework holds great potential for local transportation agencies in their efforts to screen roadway networks and pinpoint locations with significant safety needs. By utilizing this method, agencies can prioritize their resources and interventions effectively, leading to enhanced safety performance for all road users.

Chapter 4: A Review of Advancements in Artificial Intelligence Applications for Pavement Management[6]

4.1 INTRODUCTION

With the continuous advancement in automatic data acquisition devices, computer vision techniques, and machine learning (ML) algorithms over the past decades, artificial intelligence (AI) technology has increasingly been incorporated into research and practice in pavement engineering and related fields. Researchers have conducted various reviews of existing automated survey devices for collecting pavement condition data (Coenen and Golroo 2017, Ragnoli et al. 2018), computer vision-based image processing algorithms for distress detection (Zakeri et al. 2017, Gopalakrishnan 2018, Koch et al. 2015, Mathavan et al. 2015), ML models for crack detecting (Hsieh and Tsai 2020) and predicting pavement performance (Justo-Silva et al. 2021), and applying neural networks for pavement engineering (Ceylan et al. 2014). However, few publications are devoted to systematically synthesizing how general AI algorithms have been applied to the various stages of pavement management in its entire life cycle.

To fill this critical gap, this study aims to provide a comprehensive review of studies on AI applications in pavement management along with some highlighted perspectives on future research. Precisely, the objectives of the review are to:

- Review state-of-the-art articles published in archival journals from 2015 to 2020.
- Synthesize the key findings based on three broad categories of pavement management activities: distress evaluation, performance modeling, and maintenance and rehabilitation (M&R) programming.

[6]based on Xu, Y., & Zhang, Z. (2022). Review of applications of artificial intelligence algorithms in pavement management. *Journal of Transportation Engineering, Part B: Pavements*, 148(3), 03122001. https://doi.org/10.1061/JPEODX.0000373

- Transfer the latest understanding of applying advanced AI algorithms to address real-world problems in pavement management.
- Help the audience better understand the current practices, research frontiers, and future directions in applying AI-oriented techniques to pavement management.

The enormous potential of high-performance computation and autonomous learning capability makes AI a promising tool to address more generic engineering problems. Even though AI is a fast-growing field that encompasses multiple subfields, such as ML, neural networks (NN), computer vision, expert systems, robotics, fuzzy logic, and natural language processing (Strong 2016), its application to pavement management is relatively new. This paper covers some of the commonly used AI algorithms in pavement management; all the reviewed AI algorithms that have been applied to various areas of pavement management are summarized in Table 4.1.

Table 4.1 Summary of Reviewed AI Algorithms in Pavement Management

AI algorithms		Pavement Management						
		Distress Evaluation			Performance Modeling			M&R Programming
		Detection	Classification	Quantification	Roughness	Distress	Other indices	
Artificial Neural Network (ANN)	Multi-Layer Perceptron (MLP)	✓	✓		✓	✓	✓	✓
	Back Propagation Neural Network (BPNN)	✓	✓		✓			✓
	Radial Basis Function (RBF)	✓	✓		✓	✓	✓	
Deep Neural Network (DNN)	Convolutional Neural Network (CNN)	✓	✓	✓				
	Recurrent Neural Network (RNN)	✓					✓	
	Long Short-Term Memory (LSTM)				✓			
Tree-based Algorithms	Decision Tree (DT)	✓	✓		✓	✓		✓
	Random Forest (RF)	✓	✓	✓	✓		✓	✓
	Boosting	✓		✓		✓	✓	
Support Vector Machine (SVM)	Classifier	✓	✓	✓			✓	✓
	Regression				✓		✓	
Reinforcement Learning	Deep Q-Learning (DQN)							✓

4.2 DISTRESS EVALUATION

Assessing the current condition of pavements in the transportation network under study is an essential part of pavement management, enabling transportation agencies to track pavement conditions over time and predict future conditions. Raw data on pavement conditions are collected by conducting periodical condition data collections in the field either manually and/or automatically. Over the past decade, an increasing number of automated data acquisition equipment has been gradually adopted to replace manual pavement surveys due to their high accuracy, low safety risk, and less subjective properties, among other advantages of automated data collection solutions (Attoh-Okine and Adarkwa 2013).

Previously, Coenen and Golroo (2017) presented a thorough literature review of specialized data-collecting devices and techniques for automated pavement condition surveys. A rough technology appropriateness analysis was conducted to illustrate the current practice of different distress detection instruments. Ragnoli et al. (2018) evaluated the applicability and adequacy of several commonly used pavement distress survey devices in terms of their impacts on safety and comfort. According to a synthesis report from the National Cooperative Highway Research Program (NCHRP) (Pierce and Weitzel 2019), as of 2019, approximately 80 percent of the surveyed agencies have completed the transition from manual to automated pavement condition surveys. However, half of these agencies use semi-automated methods to analyze the obtained data. This means the condition assessment still involves some level of human interaction. Therefore, there is still a research gap in fully automated pavement condition assessment from survey data. Zakeri et al. (2017) studied commonly used image-based techniques for detecting and quantifying surface cracking in flexible pavements. Methods and algorithms reviewed by this paper mainly focused on noise elimination, image normalization and segmentation, feature

engineering, and distress classification. In a similar study (Gopalakrishnan 2018), the existing applications of deep convolutional neural networks (CNNs) for computer vision-based automated pavement image analysis and distress detection were studied. Koch et al. (2015) reviewed and assessed several computer vision-based methods and algorithms for detecting cracks and defects in bridges, tunnels, underground pipes, and pavements. In addition to two-dimensional image processing, the current available three-dimensional imaging methods for detecting pavement distress and monitoring pavement conditions over time were studied by Mathavan et al. (2015). More recently, Hsieh and Tsai (2020) conducted a review of ML and deep learning (DL) algorithms used for crack detection and provided the development trend in pixel-level crack segmentation.

In general, automated pavement condition assessment can be broken down into image processing, image segmentation, and distress evaluation (Sun et al. 2016). With the rapid development of AI techniques, various algorithms have been applied to pavement distress evaluation since 2015. This section presents key findings from the reviewed studies on AI algorithms and/or techniques for pavement distress evaluation which comprises: 1) distress detection, 2) distress classification, and 3) distress quantification.

4.2.1 State of Practice in Distress Detection

In this review, pavement distress detection refers to determining the existence of given distress on pavement surface. Conventional algorithms widely applied to the automatic detection of pavement distress can be grouped into three classes: 1) image thresholding, 2) patch-based classification, and 3) depth-based methods (Eisenbach et al. 2017). In addition to these conventional algorithms, an increasing number of recent studies have leveraged deep CNN frameworks and commonly used ML models for detecting the presence of pavement distress.

Deep Neural Networks

Traditional ML models usually require manual feature extraction to achieve high performance. On the contrary, the special architecture of networks enables deep neural networks (DNNs) to learn more intricate structures in high-dimensional data (Szegedy et al. 2013, Goodfellow et al. 2016). This allows DNNs to automatically perform feature extractions from raw data with no extra need for manual feature engineering (Hinton et al. 2006, Goodfellow et al. 2016). This advantage has made DNNs a widely applied algorithm in image-based pavement distress detection.

In order to eliminate substantial human interventions in traditional automated crack detection methods, Zhang et al. (2017a) developed a CNN-based tool (CrackNet) to detect cracks from 3D asphalt pavement surface images. Based on the results from comparison experiments, it was concluded that CrackNet performed more robustly than a supervised learning model (Pixel-SVM) and a non-AI crack detection method (3D shadow modeling). Overall, CrackNet achieved high accuracy in pixel-level crack detection. However, the authors pointed out that time-consuming and failures of recognizing thin cracks were the major downsides of CrackNet. To improve the learning capability of CrackNet, Zhang et al. (2018a) proposed an upgraded CNN model - CrackNet II - which eliminated handcrafted features and added more hidden layers in the network. This improved configuration enabled CrackNet II to run approximately five times faster than CrackNet. Moreover, CrackNet II was less susceptible to local noises resulting in higher accuracy in detecting fine cracks compared to CrackNet. However, it was reported that CrackNet II had difficulty in detecting global context; accordingly, it misclassified certain noise patterns as pavement distress. Later, another improved CNN-based model, CrackNet-V, was presented in (Fei et al. 2019). Compared with CrackNet, CrackNet-V was more effective in detecting fine cracks and 3-4 times faster in computation efficiency. A

comparison study indicated that CrackNet-V outperformed a ML-based model presented in (Shi et al. 2016) in terms of precision and F1-score.

In a recent study, Zhang et al. (2019) developed a RNN-based model, CrackNet-R. It was comprised of a sequence generation module, a sequence modeling module, and an output layer. The pixel probability was computed based on the average timely probability and the feature vector. Then, it was used as an indicator to quantify the likelihood that a pixel belongs to a crack. CrackNet-R achieved higher accuracy than CrackNet in terms of recognizing fine cracks and distinguishing noise in the background. This improvement was largely attributed to the nature of RNN which enabled the model to learn the pattern of the input sequence.

Gopalakrishnan et al. (2017) pointed out that the complexity of pavement distress images (e.g., inhomogeneity of cracks, diversity of surface texture, existence of background noise, and presence of multiple distresses) makes it challenging for the existing pavement crack detection methods to achieve high performance in pavement distress detection. Thus, the study proposed a CNN model (VGG-16-DCNN) to automatically detect cracks using pavement images from the Long-Term Pavement Performance (LTPP) database. To achieve high accuracy and generalization, transfer learning technique was applied by pretraining the model on the ImageNet dataset. After pretraining, VGG-16-DCNN was used to generate semantic image vectors which were utilized to train and test a final classifier for pavement crack detection. Various ML classifiers (e.g., ANN, SVM, and RF) were tested to obtain the optimal final classifier for the proposed network. The optimal classifier was found to be a single-layer ANN. In another study on pavement crack detection (Bang et al. 2019), a deep convolutional encoder-decoder network was also found to be effective in detecting pixel-level cracks from black-box images. Similarly, another

study (Ukhwah et al. 2019) confirmed the effectiveness of applying an object detection algorithm, You Only Look Once (YOLO), to detect potholes on pavement surfaces.

In addition to investigating the effectiveness of CNNs in pavement distress detection, several studies have been conducted to compare the performance of CNNs with other algorithms in distress recognition. As a case in point, a CNN-based pavement crack detection method (CNN-CDM) presented by Nhat-Duc et al. (2018) found that the CNN-based model significantly outperformed edge detection-based methods (i.e., Sobel and Canny). The results indicated that both the classification accuracy rate and precision of CNN-CDM were higher than those of the other CNN benchmark models presented in (Pauly et al. 2017) and (Gopalakrishnan et al. 2017). In another study on pavement crack identification (Zhang et al. 2016), the authors developed a supervised CNN crack detector that can automatically learn discriminative features from manually annotated images collected by smartphones. The study concluded that the proposed CNN crack detector outperformed the other two models based on SVM and Boosting methods, respectively. Similarly, the CNN-based model presented by Cha et al. (2017) performed more robustly and resulted in less noise when compared to the traditional Sobel and Canny (Canny 1986) edge detection methods. This confirmed the effectiveness of using deep learning approaches for pavement crack detection.

Tree-based Algorithms

In conjunction with DNNs, tree-based algorithms (e.g., DT, RF, and Gradient Boosting Machines) have also served as a reliable approach that has been widely used in detecting pavement distress due to their high accuracy, stability, and effectiveness in handling complex, non-linear relationships (Howard and Bowles 2012, Ali et al. 2012, Natekin and Knoll 2013, Freund and Schapire 1997, Schapire 2013). As seen in (Shi et al.

2016), a RF (i.e., CrackForest) was employed to predict the structured token for a given crack image patch. To find the optimal classifier for differentiating cracks from noises, the authors tested various classification algorithms such as k-nearest neighbors (KNN) and SVM. Among these methods, the combination of CrackForest and SVM achieved the best performance in terms of precision and recall scores. Finally, the authors concluded that the performance of the proposed CrackForest was much better than that of other methods presented in Canny (Canny 1986), CrackTree (Zou et al. 2012), CrackIT (Oliveira and Correia 2014), and minimal path selection (Avila et al. 2014).

Traditional methods for crack detection either contain too many false detections or depend largely on parameter selection (Cord and Chambon 2012). To overcome these drawbacks, Cord and Chambon (2012) applied an Adaptive Boosting (i.e., AdaBoost) model to road images for detecting the existence of pavement distress. The AdaBoost score of each sub-image was calculated based on descriptors selected by the model. Then, the scores were used to classify the sub-image into the "defect" or "no defect" group. Through a comparative study, the authors concluded that the proposed AdaBoost method outperformed a model presented in (Oliveira and Correia 2009), where statistical analysis was used.

Support Vector Machines

SVM is a widely used supervised learning algorithm to solve classification and regression problems. It has been applied to plenty of fields which include but are not limited to facial expression classification, text categorization, image classification, financial forecasting, and cancer diagnosis. As a commonly used classification algorithm, SVM has been applied to identify pavement distress. Ai et al. (2018), for example, developed a pavement crack detection method at the pixel level. The method was an integration of

probabilistic generative model (PGM) and SVM algorithm. The PGM model was used to create a probability map using pixel intensity information. The SVM model was employed to calculate the probability of a pixel belonging to a crack using neighborhood information, thus constructing the second probability map. These probability maps were then fused to achieve high accuracy. The results indicated that the proposed SVM outperformed the unsupervised multi-scale fusion crack detection method presented in (Li et al. 2018) in terms of precision, recall, and F1-score which are commonly used performance evaluation metrics for classification problems.

Miscellaneous Algorithms

On top of investigating the effectiveness of a specific AI algorithm in detecting the presence of pavement distress, some efforts have also been made to compare the performance of different algorithms in pavement distress detection. Mokhtari et al. (2016) compared the performance of four classification algorithms for detecting cracks from pavement images. These algorithms included ANN, DT, k-NN, and adaptive neuro-fuzzy inference system (ANFIS). The performance of the models was evaluated by the percentage of missed cracks and false alarms. Based on these criteria, the ANN model was observed to be significantly superior to other models. In contrast, the k-NN model missed more than half of the cracks while the DT model missed one-third of the cracks. For computation time, the DT method was reported much faster than the other three models. Finally, the authors concluded that the ANN and the ANFIS models were more suitable and compatible for pavement crack detection.

Recently, Politis et al. (2020) applied remote sensing and data mining techniques for pavement surface condition assessment at the network level. In this study, the authors compared the performance of various classification algorithms (e.g., k-NN, naïve Bayes,

SVM, and MLP) in terms of detecting pavement conditions using spectral attributes extracted from multispectral images. According to the results, the model based on the MLP method achieved the highest accuracy.

4.2.2 State of Practice in Distress Classification

Pavement distresses can be divided into multiple distinct groupings by the characteristics of the distress and the material of the surface. Using asphalt concrete surfaces as an example, the distresses can be classified as cracking (e.g., block cracking, longitude cracking, and transverse cracking), patching and potholes, surface deformation (e.g., rutting and shoving), surface defects (e.g., bleeding and raveling), and miscellaneous distresses (e.g., water bleeding and pumping) (Miller and Bellinger 2014). In some cases, a pavement section may present more than one type of distress. Hence, the objective of pavement distress classification is to automatically assign a specific type to each of the detected distresses.

Deep Neural Networks

Employing deep architectures in neural networks, CNNs have achieved remarkable success in pattern recognition for image inputs due to their superior multi-scale high-level image representations (Sun et al. 2020, Wang et al. 2016). This advantage enables CNNs to automatically detect import features from images, which makes CNNs a promising candidate for pavement distress classification. Recently, some efforts have been focused on leveraging CNNs to classify the detected pavement distresses into diffident categories.

In order to upgrade the labor-intensive and time-consuming manual inspection of pavement distress images, Eisenbach et al. (2017) proposed two CNN-based pavement distress detection and classification approaches: ASINVOS and ASINVOS-mod net. The

ASINVOS-mod net was a modified version of ASINVOS net. It replaced all the max-pooling layers in the original ASINVOS net with a filter map. Also, ASINVOS-mod net employed the size preserving convolutions to increase the explanatory power of the network. It was concluded that the proposed CNN-based methods were able to recognize various pavement distresses such as cracks, potholes, inlaid patches, applied patches, and bleedings. To further evaluate the effectiveness of ASINVOS and ASINVOS-mod net, they were compared with another CNN-based road crack detector developed by (Zhang et al. 2016) and an image thresholding method - CrackIT (Oliveira and Correia 2014). The results indicated that CNN-based methods significantly outperformed CrackIT in terms of F1-score and balanced error rate. Similarly, Li et al. (2020) employed CNN models to classify pavement image patches into five crack-related categories. Different configurations of CNNs were tested to investigate the relationship between the size of the receptive field and the training time of the proposed network. It was reported that all tested CNNs achieved high accuracy (above 94 percent). The study also revealed that the size of the receptive field can significantly affect the training time but has little impact on the performance of the network.

Another deep learning study was presented in (Zhang et al. 2018b) where a CNN-based approach with transfer learning was employed to classify pavement image pixels into cracks, sealed cracks, and background regions. The CNN model was compared to other methods including a traditional Canny edge detection (Canny 1986) method and two recently developed crack detectors presented in (Oliveira and Correia 2014) and (Shi et al. 2016). According to the results, the proposed CNN-based method achieved the highest scores in terms of recall, precision, and F1-score. In a similar study, Nie and Wang (2018) applied Faster R-CNN models with transfer learning to classify pavement distresses into

different categories. The average classification accuracy achieved by the optimal Faster R-CNN model was 96.53 percent.

More recently, a deep learning framework based on YOLOv3 was developed by Lei et al. (2020). The study aimed at detecting and classifying pavement distress using 19,665 street view images collected via Baidu API. Based on the results, the authors concluded that the proposed YOLOv3 method achieved satisfactory results with an 88 percent for average accuracy rate. In the same way, Du et al. (2020) employed a YOLOv3 network for detecting and classifying pavement distress into seven categories. The overall detection accuracy achieved was 73.64 percent. As demonstrated by the comparison study, the proposed YOLOv3 was much faster than the Faster R-CNN and SSD network. Moreover, this study found that the performance of the proposed method improved steadily as increasing the number of training data. Finally, it was concluded that deep learning methods were efficient in automated pavement distress detection and classification.

Likewise, Maeda et al. (2018) implemented a CNN-based object detection method, originally presented by Huang et al. (2017), for detecting and classifying pavement distress into eight categories using a road image dataset. In this study, an object detection framework based on a single DNN was applied to detect distress in pavement images. Two CNN-based models (Inception V2 and MobileNet) were selected as feature extractors because of their high processing speeds and accuracy. Overall, the CNN-based method can generate acceptable results for distress detection and classification. Finally, the authors pointed out that the performance of DNNs could be effectively improved by increasing more road images for certain types of distress or redefining the types of pavement distress.

Tree-based Algorithms

Tree-based algorithms have been found to be highly effective for solving classification problems over the past decade (Farid et al. 2014, Fernández-Delgado et al. 2014). Some achievements have been made in applying tree-based algorithms to pavement distress classification. As presented in (Rodriguez-Lozano et al. 2020), where an ensemble method comprised of a decision tree generated by the C4.5 algorithm, a logistic model tree (LMT), and a rule-based classification algorithm (named RIPPER) was developed to classify pavement images into different cracking types. In this study, the vertical and horizontal projects of the images were calculated and used as features for crack classification. Finally, it was concluded that the proposed ensemble method was more reliable and accurate than all the individual algorithms (C4.5, LMT, and RIPPER) if they were used alone; in addition, the proposed ensemble algorithm outperformed other ML-based models presented in (Cubero-Fernandez et al. 2017, Hoang and Nguyen 2019, Li et al. 2014) in terms of precision and recall rates.

Another example of such applications was presented in (Cubero-Fernandez et al. 2017), where a DT algorithm was applied to detect and classify cracks on pavement surfaces. The basic methodological framework for implementing crack detection and classification documented in this study (Cubero-Fernandez et al. 2017) is very similar to the methodological process depicted in (Rodriguez-Lozano et al. 2020). The results indicated that the DT model achieved an average of 88 percent success rate in detecting cracks and an 80 percent success rate in classifying detected cracks into the correct types.

Miscellaneous Algorithms

Over the past five years, several efforts have been made in comparing the effectiveness of different algorithms in categorizing pavement distress into different

groups. In a study, Hoang and Nguyen (2019) applied multiple supervised learning algorithms (e.g., ANN, SVM, and RF) to automatically recognize and classify crack patterns in asphalt pavement images. First, a Gaussian steerable filter was applied to distinguish crack regions from the background. Then, projective integrals of the images were computed and used to extract features from the images. Based on the results of a statistical hypothesis test, the accuracy rate of the SVM model was found to be significantly higher than that of the ANN or RF model.

In another comparative study of pavement crack classification, Hoang and Nguyen (2018) compared six supervised learning models including Naïve Bayesian Classifier (NBC), Classification Tree (CT), Backpropagation Artificial Neural Network (BPNN), Radial Basis Function Neural Network (RBFNN), SVM, and Least Squares Support Vector Machine (LSSVM). The LSSVM model outperformed other models in terms of the overall classification accuracy rate. Moreover, the statistical test also indicated that the performances of the LSSVM model and the SVM model were significantly better than other models.

4.2.3 State of Practice in Distress Quantification

Going one step further, distress quantification aims to quantify the severity and extent of the detected distress based on the outcomes from detection and classification. Distress quantification is an integral part of pavement condition evaluation and contributes to further decisions on M&R project and treatment selection.

Deep Neural Networks

In addition to CNNs, some state-of-the-art DNN-based object detection techniques include, but are not limited to, YOLO, Single Shot MultiBox Detector (SSD), SegNet,

Darknet, VGG-16 and VGG-19, ResNet, GoogLeNet, MobileNet, and Microsoft COCO. Not only do DNNs perform robustly on object detection and classification but they also achieve high accuracy in localizing objects of various classes (Szegedy et al. 2013).

For improving the efficiency of automated pavement distress assessment, two state-of-the-art deep learning models, YOLO-v2 and Faster Region Convolutional Neural Network (Faster R-CNN), were presented in (Majidifard et al. 2020) to automatically classify pavement distresses into nine distress types. A pavement image dataset (PID) was developed using street-view images retrieved from Google Maps, which includes both wide-view and top-down-view images. The former sets were used to classify distresses into nine categories; the latter image sets were used for quantifying the density of distresses. The performance evaluation showed that the YOLO-v2 model performed better in terms of F1 scores and confusion matrices. Finally, the authors concluded that deep learning algorithms provided a convenient, cost-effective, and accurate approach for pavement condition assessment.

In a similar study, Tong et al. (2017) applied CNN to automatically recognize, locate, quantify, and model concealed cracks that are developed under the surface of asphalt pavements. Three CNN-based models were proposed for different purposes: the recognition model for detecting concealed cracks, the location model for pinpointing the location and measuring the length of cracks, and the feature extraction model for selecting appropriate shape features to create 3D concealed crack models. It was concluded that the recognition model could detect concealed cracks using GPR images with no error and the location model was also able to achieve satisfying results for measuring the length of cracks.

Support Vector Machines

It is noted that both DNN and SVM can achieve high performance in image-based classification or object detection (Liu et al. 2017). Unlike DNNs, SVM algorithms utilize kernel functions to map data points from a low-dimensional space into a higher-dimensional space. Based on this data transformation, a separating hyperplane is created to split the data points in the higher-dimensional space by their class. This mechanism enables SVMs to work effectively on image processing.

Although automated pavement condition vehicles and devices can significantly improve the efficiency of pavement condition assessment, not all transportation agencies could afford the high cost of purchasing and operating them (Hadjidemetriou et al. 2018). To lower the overall cost of automated detection and evaluation of pavement distress, Hadjidemetriou et al. (2018) proposed a SVM model to identify and quantify pavement patches using pavement surface images. The image datasets used in this study were retrieved from three video sources. These videos were recorded by a smartphone inside the vehicle in daytime and a GoPro camera outside the vehicle in both daytime and nighttime. The images were first converted to grayscale and then decomposed into small square blocks. For each small square block, a feature vector was calculated and served as input for the SVM model. The SVM model then classified the small block into patch or no-patch. A morphological operation was applied to calibrate the results from the SVM model. After that, the whole image was classified as including a patch if more than 50 patch blocks were connected. The area of the detected patch was quantified based on the number of connected patch blocks and the area covered by each block. The authors observed that the images from the GoPro camera in daytime yielded the best performance for patch classification and quantification.

Miscellaneous Algorithms

Recently, some studies have compared the performance of various supervised learning algorithms in classifying pavement distress into different severity levels. To improve the reliability and accuracy of automated raveling assessment, Tsai et al. (2021) applied three ML-based algorithms (AdaBoost, SVM, and RF) to detect and classify raveling into different severity levels using 3D pavement data. It was reported that the RF and SVM models outperformed the AdaBoost classifier, and the RF model achieved a lower misclassification rate compared to the SVM model. Finally, the authors concluded that ML-based algorithms were capable of automatically detecting and measuring asphalt pavement raveling. It was suggested that future studies may utilize deep learning methods to improve the overall efficiency of distress evaluation in a real-world implementation.

4.2.4 Achievements and Limitations

The reviewed state-of-the-art studies where various AI algorithms were applied to facilitate the evaluation of pavement distress are summarized in Figure 4.1 by their specific usage for distress detection, classification, and quantification. It can be observed from the figure that the majority of studies in pavement distress evaluation focus mainly on leveraging AI algorithms to detect one specific pavement distress. From these studies, cracking was found to be the most targeted distress; this makes sense as cracking is the most observed distress in pavements. For distress classification, most of the reviewed studies take full advantage of DNN methods to recognize patterns from pavement distress images. Among the reviewed papers, only a few AI-oriented studies utilized the outcomes from pavement detection and classification to further quantify the extent and severity of the detected distress. The accuracy of such quantification studies depends largely on the outcomes of distress detection and classification. Although only a few articles are

published on AI-oriented distress quantification over the past few years, an increasing number of transportation agencies have switched to fully automated pavement condition data collection (Serigos et al. 2015, Pierce and Weitzel 2019). Most of the automated survey vehicles use their algorithms to automatically detect, classify, and quantify pavement defects from the obtained images. More efforts are expected to evaluate the validity and performance of these devices.

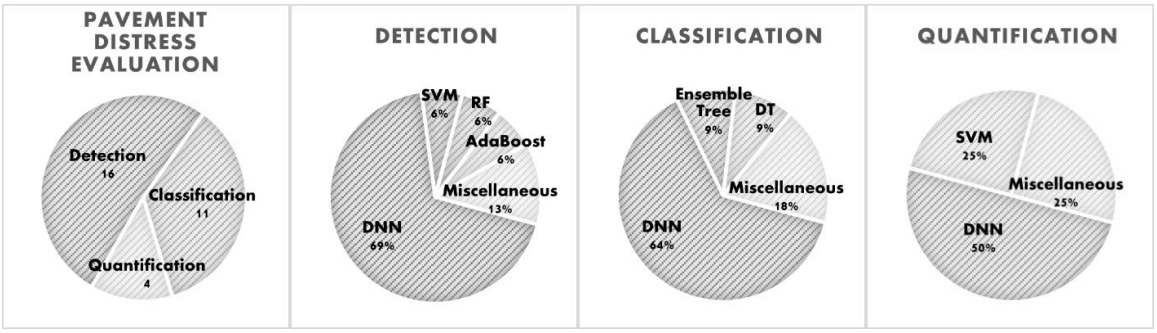


Figure 4.1 Distribution of Reviewed AI Applications in Pavement Distress Evaluation

For supervised learning algorithms, feature engineering (i.e., feature design, extraction, selection, and combination (Oakden-Rayner et al. 2017)) has a significant impact on the accuracy of the results. Among the reviewed studies, pavement distress detection and classification were typically treated as classification problems that could be addressed by supervised learning algorithms. Examples of such algorithms include but are not limited to DT, ensemble tree method (e.g., AdaBoost), RF, SVM, and KNN. Generally, the performance of supervised learning algorithms depends largely on manually generated features. Some studies have adopted dimensionality reduction techniques to reduce the complexity of a model and avoid overfitting. For example, the principal component analysis (PCA) method was used by Shi et al. (2016). In practice, however, manual feature engineering is not only a time-consuming process but also susceptible to bias and error;

hence, it limits the generalization ability of the algorithm (Khurana et al. 2017, Zhang et al. 2017b). Furthermore, there is no guarantee that the optimal features can always be found because complex high-level features are often difficult to observe directly (Oakden-Rayner et al. 2017). Consequently, such methods are biased and could overlook key factors affecting the results. To overcome this issue, an increasing number of studies on distress detection have turned to deep learning algorithms which integrate ANNs with feature learning techniques. These techniques allow a model to automatically learn the features from raw data and then use them for prediction purposes. As a commonly applied deep learning method in pavement distress detection and classification, CNN has demonstrated reliability and validity in automatically extracting features from images with no human intervention (Bang et al. 2019, Ukhwah et al. 2019, Zhang et al. 2018b, Nie and Wang 2018, Nhat-Duc et al. 2018). Moreover, this research also identifies that precision, recall, and F1-score are the top three measures used for assessing the performance of deep learning networks in pavement distress evaluation. Table 4.2 summarizes the reviewed publications using various AI algorithms for pavement distress evaluation.

Table 4.2 Summary of Reviewed Publications Using AI algorithms for Pavement Distress Evaluation

AI Algorithms		Detection	Classification	Quantification
Deep Neural Network (DNN)	Convolutional Neural Network (CNN)	Zhang et al. (2017a, 2018a); Fei et al. (2019); Gopalakrishnan et al. (2017); Bang et al. (2019); Nhat-Duc et al. (2018); Zhang et al. (2016); Cha et al. (2017); Pauly et al. (2017)	Eisenbach et al. (2017); Li et al. (2020); Zhang et al. (2018b); Nie and Wang (2018); Maeda et al. (2018)	Majidifard et al. (2020); Tong et al. (2017)
	Recurrent Neural Network (RNN)	Zhang et al. (2019)	--	--
	You Only Look Once (YOLO)	Ukhwah et al. (2019)	Lei et al. (2020); Du et al. (2020)	Majidifard et al. (2020)
Tree-based Algorithms	Decision Tree (DT)	Mokhtari et al. (2016)	Rodriguez-Lozano et al. (2020); Cubero-Fernandez et al. (2017);	--
	Random Forest (RF)	Shi et al. (2016)	Hoang and Nguyen (2019, 2018)	Tsai et al. (2021)
	Adaptive Boosting (AdaBoost)	Cord and Chambon (2012)	--	Tsai et al. (2021)
Support Vector Machine (SVM)	SVM Classifier	Ai et al. (2018); Politis et al. (2020)	Hoang and Nguyen (2019, 2018)	Hadjidemetriou et al. (2018); Tsai et al. (2021)
Miscellaneous Algorithms		Mokhtari et al. (2016); Politis et al. (2020)	Hoang and Nguyen (2019, 2018)	Tsai et al. (2021)

Before starting with applying AI algorithms to pavement distress evaluation, it is necessary to consider the potential consequences of any assumption made on the data. When using a specific dataset, a study has already assumed the validity and adequacy of the data. As observed in most reviewed studies, the performance of a deep learning network depends significantly on the availability of data used to train and test the algorithm. Table 4.3 summarizes several datasets developed in previous studies for pavement distress evaluation. In general, deep learning algorithms have high demands for the scale of the dataset. However, some of such studies have encountered difficulties in obtaining adequate data. The lack of valid data can significantly compromise the validity of the method since data is the foundation of all AI-oriented models. As revealed in (IJ 2018), the purpose of AI-oriented learning algorithms is to find patterns in data. If the data used to train the algorithm fails to reflect the truth, then any result derived from the data will be biased and questionable. Hence, any study intended to utilize AI-oriented methods should first thoroughly examine the validity and quantity of the collected data to ensure that the data can sufficiently characterize the problem of interest. In addition to that, if a study has to rely on limited data to achieve desired accuracy and generalization, other strategies should be explored to overcome this limitation. For example, transfer learning is one of the viable methods that enable models to repurpose previously learned features from other tasks to solve new problems. Several reviewed studies have successfully employed transfer learning to improve the performance of pavement distress evaluation (Majidifard et al. 2020, Bang et al. 2019, Zhang et al. 2018b, Nie and Wang 2018, Gopalakrishnan et al. 2017).

Table 4.3 Publicly Accessible Datasets for Pavement Distress Analysis

Name	Reference	Number of Images	Data Collection Device/Method	Country
Pavement Image Dataset (PID)	Majidifard et al. (2020)	7,237	Google application programming interface (API)	USA
Street-view images	Maeda et al. (2018)	9,053	Smartphones	Japan
German Asphalt Pavement Distress (GAPs)	Eisenbach et al. (2017)	1,969	The S.T.I.E.R measuring vehicle was developed by LEHMANN+PARTNER GmbH.	Germany

Furthermore, it is worth noting the quality of pavement image can significantly affect the performance of any proposed method, as the presence of noise in image data can significantly weaken the effectiveness of a distress detection or classification model. Nevertheless, compared to traditional edge detection methods, deep learning algorithms are less susceptible to noise and can still perform robustly in crack detection under various conditions of the raw images and achieve higher accuracy (Cha et al. 2017). Meanwhile, it has been found that 3D pavement images are less susceptible to lighting conditions and result in reduced noise (Zhang et al. 2017a). 3D images are superior to 2D images in many ways: the surface condition information in 3D images is more comprehensive and accurate; 3D pavement images have higher quality and are less sensitive to noises (e.g., oil stains, dark shadows, and tire marks); and the depth information stored in 3D images can largely improve the detection of pavement distresses such as cracking, surface textures, and rutting (Li et al. 2019). These advantages of 3D pavement images would allow the model to achieve high performance under different environmental conditions, thus substantially boosting the effectiveness of distress detection and classification. Although some traditional distress detection algorithms (e.g., depth-checking methods, 3D shadow modeling, and hybrid crack detection) have also been tested on 3D images, these methods

are insufficient to learn from labeled example data (Zhang et al. 2019). On the contrary, AI-oriented algorithms can fully exploit the information in 3D pavement condition images; thus, improving the overall accuracy of pavement distress evaluation (Zhang et al. 2017a, Zhang et al. 2019). Future studies in this field may also benefit from replacing 2D images with 3D surface images to maximize the performance of AI-oriented algorithms.

4.3 PERFORMANCE MODELING

Performance modeling is a process that aims at characterizing the general deterioration trends of pavements. Pavements deteriorate over time due to cumulative impacts on roadways. Such impacts include both external and internal factors. External factors typically include traffic loads, environmental stresses, frequency of M&R interventions, and construction quality. Internal factors encompass pavement types and material characteristics. Accurate performance modeling enables agencies to gain a better understanding of how pavement conditions change over time, make more informed decisions on scheduling major M&R activities, conduct more reliable budget needs analysis, and allocate available resources with a higher level of confidence.

In general, commonly collected pavement condition data can be classified into four categories: surface distress, roughness, structural capacity, and skid resistance (Stevenson 2021). By analyzing road condition data, agencies can characterize pavement performance with condition indices. In current practices, not all transportation agencies use uniform condition indices because the equipment and data processing techniques employed for pavement condition data collection vary substantially from one agency to another. Examples of widely used pavement condition indices include pavement condition index (PCI) for rating surface distress, International Roughness Index (IRI) for quantifying

roughness, Structural Condition Index (SCI) for evaluating structural capacity, and Skid Number (SN) for skid resistance.

4.3.1 State of Practice in Pavement Roughness Prediction

Pavement roughness is a key indicator for pavement performance evaluation. Roughness is typically evaluated by the International Roughness Index (IRI), which is a measurement required by the Federal Highway Administration (FHWA) for the Highway Performance Monitoring System (HPMS) (Steudle et al. 2012). Over the past decade, a significant number of studies have been published leveraging various AI algorithms to improve the accuracy of predicting IRI.

Artificial Neural Networks

Inspired by the functionality of the human brain, ANNs introduce a novel approach to processing highly complex data. In practice, ANNs have proved to be extremely effective in a wide range of research fields such as control systems, pattern recognition, signal processing, medical diagnosis, and reliability analysis. In the past years (2015-2020), ANNs have become one of the most widely applied algorithms for pavement IRI prediction.

Ziari et al. (2016) compared ANN models with the group method of data handling (GMDH) approaches for IRI prediction using pavement structure, condition, traffic, and climate data from the LTPP database. Based on the prediction assessment, the authors concluded that the proposed ANN models were able to make reliable IRI predictions for both short-term and long-term periods; in contrast, GMDH models were incapable of accurately predicting IRI values.

To predict one-year-ahead pavement roughness, Georgiou et al. (2018) developed models based on ANN and SVM algorithms using time series IRI values. It was shown that the proposed ANN model with one hidden layer could achieve more accurate predictions. For the SVM regression model, the RBF was used as the kernel function. A statistical evaluation indicated that the ANN models slightly outperformed the SVM models. Finally, the authors stated that both ANN and SVM were effective in accurately predicting the one-year-ahead IRI values.

It is noted that the effect of climate factors varies from place to place (Hossain et al. 2019). For this reason, separate ANN models were created to predict IRI for ten sampled flexible pavements from four climatic zones throughout the U.S. (Hossain et al. 2019). The optimal model obtained in this study was an ANN with 7 features on the input layer, two hidden layers each with 9 neurons, and an output layer for generating the predicted IRI. The final average RMSE for all ten samples obtained by this model was 0.015. More recently, a similar study using ANN models with traffic and climate data to predict IRI for rigid pavements was conducted by the same research group (Hossain et al. 2020). The coefficient of determination (R^2) achieved by the proposed ANN models were all close to one. These studies illustrated that the effectiveness of the ANN algorithms in IRI prediction is independent of the type of pavements.

As discussed earlier, ANNs have been proven to be an effective algorithm for predicting pavement roughness by many studies. However, this does not mean that ANNs can always outperform other models. In a recent study, Yamany et al. (2020) compared an ANN model with two regression models in pavement performance prediction. Environmental conditions, traffic volume, and pavement age data from eight midwestern states in the LTPP database were used as inputs. The ANN was found to be the optimal model when all eight states were considered. However, for individual states, regression

models made more accurate predictions than ANN-based models in three out of eight states. Finally, the authors suggested that local agencies should only select appropriate models that work best for their pavement performance prediction.

Hybrid Methods

To boost the efficiency of ANN algorithms, a few efforts have been made to develop hybrid models which integrated ANNs with other algorithms. These hybrid models represent an innovative approach to pavement performance prediction. In recent research on IRI prediction (Dong et al. 2019), a hybrid model was developed based on feature fusion techniques that combined the cross-sectional features and the time-series ones. Particularly, a Long Short-term Memory (LSTM) model was utilized to train the time-series features, and the cross-sectional features were processed separately by a Back Propagation Neural Network (BPNN) model. The dataset encompassed more than 2,000 data points retrieved from the LTPP database. The considerable scale of the dataset enabled the model to achieve favorable prediction results. Consequently, the R^2 value for the IRI prediction obtained by this hybrid deep learning model was 0.867. In a similar study (Mazari and Rodriguez 2016), the authors developed a composite model integrating gene expression programming (GEP) with ANNs to predict the IRI using pavement structure and traffic data obtained from the LTPP database. It was concluded that the hybrid model was effective for IRI prediction.

Tree-based Algorithms

As one of the most versatile algorithms (Howard and Bowles 2012), RFs have been widely applied to solve problems including classification, regression, and clustering. For a classification problem, each decision tree in a RF outputs a class value, and the class that

has the most votes will be the final decision of the forest. For regression modeling, the final prediction of a RF is an average of the values from all trees in the forest. RFs can provide high-accuracy prediction and reliable generalization using ensemble algorithms and random sampling methods (Ali et al. 2012). In addition to the widely applied ANN prediction models, RF is another type of prevailing supervised learning algorithm that has been utilized to make predictions on pavement roughness.

As an example, a random forest regression (RFR) model was developed by Gong et al. (2018b) for predicting the IRI of asphalt pavements. The input features encompassed multiple distress scores, layer thickness, traffic flows, and climatic conditions. Based on the feature importance ranking, the initial roughness was found to be the most important factor for predicting future IRI values. According to the goodness-of-fit testing, the R^2 value for the training dataset was 0.998 and for the testing was 0.974. Furthermore, a comparative study indicated the proposed RFR outperformed the linear regression method.

Miscellaneous Algorithms

Instead of focusing on the performance of an individual algorithm, some studies have compared the effectiveness of multiple AI algorithms in roughness prediction. To identify the most significant factors affecting pavement performance in warm regions, Zeiada et al. (2020) conducted a study using five algorithms including ANN, regression DT, SVM, ensemble trees, and Gaussian process regression (GPR). A feature selection method called Forward Sequential Feature Selection (FSFS) was used in combination with the ANN model to reduce the number of input features. The results indicated that ANN models achieved the best performance among all five models. Similar results were presented in an earlier study (Zeiada et al. 2019) which utilized an ANN-FSFS model to investigate important pavement performance factors in cold regions.

As increasing numbers of algorithms are applied to pavement performance prediction, it is inevitable for concerned researchers to evaluate the performance of these prediction models. In a study of model evaluation, Kargah-Ostadi and Stoffels (2015) proposed a framework to compare the performance of various pavement prediction algorithms including ANN, RBF, and SVM. In this study, the learning processes of the selected algorithms were assessed regarding effectiveness, efficiency, and reliability. The authors pointed out that a model with high accuracy might lack generalization capability. Therefore, to select proper algorithms for pavement performance prediction models, it is necessary to evaluate not only the accuracy but also the generalization ability and the complexity of the algorithms.

4.3.2 State of Practice in Pavement Surface Distress Prediction

Surface distress is another factor that has a significant impact on the performance of pavements. Pavement Condition Index (PCI), originally developed by the U.S. Army Corps of Engineers, is a composite indicator that combines different types of pavement surface distress to represent the overall surface distress condition of a pavement. The calculation procedures and corresponding equations of PCI for each type of pavement are documented in (Shahin 2005). The following sections present some recent studies that have applied various AI algorithms to predict PCI based on available physical measurements of pavement distress.

Artificial Neural Networks

In general, it is more challenging to predict a composite index (e.g., PCI) than an individual index (e.g., IRI) since a composite index is a combination of multiple individual indices. The interconnected neurons between different layers enable ANNs to effectively

deal with nonlinear problems and enhance the generalization ability of ANNs. Consequently, ANNs have been widely applied in pavement PCI prediction.

As seen in a recent study, Karballaeezadeh et al. (2020) proposed an ANN-based method for PCI prediction using non-destructive falling weight deflectometer (FWD) testing data collected from 236 freeway segments. Two types of neural networks, MLP and RBF, were employed in this study. To achieve the best performance, the MLP network was integrated with two optimization algorithms: Levenberg-Marquardt (LM) and scaled conjugate gradient (SCG). Similarly, the RBF was combined with a genetic algorithm (GA) and imperialist competitive algorithm (ICA). Later, an integrated model (CMIS) was created by combing these four networks (MLP-LM, MLP-SCG, RBF-GA, and RBF-ICA) with weighted coefficients. Based on the results, the integrated CMIS model achieved the highest scores.

In the same way, Mahmood et al. (2020) applied ANN to predict network-level pavement deterioration trends for flexible pavements. The study employed a feed-forward network that was optimized by the Levenberg-Marquardt algorithm. To identify the optimal configuration, the authors explored multiple ANN architectures by modifying the number of hidden layers and the number of neurons in each hidden layer. Overall, the best performance was achieved by the network with three hidden layers. It was concluded that the ANN model was efficient in predicting PCI. Meanwhile, the authors stated that the difficulty in interpretation and lack of generalization would be the major downsides of the method.

Tree-based Algorithms

Gradient boosting (GB) is a popular boosting method that adds the predictions from weak classifiers together to obtain an ultimate powerful classifier. GB can overcome the

major drawback of DT (i.e., overfitting) by creating a single tree each time and then using the new tree to adjust the weights of observations and calibrate errors caused by previous trees. Recently, some studies have explored the validity of using GB algorithms for PCI prediction.

Piryonesi and El-Diraby (2020) used DT and gradient boosted trees (GBT) to predict the short-term deterioration in PCI with cost-effective data. The PCI values were first computed by a Python program using the method documented in (ASTM 2009). Then, the importance of selected features was evaluated by various algorithms such as information gain, Chi-squared, Gini index, GBT, among others. The results found the PCI prediction was significantly affected by several elements including the initial PCI value, latest M&R treatment, pavement age, AADT, and climate factors. After selecting the input features, two types of DT algorithms and their corresponding GBT models were trained and tested. The GBT models achieved higher accuracy than the individual classifiers. Based on the observations, the authors concluded that using data from similar climatic regions can help improve the accuracy of performance prediction models. Furthermore, the study recommended that state agencies record major M&R activities because information on M&R history can help improve the accuracy of pavement condition prediction.

Similarly, Barua et al. (2020) applied a gradient boosting machine (GBM) to make predictions on pavement deterioration concerning PCI. The dataset was comprised of pavement records for both runways and taxiways at Chicago O'Hare International Airport. To identify the optimal configuration of the GBM model, grid search was utilized for hyperparameters tuning; also, cross-validation was applied to prevent overfitting. The R^2 and RMSE achieved by the taxiway GBM model were 0.86 and 8.1 for the testing data. The runway GBM model performed slightly better with R^2=0.91 and RMSE=4.0 for the testing data. Furthermore, the authors compared the performance of the proposed GBM

models with other ML algorithms including linear regression, quadratic regression, ANN, and RF. The results indicated that the GBM models significantly outperformed other models.

4.3.3 State of Practice in Prediction of Other Pavement Condition Indices

In addition to pavement roughness and surface distress, several recent efforts have been made to apply AI algorithms for predicting other pavement functional and structural performance indices.

Artificial Neural Networks

As noted previously, ANNs have achieved great success in pavement performance prediction over the past few years. Being a popular method for performance modeling, ANNs can not only model the complex nonlinear relationships between dependent and independent variables but also detect the possible interactions between predictor variables (Tu 1996).

This can be seen in (Janani et al. 2020) where an ANN model was utilized to explore the correlation between functional characteristics (i.e., roughness and surface distress) and structural performance (i.e., structural number) for flexible pavements. The ANN model was configured with a 7-25-1 architecture (i.e., 7 neurons on the input layer, 25 neurons on the hidden layer, and one output layer for generating the effective structure number). It was concluded that the ANN can make an accurate prediction for the structural number based on material and function attributes. Similarly, another study proved that ANN models can achieve an accurate prediction of the effective structural number based on roughness data from the LTPP database (Sollazzo et al. 2017).

To improve the accuracy of rutting prediction in flexible pavement, Gong et al. (2018a) developed two neural networks (NN3 and NN20). A conclusion was made that the proposed NN20 with 20 input parameters achieved the best performance; moreover, the proposed method could effectively improve the R^2 by more than 30 percent and reduce the standard error by approximately 50 percent as compared to the calibrated transfer function in the Mechanistic-Empirical Pavement Design Guide (MEPDG).

Support Vector Machines

SVMs are widely used in classification problems as illustrated previously. In addition to classification, SVMs can also be applied to regression problems. Specifically, the types of SVMs that are used for regression are known as Support Vector Regression (SVR). The sparse solution and good generalization enable SVRs to achieve accurate predictions in regression (Awad and Khanna 2015).

Recently, Karballaeezadeh et al. (2019) developed a model, named SVR-PF, by integrating a SVM regression model with a particle filter to predict the remaining service life (RSL) of pavements using pavement thickness and surface temperature data. The predictions from the SVR-PF model were compared with the actual RSL values from HWD filed test. The results indicated that the SVR-PF method achieved a 95 percent accuracy in RSL prediction. Furthermore, the performance of the SVR-PF method was compared with the other two algorithms, normal SVM and MLP neural networks. The authors concluded that SVR-PF outperformed the other algorithms in terms of RMSE, MSE, correlation coefficient, and Nash-Sutcliffe model efficiency.

Hybrid Methods

Some studies have reported that hybrid models that integrate two AI algorithms can largely improve the accuracy of pavement performance prediction. In an early study, although it was conducted in 2013 still worth to be highlighted due to its uniqueness, Tabatabaee et al. (2013) presented a two-stage hybrid model which integrated support vector classification (SVC) with a RNN to predict pavement performance in terms of the present serviceability index (PSI). The SVC model was applied to classify pavement sections into different structure types. Then, the RNN model predicted PSI values using the outcomes from the SVC model and other pavement-related variables (e.g., age, traffic, maintenance history, and previous-year PSI). The results indicated that the SVC-RNN model can capture the underlying relationship between input parameters and PSI, thus leading to more accurate predictions. Furthermore, the proposed SVC-RNN model also outperformed a single-stage RNN.

Tree-based Algorithms

The eXtreme Gradient Boosting (XGBoost) is an advanced implementation of gradient boosted decision trees. It is designed for achieving high execution speed and model performance (Brownlee 2016). Compared to ANNs, XGBoost has many advantages including higher efficiency, feasibility, accuracy, and shorter processing time (Chen and Guestrin 2016). XGBoost has achieved great success in classification and regression predictive modeling for structured or tabular datasets (Brownlee 2016). Some researchers have successfully employed XGBoost to predict pavement performance indices.

As a case in point, an XGBoost was presented by Mousa et al. (2019) to predict the long-term effectiveness of structural overlays in terms of pavement service life (PSL). Four pavement condition indices were employed to evaluate the performance of the overlays,

which encompassed PCI, random cracking index (RCI), roughness index (RI), and rutting index (RUTT). An XGBoost classifier was applied to classify the lowest PSL value into three groups. It was concluded that the proposed XGBoost achieved acceptable results. Likewise, Gong et al. (2019) utilized an XGBoost model for alligator cracking (AC) and longitudinal cracking (LC) prediction using data from a NCHRP Project. Two separate models were built for AC and LC prediction, respectively. The grid search method and cross-validation were employed for hyperparameter tuning. Moreover, RFRs were used as comparative models. According to the results, the proposed XGBoost models outperformed RFRs; both XGBoost models and RFRs significantly outperformed the transfer functions in MEPDG. Meanwhile, the authors observed that overfitting has largely affected the performance of the proposed XGBoost models. To mitigate this issue, it was recommended that the number of trees be reduced, or a larger dataset be used as possible solutions.

4.3.4 Achievements and Limitations

Among the studies reviewed in this section, as illustrated in Figure 4.2 and Table 4.4., most of the studies reviewed have confirmed the effectiveness of AI-oriented methods in pavement performance modeling. The majority of efforts are dedicated to pavement roughness prediction in terms of IRI. This is mainly because IRI, being one of the most commonly used pavement condition indicators, has been recognized and/or used as 1) a standard measure for road roughness by transportation agencies throughout the world (Múčka 2017); and 2) an explanatory variable to predict other pavement condition indices (Al-Omari and Darter 1994, Sollazzo et al. 2017, Elhadidy et al. 2019, Park et al. 2007). The fast development of advanced AI algorithms leads to easier accessibility of ready-to-use tools (e.g., ML libraries and packages). As a result, an increasing number of studies

have applied various AI algorithms for the prediction of composite indices that represent the overall condition of pavements (e.g., PCI).

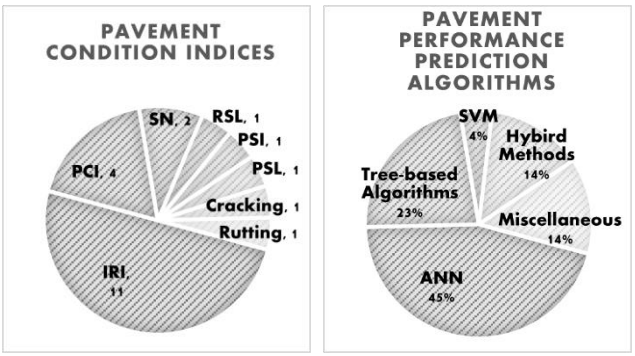


Figure 4.2 Distribution of Reviewed AI Applications in Pavement Performance Prediction

Table 4.4 Summary of Major Publications Presented in Pavement Performance Modeling

Index	Reference	Algorithms	Metrics	Data Source	Number of Data	Features
IRI	Ziari et al. (2016)	ANN	RMSE, R^2, VAF, MAPE, NDEI, and CF	LTPP	205	Pavement thickness, traffic volume, climate factors (precipitation, temperature, and freezing index), and pavement age
	Georgiou et al. (2018)	ANN (BPNN), SVM	R^2, MAE, and RMSPE	Private data	--	Discrete time series - annual roughness data over 7 years
	Hossain et al. (2019)	ANN	RMSE	LTPP	200	Climate factors (temperature, freezing index, humidity, and precipitation), and traffic volume
	Hossain et al. (2020)	ANN	RMSE, MAPE, and R^2	LTPP	200	Climate factors (precipitation, temperature, freezing index, and humidity), and traffic volume
	Yamany et al. (2020)	ANN	MAD, MAPE, MSE, RMSE, and R^2	LTPP	966	Climate factors (precipitation, freezing index, and temperature), pavement age, and traffic volume
	Dong et al. (2019)	LSTM-BPNN	R^2	LTPP	2,243	Pavement structure, condition, construction, material, and geographic data; climate factors (precipitation and temperature); traffic volume; previous IRIs
	Mazari and Rodriguez (2016)	GEP-ANN	RMSE and r	LTPP	98	Structural number, traffic volume, initial IRIs, and pavement age
	Gong et al. (2018b)	RFR	R^2	LTPP (SDR31)	$\geq$ 11,000	Pavement thickness, distress (fatigue, cracking, patch, ravel, block, pothole, polish, rut, IRI, etc.), traffic volume, and climate factors (freeze and precipitation)
	Zeiada et al. (2020)	ANN, DT, SVM, ensemble trees, and GPR	MSE, RMSE, MAE, and R^2	LTPP	115	Initial IRIs, relative humidity, average wind velocity, average albedo, average emissivity, traffic volume, structure condition index, sunshine, precipitation, and average temperature
	Kargah-Ostadi and Stoffels (2015)	ANN, RBF, SVM	MSE STD, AVG MSE, and R^2	LTPP (SPS-1 and SPS-5)	3,402	Previous IRIs, traffic volume, pavement age, thickness, structure information, climate factors (precipitation, freezing index, number of freeze-thaw cycles, and temperature)

Index	Reference	Algorithms	Metrics	Data Source	Number of Data	Features
PCI	Karballaeezadeh et al. (2020)	ANN (MLP & RBF)	APRE (%), AAPRE (%), RMSE and SE	Freeways in Iran	236	Pavement surface deflections
	Mahmood et al. (2020)	ANN (MLP)	MSE	LTPP	838	Pavement age, cracking area, cracking length, traffic volume, maintenance effect, climatic zones, and functional classes (collector and arterial)
	Piryonesi and El-Diraby (2020)	DT & GBT	Cross-validation accuracy, confusion matrix (precision and recall)	LTPP	943	Initial PCI, road age, material type and thickness; climate factors (precipitation, freezing index, temperature, and number of freeze-thaw cycles); functional classes, traffic volume, and M&R history
	Barua et al. (2020)	GBM	R^2 and RMSE	Chicago O'Hare International Airport	Runways: 1,091; taxiways: 2,090	Pavement age, material type, time elapsed, since last inspection, PCI from the last inspection, number of minor M&R activities, major M&R action indicator, amount of rainfall, and number of freeze-thaw cycles
SN	Janani et al. (2020)	ANN	R^2 and MSE	Tamil Nadu, India	--	Length of road section, traffic volume, subgrade properties, pavement temperature, and IRI
SN_{eff}	Sollazzo et al. (2017)	ANN	MSE and R^2	LTPP	342	Pavement thickness, traffic volume, climate factors (temperature), age, initial IRI, effective structural number, and pavement surface temperature
Rutting	Gong et al. (2018a)	ANN	R^2 and SE	NCHRP Project 01-37A	440	Predicted rutting values, material properties, structure, traffic, and climate conditions
RSL	Karballaeezadeh et al. (2019)	SVM	R, MSE, RMSE, and NSE	Private data	147	Pavement thickness and temperature of asphalt surface
PSI	Tabatabaee et al. (2013)	SVM & RNN	R^2 and RMSE	MnROAD	--	Structure, age, season, traffic, M&R indicator, treatment history, to be treated, previous-year PSI
PSL	Mousa et al. (2019)	XGBoost	confusion matrix	LaDOTD PMS database	141	Existing condition indices, pavement age, overlay thickness, and traffic volume
Cracking	Gong et al. (2019)	XGBoost	R^2 and SE	NCHRP Project 01-37A	AC: 461; LC: 414	Damage indices, pavement thickness, materials properties, climatic conditions, and traffic volume

Compared with traditional regression methods, AI algorithms provide a more accurate and efficient approach to predicting future pavement conditions. Among all AI-oriented models for pavement performance modeling, ANN-based models are the dominant method. In many studies, the performance of ANN-based models was compared with models based on other methods. Most of these studies found that ANN-based models have superior performance in terms of accuracy and generalization, which verifies ANNs' effectiveness in predicting future pavement roughness.

Similar to what has been observed in the pavement distress evaluation, most reviewed performance modeling studies encountered the problem of insufficient historical data, except for a few studies that obtained a relatively large set of data (Gong et al. 2018b, Dong et al. 2019, Kargah-Ostadi and Stoffels 2015, Majidifard et al. 2020, Maeda et al. 2018, Lei et al. 2020). The lack of adequate data is the major reason for overfitting in pavement performance modeling. Being limited by the small scale of the training and testing datasets, the prediction models cannot accurately reveal the relationships between the input parameters and the labeled target; this leads to a poor generalization of the models on new datasets. Therefore, the capability to obtain enough data plays a key role in effectively improving the generalization and accuracy of pavement performance prediction models. In practice, several studies adopted proactive strategies to address the issue of insufficient data. For instance, Hossain et al. (2019, 2020) first identified the statistical distributions of the input parameters and the IRI values based on available historical records. With these obtained distributions, a synthetic dataset was generated and applied to validate the proposed prediction models. This could be an effective method for future studies with constraints on data availability.

It is worth noting that most studies regarding pavement roughness prediction use data retrieved from the LTPP database. For one thing, pavement performance prediction is

an integral part of the goal of the LTPP program, which is partially established to gain a better understanding of the long-term impacts of various factors affecting pavement performance throughout its service life. For another, the findings from these performance prediction studies support the LTPP program to better fulfill its objectives. Therefore, it will become a long-term win-win strategy, if the LTPP data quality (e.g., accuracy, consistency, and reliability) can be fundamentally improved and the data quantity can be significantly increased.

When it comes to modeling pavement performance, parameters affected by local factors (e.g., traffic and climate) should be treated carefully. The prediction models built with data collected from one region may not be effective when applied to another region; maintenance strategies, construction materials, and climatic conditions vary from one place to another. This is also true for AI-oriented models; even though AI algorithms have been proven more accurate than traditional methods in pavement performance modeling. As observed from the reviewed studies, most of the AI-oriented prediction models use climatic factors and traffic volume as part of the input. By minimizing the variations induced by these factors, the models are more likely to achieve higher prediction accuracy. Accordingly, a customized prediction model should be developed or calibrated using local data and considering local conditions. If such data are not available, data from a region with similar factors affecting pavement performance (e.g., pavement condition, traffic capacity, climate elements, and other parameters) might be considered as an alternative.

Furthermore, as stated by (Doshi-Velez and Kim 2017), “a single metric, such as classification accuracy, is an incomplete description of most real-world tasks”. The performance of proposed models should be evaluated by various performance metrics that focus not only on prediction accuracy but also on generalization (Kargah-Ostadi and Stoffels 2015). By doing so, the prediction model can have a better chance to have

satisfactory performance when it is applied to a new set of data. In addition to accuracy and generalization, interpretability should also be deemed as an essential metric for pavement performance models. An AI model that is not capable of explaining how the solutions are derived from the inputs may easily become a black-box model. This is especially prevalent in models based on neural networks; Yamany et al. (2020) reported that ANN-based models are usually more difficult to interpret compared to other methods in pavement performance prediction. In particular, caution should be exercised for the notion that, as long as an AI model can generate accurate outputs, there is no need to concern if it is understandable or interpretable. This might be true for those low-risk systems (e.g., advertisement systems) in which the outputs of the model do not lead to any significant issue or those well-studied systems (e.g., postal code sorting) that have been thoroughly tested in real-world implementation (Carvalho et al. 2019). Except for those low-risk systems and well-studied problems, any black-box AI model that simply pursues high accuracy without giving a cursory consideration to the interpretation of the model is questionable in terms of reliability. The outcomes of such AI models fail to provide meaningful insights to solve the problem and can even lead to harmful consequences such as documented in (Varshney and Alemzadeh 2017). Particularly, in pavement management domain, such black-box models can result in misinterpretations of pavement performance and under-informed decisions which can lead to a backlog of necessary M&R interventions wasting limited resources and funding. To avoid such issues, future studies may use comprehensive metrics to evaluate the performance of AI-oriented models. Finally, examples of some general considerations that may facilitate future research are listed in Table 4.5. It is challenging for any AI-oriented model to achieve a meaningful result if such elements are not given proper consideration.

Table 4.5 Examples of General Considerations when Applying AI-oriented Methods to Pavement Management

Methods	Purpose	Examples
Algorithm Selection	Choose proper models based on the problem of interest	• Classification • Regression • Clustering • Reinforcement Learning
Data Collection and Preparation	Acquire adequate data to develop the model	• Check duplicates • Substitute/delete missing values
Feature Engineering	Improve the accuracy of the model by transforming raw data into meaningful features (Brownlee 2014)	• Handle outliers • Binning • Log transform • One-hot encoding • Scaling (e.g., normalization and standardization)
Hyperparameter Tuning	Obtain the optimal model that minimizes the loss function based on the input data	• Grid search • Random search • Bayesian optimization
Cross-validation	Improve the generalization ability of the model	• Holdout method • K-fold cross-validation • Leave-one-out cross-validation

4.4 M&R PROGRAMMING

As discussed previously, pavements deteriorate over time due to stresses induced by traffic loads, climatic factors, and aging of materials. Theoretically, to maintain the serviceability of pavements at a desired level, local transportation agencies should apply timely M&R treatments to pavements where such treatments are needed. However, most transportation agencies are experiencing difficulties in filling the growing gap between M&R needs and available funding. Given the constraint on available budgets, a primary objective for pavement preservation programs is to prioritize candidate projects and implement optimal M&R treatment interventions at the right time (Davies and Sorenson 2000, Wang et al. 2003). As revealed by (Hoang and Nguyen 2019), if appropriate

maintenance interventions can be implemented at the right time and right place, the overall M&R costs could be greatly reduced. Consequently, local agencies can make more effective decisions on allocating limited funding to maximize the overall pavement performance at the network level.

4.4.1 State of Practice in M&R Programming

In recent years, there is an increasing trend in research efforts focused on applying AI algorithms to enhance the decision-making process for pavement M&R planning and programming.

Artificial Neural Networks

As a versatile algorithm, ANN has achieved high performance in tackling complex multivariate nonlinear relationships. Leveraging this advantage, some studies have successfully demonstrated the effectiveness of ANN in developing recommendations for selecting M&R projects and treatment actions. As seen in (Janani et al. 2020), the authors proposed a M&R project prioritization method based on predictions generated from an ANN model. To shortlist road segments in the network for pavement preservation, an index for project prioritization, Maintenance Priority Index (MPI), was developed. The MPI was calculated based on traffic factors and road condition indices which were comprised of roughness and deflection index. The deflection index was predicted by an ANN model using various pavement functional parameters. Based on the predicted MPI values, the road segments can be ranked in priority order. Although the ANN model was not used directly to generate the MPI values, the outputs from the ANN model largely facilitated the calculation of MPI values.

Similarly, Elbagalati et al. (2018) proposed a pattern recognition tool based on a feedforward ANN with a back-propagation optimization algorithm to enhance the decisions on pavement M&R treatments at the network level. To construct the proposed ANN model, various parameters were obtained from a PMS database, including structural condition index (SCI), traffic level, functional class, roughness, and surface distresses. Based on the data, an ANN was developed with an 8-20-7 architecture. To overcome the issue caused by the originally imbalanced dataset, random oversampling (ROS) and random undersampling (RUS) techniques were applied. The overall pattern prediction accuracy achieved by the model was 96.9 percent and the precision was over 93.2 percent.

To identify the optimal pavement M&R treatment strategies at the network level for low-traffic roads in Colorado, a decision-making support tool using ANN-based pattern recognition algorithms was presented by Hafez et al. (2019). In this study, two individual ANN-based models were developed: one at the state-wide level and the other at the regional level. Parameters used to construct the proposed ANN-based models included M&R recommendations from local experts, multiple pavement condition indices, drivability life, and the length of road segments. The outputs of the networks were seven M&R treatment alternatives. A softmax transfer function was employed to compute the probability of each M&R treatment alternative. According to the results, the overall accuracy of the regional model was 93 percent, which was 30 percent higher than that of the state-wide model.

Likewise, Domitrović et al. (2018) applied a BPNN to select an appropriate maintenance strategy for the national road network in Croatia. The proposed BPNN took various pavement distress parameters as inputs, including IRI, rut depth, texture depth, cracks, and patches. To improve the performance, multiple activation functions were applied to the network, including linear, Gaussian, hyperbolic tangent, and logistic function. This study found that IRI was the most important factor for maintenance strategy

selection. Furthermore, a high correlation was observed between the predictions from the proposed BPNN model and the observed data, which confirmed the effectiveness of the proposed BPNN model. Overall, the reported percentage of selecting the correct maintenance strategy was 95 percent.

Reinforcement Learning

Reinforcement Learning (RL) is an individual ML paradigm that is different from supervised and unsupervised learning. For supervised learning, the primary objective is to extrapolate or generalize the knowledge learned from a training set of labeled examples so that it can respond correctly to new data. For unsupervised learning, the target is mainly to identify the structure (e.g., similarities and differences) of the unlabeled data. The goal of RL is to maximize the total cumulative reward. Instead of learning from examples of correct behavior, RL is learning from trial and error based on the feedback (i.e., rewards) obtained from the interaction between agents and the environment (Sutton and Barto 2018). Over the past few years, an increasing number of studies have applied reinforcement learning (RL) to transportation engineering (Sun and Zhang 2020, Cao et al. 2020, Li et al. 2016).

In a recent study, Yao et al. (2020) employed a deep Q-learning network (DQN) to assist in decisions on selecting pavement maintenance treatment strategies. The objective was to maximize the cost-effectiveness over a long-term pavement preservation planning horizon. The DQN models were constructed based on four components: state, action, reward, and state transition probability matrix. The action space contained a total of 38 elements which were combinations of maintenance types, maintenance materials, and distress treatments. Two DQNs with the same configurations were developed, an evaluation Q-network and a target Q-network. The ε-greedy algorithm was applied in the

evaluation Q-network to select a candidate action. The target Q-network was utilized to train the approximating function and obtain the optimal policy with the highest Q-value. Based on the results, the authors concluded that the proposed DQN could generate optimized maintenance strategies that can improve the long-term cost-effectiveness of pavement maintenance over a 15-year planning horizon.

Tree-based Algorithms

DT is an essential component of other tree-based algorithms. It is a commonly used supervised learning algorithm for classification problems. DT uses a branching method to implement decision rules, which is similar to the human decision-making process. Thus, it provides a white-box analysis approach that is easy to understand and interpret. This advantage also makes DT a suitable candidate for supporting decisions on M&R programming.

France-Mensah and O'Brien (2018) presented a study to compare the performance of DT-based methods with conventional methods in pavement M&R project prioritization and budget allocation. The comparison was among cost-benefit analysis (CBA), integer-linear programming (ILP), and a hybrid method integrating a DT model with the needs-based allocation (DTN). The proposed DTN model was based on a previous study conducted by Chi et al. (2013). The results were evaluated based on effectiveness, equity, and percentage of pavement sections in good or better condition. It was reported that the optimization method ILP outperformed the other two in terms of effectiveness and overall percentage of pavement sections in good or better condition in the network. In contrast, CBA and DTN achieved higher scores in the equity index. DTN was found to slightly outperform CBA in terms of effectiveness.

To facilitate decisions on pavement preservation plans, Han et al. (2020) presented a data mining method - the improved weight random forest (IWRF). In this study, the performance of a regular RF algorithm was enhanced by integrating with the analytic hierarchy process (AHP) and correlation analysis. This correlation analysis served as a preliminary feature selection by filtering significant parameters from noise data. The outcomes from the correlation analysis were randomly selected as inputs to train individual DT classifiers. Later, these individual trees served as sub-classifiers to form a RF classifier. To diminish potential bias from the proposed method, an AHP method was applied to adjust the weights of sub-classifiers. Finally, the overall accuracy achieved by the proposed IWRF model was more than 90 percent, which was slightly higher than that of the regular RF model.

Miscellaneous Algorithms

In addition to those individual algorithms, some researchers have performed comparison studies on the effectiveness of various AI algorithms in facilitating M&R programming. As seen in (Milad et al. 2020), a cloud-based platform, azure machine learning (AML) systems, was utilized to predict pavement maintenance treatment actions. In this study, the performance of three classification algorithms (i.e., SVM, RF, and ANN) was compared using various input parameters such as pavement distress severity, density, functional class, and traffic level. Based on these inputs, the AML models were trained and tested to generate a specific treatment action (e.g., chip seal, crack fill, cold mix patching, micro-surfacing, etc.) for a given input condition of a pavement segment. Finally, it was concluded that the ANN-based model outperformed the other models and achieved an accuracy of 99 percent.

Similarly, a recent study (Morales et al. 2020) also employed AI techniques to prioritize maintenance intervention alternatives and generate the probability of occurrence for each alternative. The authors compared the effectiveness of four supervised learning algorithms including DT, KNN, SVM, and ANN. Based on the results, it was reported that the best performance was obtained by the DT model.

4.4.2 Achievements and Limitations

This section summarizes the major findings from the latest accomplishments in applying AI techniques for pavement M&R programming, as listed in Table 4.6. The distribution of different AI algorithms presented in this section is summarized in Figure 4.3.

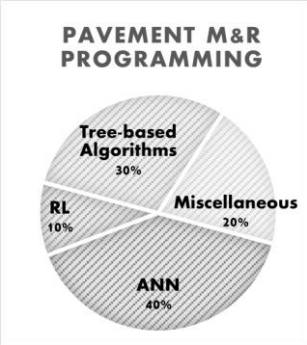


Figure 4.3 Distribution of Reviewed AI Applications in M&R Programming

Table 4.6 Summary of Major Publications Presented in M&R Programming

Reference	Algorithms	Metrics	Number of Data	Features
Janani et al. (2020)	ANN	R and MSE	--	Length of the section, traffic levels, California bearing ratio, pavement temperature, subgrade property, rutting index, and IRI
Elbagalati et al. (2018)	ANN	Confusion matrix	5,174	Surface distresses, functional class, traffic loads (ESALs), and SCI
Hafez et al. (Hafez et al. 2019)	ANN	Confusion matrix	884	Fatigue index, longitudinal index, transverse index, ride index, rut index, length of segment, and DL
Domitrović et al. (2018)	ANN (BPNN)	R^2, Max AE, Coefficient of correlation, accuracy	471	IRI, rut depth, texture depth, surface cracks, and patches
Yao et al. (2020)	DQN	MSE in Q-values	1,974	Pavement structure and material, maintenance records, traffic levels, condition indicators, temperature, road inventory, and geometries.
Chi et al. (2013)	DT	accuracy	712	Condition score (CS), CS drop, and original section category
Han et al. (2020)	RF	confusion matrix and accuracy	7,206	Rutting index, anti-slip index, distress index, and driving quality
Milad et al. (2020)	SVM, ANN, and RF	confusion matrix and accuracy	314	Distress severity, density, road function, ADT, and treatment actions
Morales et al. (2020)	DT, KNN, SVM, and ANN	learning curves and confusion matrix	1,241	Surface distress data, IRI, rut depth, percentage of cracking

Compared to pavement distress evaluation and performance prediction, fewer efforts are made in accessing M&R needs and M&R treatment programming. This is mainly because M&R programming is more complicated than distress evaluation and performance prediction. In practice, M&R programming depends not only on the outcomes from distress evaluation and performance prediction but also on subjective decisions made by experienced engineers. Specifically, the decisions on selecting preliminary M&R candidate projects are based on outcomes of various assessments and analyses, such as assessing current pavement condition, forecasting how fast the pavement deteriorates given the current state, estimating the condition at a specific time in the future, and evaluating

possible effects of the treatment based on historical records. Taking all these factors into consideration, engineers from the local agency will make the final decision on M&R programming. Hence, any inaccuracy that occurred in these prior analyses can have a significant impact on the effectiveness of accessing M&R needs and prioritizing M&R projects.

Given the complexity of the M&R programming, in a step toward obtaining the optimal M&R treatment strategy, it is worth trying to disassemble the whole research problem into several phases and apply AI algorithms to address the sub-problems as presented in (Janani et al. 2020, Morales et al. 2020).

Furthermore, pavement distress evaluation and performance prediction can be addressed by many widely applied supervised or unsupervised learning algorithms. However, pavement M&R programming is a constrained optimization problem, which is beyond the scope of conventional supervised or unsupervised learning models. The objective of M&R programming is to identify roadway sections to be maintained in the network and select the optimal M&R treatment strategy for each of the candidate sections given the presence of constraints either to minimize the cost (for budget planning problems) or to maximize the effectiveness (for budget allocation problems) (Gao et al. 2012). Therefore, to improve the effectiveness of M&R programming, more efforts should be dedicated to algorithms that are competent to solve optimization problems.

Over the past years, RL has been proven to be an efficient tool that can address such optimization problems (Xu et al. 2017, Bello et al. 2016, Zhou et al. 2019, Mao et al. 2016). A recent study (Yao et al. 2020) justified the feasibility of applying RL algorithm to assess pavement maintenance needs and select optimal M&R treatment strategies. In the future, more efforts can be made to fully explore the potential of RL in pavement M&R programming.

4.5 SUMMARY

In this chapter, a non-exhaustive search method was used to review various peer-reviewed journals on AI applications in pavement management. Accordingly, this review covered the major work published as state-of-the-art journal papers from 2015 to 2020 in applying AI algorithms to pavement distress evaluation, performance modeling, and M&R programming. Most of the reviewed studies achieved satisfactory results which justified the effectiveness of leveraging AI algorithms for pavement management.

Key conclusions drawn from this comprehensive review are presented as follows:

- Distress detection is the area that received most of the attention in pavement distress evaluation in terms of applying AI techniques and algorithms. Distress classification is also a major topic area in this sense. In the future, it can be anticipated that DNNs will become a primary method for both distress detection and classification.
- Major gaps are found in integrating the three specific tasks in distress evaluation: distress detection, distress classification, and distress quantification. Currently, for distress evaluation, most of the reviewed studies have focused on performing a single task with respect to either distress detection, classification, or quantification. In research, it can be valid and efficient to break down distress evaluation into separate tasks and only focus on a particular task; in practice, however, distress evaluation is a process that encompasses all three tasks. Through integrating distress detection, classification, and quantification, future studies can further automate the process of generating pavement condition indices by using the type, severity, and extent information obtained from AI-oriented distress evaluation and measurement.

- It is noted that not all reviewed AI-oriented pavement performance models have achieved the expected results. The performance of AI-oriented models can be affected by multiple factors including, but not limited to, data accuracy, model selection, feature engineering, hyperparameter tuning, and model validation. In future studies, these aspects should be taken into consideration to maximize the effectiveness of AI-oriented methods.
- Furthermore, the availability and quality of training data have a significant impact on the performance of AI-oriented models in pavement management; as shown in many studies, the accuracy and generalization of AI-oriented models improve dramatically with the increase of available high-quality training data. In addition to accuracy and generalization, interpretability should also be taken into consideration to avoid black-box models when applying AI algorithms to pavement management domain. Only by doing so can AI-oriented models provide meaningful insights into the problem of interest.
- This review finds a major research gap in applying AI techniques to M&R programming. Due to the complex nature of M&R decision-making process, only a limited number of AI-oriented techniques have been explored in this field. Widely applied supervised and unsupervised learning algorithms are not adequate to handle optimization problems that are typically associated with M&R programming. To improve the effectiveness of M&R programming, more efforts should be dedicated to reinforcement learning algorithms that can address optimization problems.

In summary, it can be concluded that AI algorithms have made noticeable achievements in most activity areas of pavement management. With the implementation of automated pavement survey devices, a significant amount of more consistent condition data can be collected more cost-effectively. AI techniques provide an accurate and efficient

approach to deciphering the data for various purposes within the scope of pavement management. As AI technologies continue to improve, it is inevitable that the types of algorithms that can be employed, the magnitude and accuracy of condition data that can be automatically collected and processed, and the performance of AI-oriented models that can be developed, will continue to advance. In the future, it can be expected that more efforts will be dedicated to optimizing the performance of AI-oriented methods, models, and algorithms in pavement management.

Chapter 5: An Exploration of the Impact of Pavement Conditions on Roadway Safety Using Deep Neural Networks

5.1 Problem Statement and Objectives

The preceding chapters present individual research endeavors focusing on the application of data-driven methods in two distinct domains: roadway safety and pavement management. Building upon these studies, this chapter presents an endeavor that integrates these two topics, thereby illustrating the correlation between roadway safety and pavement condition.

Previous studies have demonstrated that pavement conditions significantly impact roadway safety. Among all pavement surface conditions, pavement friction is the one that has been found to have a direct impact on roadway safety (Geedipally et al. 2019, Cafiso et al. 2007, Merritt et al. 2015). As an indicator for skid resistance, friction reflects the tire-pavement interaction that is designed to add resistance in preventing vehicles from sliding, especially during braking, navigating curves, or aggressive steering (Cafiso et al. 2021). Furthermore, friction is particularly crucial in wet weather conditions as a thin film of water on the pavement can greatly reduce tire contact and decrease pavement friction, leading to skidding or hydroplaning (Intini et al. 2019, FHWA 2016). Due to this reason, poor pavement friction is ranked as a major contributing factor to roadway departure crashes (FHWA 2016); as reported by (Donnell et al. 2019), over half of motor vehicle crash fatalities are caused by roadway departure crashes. Hence, improving pavement friction can prevent around 70% of crashes that are associated with wet pavement (FHWA 2016). Findings from other studies have also demonstrated significant reductions in wet-road crashes, run-off-the-road crashes, and total crashes as a result of improved pavement friction (Lyon and Persaud 2008, Merritt et al. 2015, Lyon et al. 2018, Lyon et al. 2020, Mayora and Piña 2009, Geedipally et al. 2019, Alhasan et al. 2018). In addition, research

conducted by Geedipally et al. (2019, 2022) found that the increase in skid resistance in terms of skid number could dramatically reduce the occurrence of both wet-weather run-off-the-road and total crashes. The significance of friction in mitigating run-off-the-road crashes is emphasized in the AASHTO Strategic Highway Safety Plan (Neuman et al. 2003), recommending the use of skid-resistant pavements as a key strategy. The findings strongly support the argument that enhancing skid resistance can yield a positive impact on overall road safety.

In addition to examining skid resistance, several studies have investigated the correlation between widely used pavement condition scores and crash frequency. A study conducted in Tennessee revealed that International Roughness Index (IRI) and Present Serviceability Index (PSI) were significant influencing factors in various types of crash prediction models (Chan et al. 2010). Furthermore, Jafari Anarkooli et al. (2021) reported a statistically significant increase in crashes with higher IRI on two-lane rural roads in Canada. More recently, a comprehensive statistical analysis further supported these findings by confirming that an increase in friction or a decrease in roughness (as measured by IRI) corresponded to a reduction in crash frequency (Cafiso et al. 2021). Li and Yu (2021) proposed a robust safety performance function (SPF) based on negative binomial regression models using highway data collected in Oklahoma. The study utilized various variables including crash exposure indicators, pavement surface characteristics, and roadway geometry features. Among these variables, pavement surface conditions such as friction, texture, IRI, and rutting were identified as statistically significant influencing factors for predicting highway crash frequency. Together, these studies emphasize the importance of considering pavement surface conditions in understanding and predicting crash frequency.

Nevertheless, while multiple studies have predominantly employed statistical methods to explore the relationship between pavement conditions and crash frequency, there has been limited focus on utilizing deep learning methods to address this problem. To fill this gap, the present study provides a demonstration of how AI-oriented methods can be effectively employed to establish the link between these two areas. To be precise, the objective of this study is to develop a robust deep neural network (DNN) model capable of predicting crashes by leveraging various input features. These features encompass pavement condition scores (e.g., ride score and distress scores), traffic volume (e.g., Average Annual Daily Traffic (AADT) and truck AADT), and roadway geometric characteristics (e.g., median width and lane width). By incorporating these diverse variables into the proposed DNN model, the study aims to enhance the understanding of the relationship between these factors and the occurrence of crashes.

Furthermore, to gain a comprehensive understanding of the impact of individual variables on crash frequency, a parametric analysis was performed using the proposed DNN model. This analysis focused on the top five features that displayed the highest influence on crash predictions, namely: AADT, Truck AADT, median width, directional distribution factor, and ride score. During the parametric analysis, each input feature was systematically modified while keeping the remaining input features constant (e.g., the mean of the original data). Ultimately, this analysis contributes to improved interpretability of the results through a detailed investigation into the isolated effects of each variable on crash frequency, enabling a more accurate assessment of their respective contributions.

5.2 METHODOLOGIES

5.2.1 DNN

Neural Networks (NNs) are learning machines that function in a similar way to human brains. In practice, NNs have proved to be extremely effective in a wide range of research fields such as speech recognition, pattern recognition, disease diagnosis, optimization method, natural language processing, and computer vision (Abiodun et al. 2018).

Multilayer perceptron (MLP) is one of the most used NNs (Ramchoun et al. 2016, Popescu et al. 2009, Noriega 2005). A MLP is a network of neurons that can learn the mapping between the input features and the output target using the training data. The basic computational units of MLPs are neurons that can process information by summing up weighted input signals and generating an output signal based on an activation function (or transfer function). The weight of an input represents the strength of the connection between the input and the neuron. The weighted inputs are aggregated with a bias and then sent to an activation function that controls the threshold of determining whether a neuron should be activated or not. Moreover, the activation function introduces non-linearity into the output and governs the strength of the output signal. Some examples of commonly used activation functions in NNs include binary step, hyperbolic tangent (Tanh), logistic (Sigmoid), and rectified linear unit (ReLU) (Szandała 2021, Sharma et al. 2017). Among them, ReLU has been proven to be a highly efficient non-linear activation function; in fact, a study reported that ReLU was the most utilized activation function in DNN (Szandała 2021, Ping et al. 2017).

MLPs typically have a feedforward architecture in which the signals are transmitted from the inputs to the output (Popescu et al. 2009). As shown in Figure 5.1, the very first

layer that directly receives input data (i.e., features) from external sources is the input layer. The last layer that presents the computational result is the output layer. All the other layers that are constructed after the input layer and before the output layer are referred to as hidden layers. Neurons on a hidden layer take inputs from neurons on the preceding layer and propagate the processed outputs to neurons on the following layer. The network is an integration of all three types of layers. In particular, when the number of hidden layers is greater than one, the NN is referred to as a deep neural network (DNN). The multiple hidden layers enable DNNs to become deep architecture models which are capable of learning more intricate structures in high-dimensional data (Szegedy et al. 2013, Goodfellow et al. 2016).

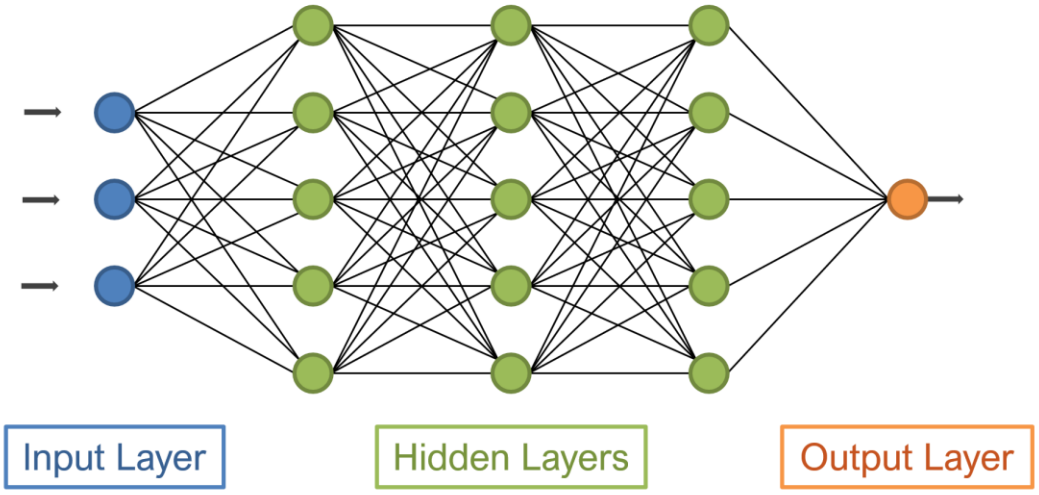


Figure 5.1 An illustration of a Deep Neural Network with Three Hidden Layers

The interconnected parallel structure enables DNNs to deal with nonlinear problems effectively and enhances the capability of generalization which represents a model's ability to generate appropriate results for new data. It has been ranked as one of the most important metrics to evaluate the effectiveness of a ML algorithm (de Mello and

Moacir 2018). An algorithm with a robust capability of generalization can function well not only on the training data but also on data that have never been encountered before. Therefore, generalization often reflects how "intelligent" an algorithm can be.

5.2.2 Random Forest

Random forests (RFs) are ensemble algorithms that are developed to overcome overfitting caused by a single decision tree. A RF grows a mass of decision trees in which a common setting for the number of trees is 500 (Belgiu and Drăguţ 2016). These decision trees are combined to generate a more accurate prediction. For a classification problem, each decision tree in a RF outputs a class value, and the class that has the most votes will be the final decision of the forest. For regression modeling, the final prediction of a RF is an average of the values from all trees in the forest. RFs have been widely applied to solve problems including regression, classification, and clustering; some computer scientists remarked that RFs are the most versatile ML algorithms (Howard and Bowles 2012). In addition, RF is capable of handling missing data and can be applied to various types of data. It can provide high-accuracy prediction and reliable generalization using ensemble algorithms and random sampling methods (Ali et al. 2012).

5.2.3 Performance Metrics

To access the goodness-of-fit of the AI/ML algorithms, this study employs three commonly used statistical measures for regression problems, which include the coefficient of determination (R-squared or R^2) and root mean square error (RMSE). Table 5.1 lists the definitions and equations for each metric. According to a recent study (Chicco et al. 2021), R-squared is superior to other statistical metrics used in regression analysis because it is more informative and easier to interpret with a definite range [0, 1]. In contrast, other

performance metrics such as MSE, RMSE, and mean absolute error (MAE) have a natural limitation of interpretability due to the lack of an upper boundary (Chicco et al. 2021).

Table 5.1 Commonly Used Performance Evaluation Metrics for Regression Models

Metrics	Term	Description	Equation
R-squared (R^2)	Coefficient of Determination	a measure to evaluate how close the data are to the fitted regression line	$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y})}$
$RMSE$	Root Mean Square Error	the standard deviation of the prediction errors	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

Note: y_i = actual observations, $\hat{y}_i$ = model predictions, and n = number of data points.

5.2.4 k-Fold Cross-Validation

Cross-validation is a data resampling approach that has been widely applied in machine learning methods to assess the generalization ability of predictive models (Berrar 2019). It aims to provide an estimation of how the model is expected to perform in general when making predictions on data that were not part of the training process. This method is widely applied because of its simplicity and its effectiveness to yield less biased results. By using cross-validation, it helps to obtain a more reliable evaluation of the model, thereby enabling better generalization to unseen data.

Cross-validation typically employs a parameter – denoted as "k" – to dictate the number of groups into which a given dataset is to be partitioned. Consequently, this approach is commonly referred to as "k-fold cross-validation" where "fold" refers to the number of resulting subsets that have approximately equal sizes (Berrar 2019). In the k-fold cross-validation process, the initial fold is utilized as a validation set, while the

remaining $k-1$ folds are used as the training set for the predictive model. The performance is measured based on the data points within the held-out fold (i.e., validation set). Then, the same procedure is repeated k times, with each iteration selecting a distinct group as the validation set. Correspondingly, a total of k estimates of the test error is obtained. By taking an average of these test errors, the k-fold cross-validation estimate can be computed (Gareth et al. 2013). Although there is no formal rule, in practice, k=10 is recommended as empirical evidence has shown that this value usually provides test error rate estimates that effectively avoid both high bias and excessive variance (Gareth et al. 2013, Berrar 2019, Kohavi 1995).

5.3 DATA

5.3.1 Data Sources and Study Scope

The scope of the study is confined to the interstate highway (IH) routes within the state of Texas, encompassing a total length of 2,819 miles (Figure 5.2). The analysis involves a period of three years, specifically from 2016 to 2018. It is assumed that there were no substantial variations in roadway characteristics and the surrounding environment (e.g., traffic volume, weather conditions, traffic control, land use, and geometric design) throughout the designated study period. To perform the proposed study, three data sources – all developed and maintained by the Texas Department of Transportation (TxDOT) – were identified and utilized including the Pavement Management Information System (PMIS), the Crash Records Information System (CRIS), and the TxDOT Roadway Inventory (2018).

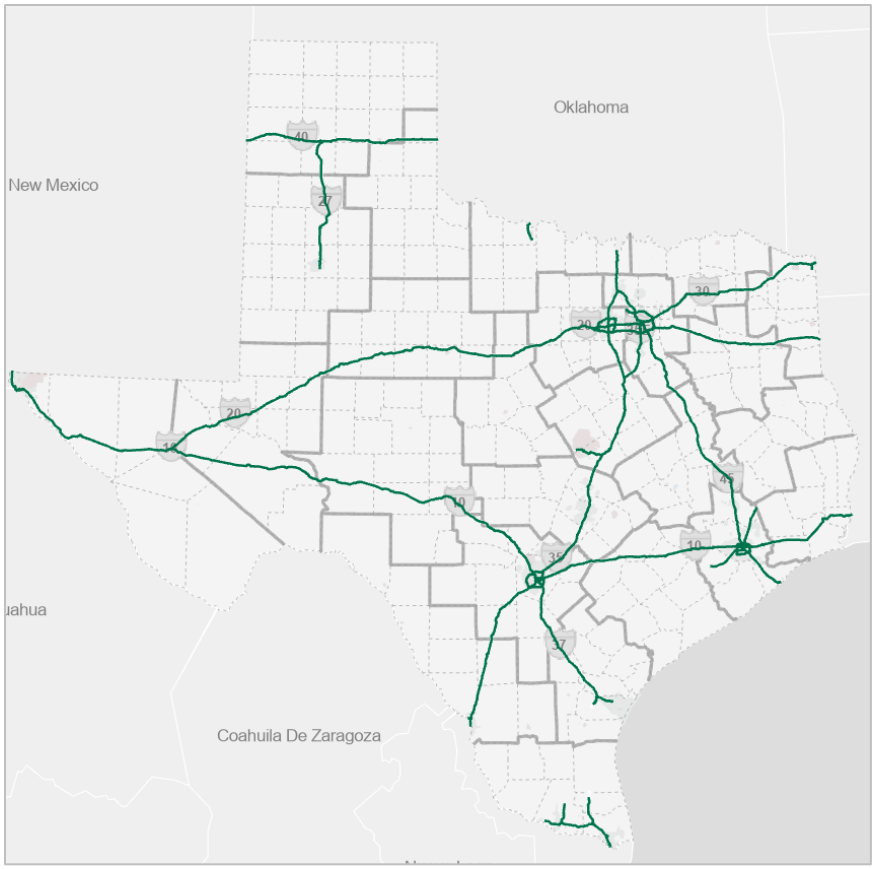


Figure 5.2 An Illustration of IH Routes in the Analysis

5.3.2 Variables of Interest

The primary objective of this study is to utilize a DNN-based approach to predict the number of crashes (i.e., crash frequency) that occurred in each road segment during the analysis period. This approach incorporates a comprehensive set of variables, including pavement condition indicators, traffic volume data, and roadway geometric characteristics. To accomplish this objective, pertinent variables from the PMIS dataset and the Roadway Inventory were reviewed and selected to serve as input features for supporting the development of the proposed predictive model. The details of these attributes are presented in Table 5.2.

Table 5.2 Variables Relevant to Crash Prediction

Type	Variables	Description	Mean (Std)	Min	Max	Source
Pavement Condition Data	DS_16	Distress score in 2016	93.27 (14.83)	6.00	100.00	PMIS
	DS_17	Distress score in 2017	92.49 (14.98)	1.00	100.00	PMIS
	DS_18	Distress score in 2018	92.50 (13.93)	1.00	100.00	PMIS
	CS_16	Condition Score in 2016	92.20 (15.94)	6.00	100.00	PMIS
	CS_17	Condition Score in 2017	91.50 (15.92)	1.00	100.00	PMIS
	CS_18	Condition Score in 2018	91.82 (14.85)	1.00	100.00	PMIS
	RS_16	Ride score in 2016	4.11 (0.54)	1.80	5.00	PMIS
	RS_17	Ride score in 2017	4.17 (0.59)	1.87	5.00	PMIS
	RS_18	Ride score in 2018	4.26 (0.55)	0.99	5.00	PMIS
	avgIRI_16	Average IRI score in 2016 [inches/mile]	80.25 (29.25)	22.00	240.00	PMIS
	avgIRI_17	Average IRI score in 2017 [inches/mile]	77.84 (30.79)	18.00	231.00	PMIS
	avgIRI_18	Average IRI score in 2018 [inches/mile]	72.80 (28.04)	11.00	223.00	PMIS
Traffic Data	AADT_16	AADT in 2016	24721.82 (25347.17)	1,131.00	143,953.00	PMIS
	AADT_17	AADT in 2017	26887.95 (27380.86)	2,508.00	163,339.00	PMIS
	AADT_18	AADT in 2018	27447.40 (27518.35)	1,994.00	160,348.00	PMIS
	avgAADT	3-year average AADT	26352.39 (26661.29)	2,648.00	155,880.00	PMIS
	ADT_CUR	AADT for both directions in 2018 (RI)	55865.87 (53833.79)	4,689.00	314,266.00	RI
	ADT_HY_1	AADT for both directions in 2017 (RI)	54480.77 (52947.71)	5,478.00	314,266.00	RI
	ADT_HY_2	AADT for both directions in 2016 (RI)	54077.48 (54109.17)	3,987.00	320,695.00	RI
	TRK_AADT_PCT	Percentage of trucks in AADT [%]	28.11 (14.48)	1.50	95.80	RI
	AADT_TRUCKS	Number of trucks in AADT	10106.16 (5082.64)	376.00	37071.00	RI
	SPD_MAX	Maximum speed limit [mph]	71.88 (6.54)	30.00	80.00	RI
	K_FAC	Peak factor [%]	9.75 (1.35)	5.30	19.80	RI
	D_FAC	Directional distribution factor [%]	55.98 (4.34)	50.00	79.00	RI
Road Characteristics	lenOfSeg	Length of road segment	0.50 (0.03)	0.12	0.80	PMIS
	MED_WID	Width of median excludes inside shoulders [feet]	43.49 (31.14)	1.00	455.00	RI
	HP_MED_W	Width of median includes both inside shoulders [feet]	49.80 (30.64)	3.00	459.00	RI
	RB_WID	Width of roadbed includes shoulder-width and surface-widths [feet]	44.81 (11.24)	28.00	120.00	RI
	SUR_W	Width of surface excludes shoulder-widths [feet]	29.03 (9.44)	22.00	106.00	RI
	NUM_LANES	Number of through lanes	2.42 (0.76)	1.00	7.00	RI
	LANE_WIDTH	Width of lane [feet]	11.99 (0.72)	8.00	24.00	RI
	R1	Binary indicator for Rural Interstate	Yes: 59.71%; No: 40.29%	0 = No	1 = Yes	RI
Crash Data (Target)	Total Crashes	3-year crash frequency	16.25 (30.05)	0.00	291.00	CRIS

Note: RI = TxDOT Rodeway Inventory Database

Various pavement condition variables were extracted from the PMIS database. Among these, the Distress Score (DS) quantifies the extent of surface distress observed in the data collection section. The Ride Score (RS) provides an assessment of the ride quality experienced by users in the section. Similarly, the IRI is a standardized measure to quantify the ride quality of a pavement segment. The Condition Score (CS) is a comprehensive indicator that evaluates the overall quality of the pavement, encompassing both the visual observation of distress and the tactile experience of ride comfort.

Regarding the traffic volume pertinent variables listed in Table 5.2, it is worth noting that the AADT stored in the PMIS dataset only counts for a single direction of the road while traffic data from the Roadway Inventory dataset reflect the traffic volume in both directions. Accordingly, the directional distribution factor specifies how the traffic is distributed.

5.3.3 Data Preparation

In this study, the TxDOT Roadway Inventory geodatabase was utilized as the master map. Pavement condition data based on left main lanes in IH routes, which was originally retrieved from PMIS 2016-2018 datasets, were mapped on the master map. The process of data preparation started with examining data completeness of pavement condition variables and removing missing data. After these operations, the total number of road segments for left main lanes in IH routes in PMIS datasets were 6,860 (2018), 6,774 (2017), and 6,735 (2016). These segments were then spatially integrated using the linear referencing method predefined by the Roadway Inventory geodatabase in ArcGIS Pro, resulting in 6,198 road segments. Later, for each road segment generated from the previous step, crashes from CRIS 2016-2018 were aggregated, in the form of total crash frequency, using Python programming. Finally, the integrated PMIS dataset was spatially joined with

useful variables from the Roadway Inventory dataset in ArcGIS Pro, thus generating a single dataset that encompasses all the selected variables presented in the previous section. This comprehensive dataset contains 5,632 data points (i.e., road segments) that support the development of the DNN crash prediction models, as illustrated in Figure 5.3.

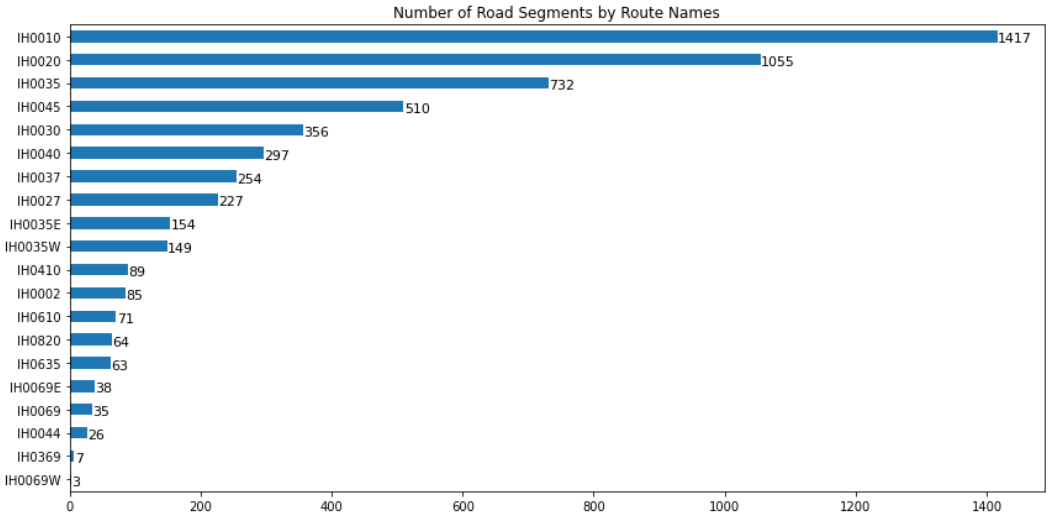


Figure 5.3 Number of Road Segments by IH Routes

5.3.4 Feature Selection

The aforementioned dataset contained a variety of variables that have the potential to serve as inputs for the proposed crash prediction model. Upon close examination, however, it can be noticed that some of these variables were highly correlated with each other. For example, the three-year average AADT (i.e., avgAADT) was computed directly from a simple linear combination of AADTs in individual years; hence, a high correlation is expected to present between the average AADT and individual AADTs. Similarly, pavement condition scores are expected to be highly correlated with pavement distress scores in the same year; because condition scores are a composite indicator computed from distress scores and ride scores (TxDOT 2021d).

The strong correlation between independent variables implies multicollinearity (Alin 2010, Daoud 2017). Multicollinearity not only leads to unreliable coefficients but also increases the difficulty to assess the marginal effect of independent variables (Alin 2010). Pearson correlation coefficient is a statistical measure that can be employed to diagnose multicollinearity in data. It quantifies the strength (ranging from -1 to 1) of the linear relationship between two features. To identify the most appropriate input features, the correlation matrix that contains pairwise Pearson correlation coefficients and feature importance were obtained, as shown in Figure 5.4 and Figure 5.5.

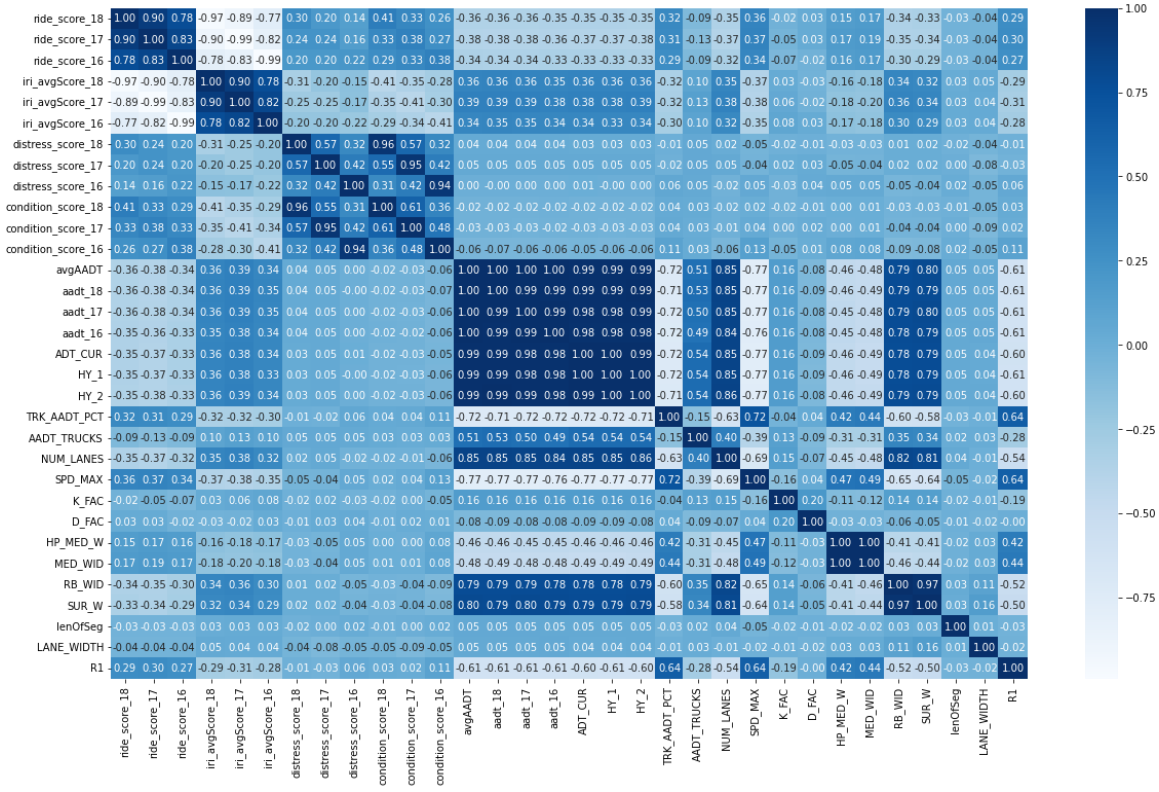


Figure 5.4 Feature Correlation Matrix

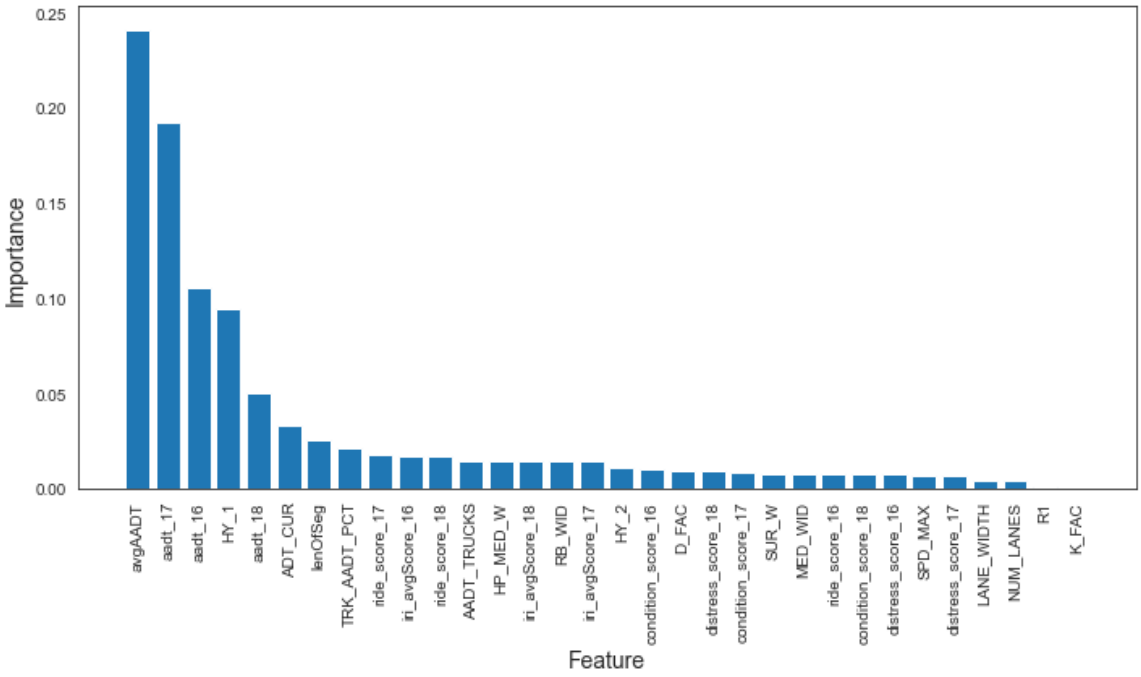


Figure 5.5 Feature Importance derived from Random Forest

To remove multicollinearity, a cutoff threshold for the correlation coefficient was set as 0.7. In other words, it is assumed that an absolute correlation coefficient greater than 0.7 indicates the presence of multicollinearity. Further, feature importance based on a tree-based algorithm (i.e., RF) was obtained and used for supporting decisions on final features. Feature importance is an approach that is mostly used for assessing the effectiveness of input features in terms of their predictive utility in relation to the target variable.

In this step, the entire dataset was used to train a random forest regressor (RFR), and its performance was evaluated using cross-validation. The analysis revealed that the 3-year average AADT (avgAADT) emerged as the most important feature, as depicted in Figure 5.5. This outcome was anticipated, given the strong correlation between crash frequency and traffic volume. Consequently, other features that present significant correlations with avgAADT were removed from the data, including all the traffic volume

variables, RB_WID, SUR_W, SPD_MAX, and NUM_LANES. Similarly, pavement condition scores and IRI average scores were removed from the data; the former was highly correlated with pavement distress scores in the same year while the latter was negatively correlated with ride scores. Regarding ride scores, as shown in Figure 5.4, three ride scores were significantly correlated with each other. The feature importance indicated that the score in 2017 was relatively more important than that of the other two years. Hence, ride scores in 2016 and 2018 were removed. Further, MED_WID was removed due to its high correlation with HP_MED_W which was found to be more important in the evaluation. K_FAC was removed because it did not contribute to the predictive model.

In summary, a total of 11 features were selected based on the correlation matrix and feature importance, including pavement condition indicators (RS_17, DS_18, DS_17, and DS_16), traffic data (avgAADT, AADT_TRUCKS, and D_FAC), and roadway characteristics (HP_MED_W, lenOfSeg, LANE_WIDTH, and R1). Table 5.3 presents the statistical summary of these features which were used as inputs for developing the proposed predictive model.

Table 5.3 Statistical Description of Selected Features

Features	Mean	Std.	Min.	25%	50%	75%	Max.
avgAADT	26352	26661	2648	8373	15832	31566	155880
AADT_TRUCKS	10106	5083	376	6015	10107	12849	37071
RS_17	4.177	0.591	1.869	3.798	4.262	4.656	5.000
DS_18	93	14	1	90	100	100	100
DS_17	92	15	1	92	100	100	100
DS_16	93	15	6	94	100	100	100
HP_MED_W	49.80	30.64	3.00	36.00	46.00	64.00	459.00
lenOfSeg	0.50	0.03	0.12	0.50	0.50	0.50	0.80
LANE_WIDTH	11.99	0.72	8.00	12.00	12.00	12.00	24.00
D_FAC (%)	55.98	4.34	50.00	53.00	55.00	60.00	79.00
R1	0.60	0.49	0	0	1	1	1

5.3.5 Data Transformation

When preparing the data, it was found that the distribution and scale of features vary dramatically from one to another. For instance, the range of avgAADT is from 2,648 to 155,880, while the scale of RS_17 is from 1.869 to 5.000. The huge discrepancies in the scales observed among input variables can introduce complexities in the development of DNN models. When dealing with substantial input values (e.g., avgAADT), the resulting model may acquire disproportionately large weight values. Consequently, the stability of the model is compromised, leading to inadequate performance in the learning phase and increased susceptibility to variations in input values. This can significantly impair the generalization of DNN models (Brownlee 2019). Therefore, it is necessary to implement appropriate data transformation techniques before using the data to train a learning algorithm.

Normalization and standardization are widely employed techniques for transforming data. Standardization is commonly used when a variable follows a normal distribution, as it adjusts the data to conform to the properties of the standard normal distribution by scaling them to have a mean of 0 and a standard deviation of 1. On the other hand, normalization is a method that maps data to a specific scale, typically ranging from 0 to 1 (or from -1 to 1). In this study, data normalization was utilized on selected features due to the absence of normal distribution in any of the examined features. In practice, the normalization of input features can not only improve model performance but also significantly reduce the computational time (Sola and Sevilla 1997). For the same reason, the target was also normalized.

5.4 RESULTS AND DISCUSSION

In this study, a portion of the IH-10 route was used as the testing dataset to evaluate the model performance. Specifically, 15% of IH-10 road segments were randomly selected and merged with other IH routes to serve as the training dataset. While the remaining 85% of IH-10 data (around 1,204 data points) was utilized as the testing dataset. Both the input features and the output of the network were normalized using the *MinMaxScaler* provided by the *scikit-learn* Python library. The distributions of features and targets in the training dataset before and after applying *MinMaxScaler* are shown in Figure 5.6 and Figure 5.7.

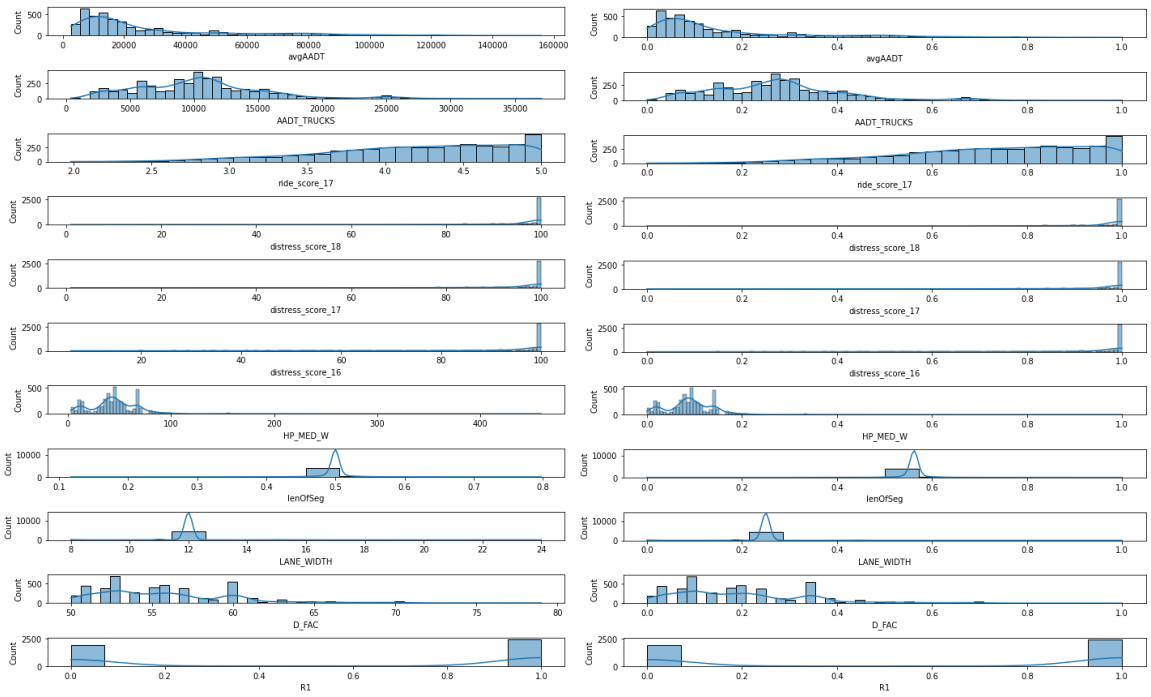


Figure 5.6 Distributions of Features in Training Dataset before and after Data Normalization

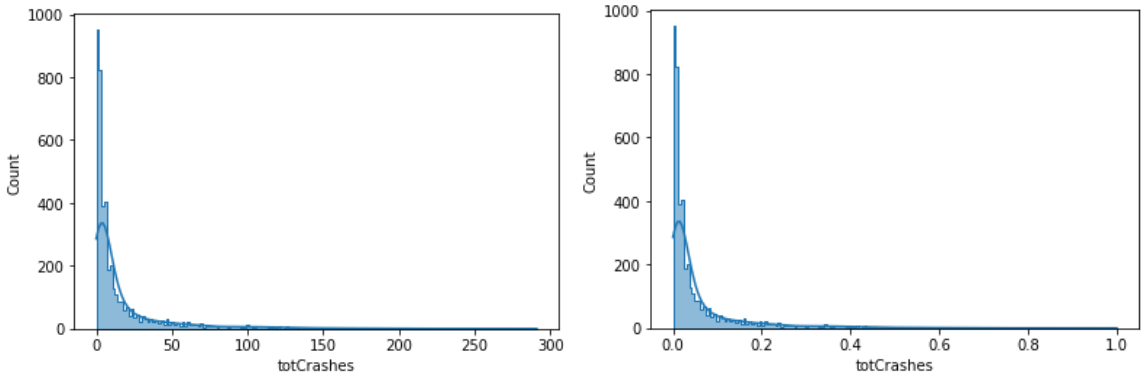


Figure 5.7 Distribution of Target in Training Dataset before and after Data Normalization

5.4.2 Network Architecture

DNNs are characterized by their complex architectures and numerous hyperparameters that need to be configured before training. Hyperparameters are a set of external configuration variables that are used to control the learning process in AI/ML algorithms. These hyperparameters, such as learning rate, batch size, and regularization methods, have a significant impact on various aspects of the model, including its performance, convergence speed, and generalization ability. Hyperparameter optimization (also known as hyperparameter tuning) is a data-driven approach that systematically searches and evaluates different hyperparameter configurations to identify the optimal combination that maximizes the performance of a learning model on a specific task. This process is pivotal in model development as it enables the discovery of hyperparameter settings that yield superior results (e.g., maximize accuracy or minimize error), ultimately leading to more accurate and efficient learning models.

This study employed a hybrid hyperparameter optimization approach that combined grid search with random search to identify the optimal configurations of hyperparameters in the DNN model. Random search involves defining a bounded domain

of hyperparameter values as the search space and randomly selecting values within that domain. On the other hand, grid search is a brute-force searching approach that creates a grid of hyperparameter values and evaluates each candidate on the grid one by one. To implement these techniques, the *scikit-learn* Python library provides *RandomizedSearchCV* for random search and *GridSearchCV* for grid search. Both approaches employ cross-validation to evaluate models for a given hyperparameter vector, which explains the inclusion of the *"CV"* suffix in their names.

To implement the hybrid hyperparameter optimization approach, a Sequential neural network (i.e., MLP) was constructed using the Python *Keras* libraries, leveraging Google's *TensorFlow* module. Then, the *RandomizedSearchCV* algorithm was utilized to estimate the optimal configurations for the Sequential model. Random search was chosen for this initial step of hyperparameter optimization due to its superior performance in identifying optimal models compared to grid search (Liashchynskyi and Liashchynskyi 2019, Bergstra and Bengio 2012). Additionally, random search is known for its computational efficiency, which means it requires less computation time than grid search (Bergstra and Bengio 2012). The hyperparameter tuned by RandomizedSearchCV included number of hidden layers, number of nodes in each layer, learning rate, dropout rate, optimizer, batch size, and training epochs. The search space for each hyperparameter in the *RandomizedSearchCV* is detailed in Table 5.4.

Table 5.4 Hyperparameter Optimization using RandomizedSearchCV

Hyperparameter	**Range**	**Optimal Value**
Number of hidden layers	2, 3, 4	3
Number of nodes/neurons	[16, 512]	311-103-362
Optimizer	Adam, SGD, RMSprop	Adam
Learning rate	[0.0001, 0.1]	0.018
Dropout rate	[0.0, 0.4]	0.2
Batch size	[16, 128]	54
Training epochs	[50, 200]	139

The best hyperparameter values identified by *RandomizedSearchCV* were listed in the last column in Table 5.4. It was found that the best DNN model was a network comprised of three hidden layers with 311, 103, and 362 nodes in each layer, respectively. The *ReLU* activation functions were applied to all three hidden layers. A linear activation function was used for the output layer since the study aims to solve a regression problem in which the target (i.e., outputs) is the number of crashes (i.e., crash frequency).

Subsequently, using the results obtained from *RandomizedSearchCV* as a starting point, *GridSearchCV* was employed to further refine the model. In particular, the optimal architecture of the network (e.g., 311-103-362) as well as the best optimizer identified by *RandomizedSearchCV* was used as fixed inputs in the subsequent search powered by GridSearchCV. The remaining hyperparameters were further fine-tuned via the traditional optimization method (i.e., grid search), including learning rate, dropout rate, batch size, and training epochs. Table 5.5 lists the search space for these hyperparameters.

Table 5.5 Hyperparameter Optimization using GridSearchCV

Hyperparameter	**Range**	**Optimal Value**
Learning rate	0.0001, 0.001, 0.018, 0.1	0.001
Dropout rate	0.05, 0.1, 0.2	0.1
Batch size	50, 64, 100	64
Training epochs	100, 120, 140, 160	120

In this study, the identification of the best DNN model comprised two phases. The initial phase aimed to examine the effectiveness of selected features. This step was important because the features were selected based on tree-based algorithms, there was no guarantee that all of them would be useful in support of DNN models. After evaluating the feature importance from the identified DNN models, two features (i.e., DS_18 and lenOfSeg) were removed since they were found to be the least contributors to the predictive model. While the length of road segments is considered a significant crash exposure

indicator in the Highway Safety Manual (HSM) developed by AASHTO (2010), it was not found to be a significant contributing factor in the proposed DNN model. This discrepancy is most likely attributed to the consistent and relatively equal length of segments in PMIS (typically around 0.5 miles) which minimizes variations in segment length across the dataset.

Upon retraining, no significant changes in the performance of the DNN model had been observed. In contrast, after the removal of these features, the training performance had been found to be more stable and robust. This observation is consistent with what has been reported in other studies (Hooker et al. 2019). In the second phase, the remaining features were used to train the hybrid search algorithm. Based on the customized testing data, the performance metrics achieved by the best DNN model were RMSE=14.645 and R^2 =0.680.

5.4.3 Model Comparison

In order to gain a better understanding of the proposed DNN model, the study also created a baseline model using Random Forest Regressor (RFR) as a comparison method. The RFR was chosen because this method outperformed other algorithms in terms of RMSE and R^2 score based on the performance evaluation using k-fold cross-validation. To select a model for the comparative study, various ML algorithms were trained on the entire dataset, which included RFR, XGBoost, GradientBoosting, Ridge regression, HuberRegressor, and linear regression. Table 5.6 listed the performance of each model based on k-fold cross-validation. It is worth noting that an equitable ground for comparison was established by utilizing the same set of input features for both the DNN model and the aforementioned models. This strategic decision ensures a fair and unbiased evaluation of their respective performances.

Table 5.6 Performance of Candidate Models for the Comparative Study

Algorithms	**R^2 (+std)**	**RMSE (+std)**
LinearRegression	0.620 (0.037)	18.433 (1.537)
Ridge	0.620 (0.037)	18.434 (1.541)
HuberRegressor	0.581 (0.031)	19.380 (1.733)
RandomForestRegressor (RFR)	**0.665 (0.043)**	**17.292 (1.493)**
GradientBoostingRegressor	0.663 (0.040)	17.368 (1.732)
XGBRegressor	**0.665 (0.039)**	17.301 (1.715)

This study employed *Optuna* to identify the best hyperparameter configurations for RFR. *Optuna* is an automatic hyperparameter optimization framework that can be integrated with various machine learning packages to perform a variety of tasks (Akiba et al. 2019). This integration enables the utilization of pruning algorithms to enhance the efficiency of hyperparameter search processes (Akiba et al. 2019). *Optuna* is featured by its "*define-by-run*" characteristic that allows users to dynamically define a search space for optimization (Akiba et al. 2019).

In general, the implementation of *Optuna* consisted of five key steps:

- Defining the hyperparameters to be tuned.
- Specifying the search space for each hyperparameter.
- Defining the objective function.
- Specifying the scoring function, whether it is maximized or minimized.
- Determining the number of trials to be conducted.

The RFR hyperparameters and their corresponding search spaces optimized via *Optuna* are listed in Table 5.7. The objective of minimizing RMSE was pursued in this study, with a total of 100 trials conducted. Based on this setting, the optimal hyperparameter values identified by *Optuna* are presented in the last column in Table 5.7.

Table 5.7 Search Space for Hyperparameters Tuned by Optuna

Hyperparameter	Description	Range	Optimal Value
n_estimators	Number of trees in the forest	[100, 1000]	398
max_depth	Max. depth of tree	[3, 10]	10
min_samples_split	Min. number of samples required to split a node	[2, 10]	10
min_samples_leaf	Min. number of samples required to be at a leaf node	[1, 10]	3
max_features	Number of features in searching for best split	[0.1, 1.0]	0.56

The evaluation of performance on the testing dataset revealed that the best RFR model attained an RMSE of 15.185 and an R^2 value of 0.656. Figure 5.8 illustrates the comparison of performance metrics between the RFR model and the proposed DNN model. The results demonstrate that the proposed DNN model surpasses the RFR model, achieving higher RMSE and R^2 scores on the testing dataset, indicating superior performance in terms of generalization and accuracy.

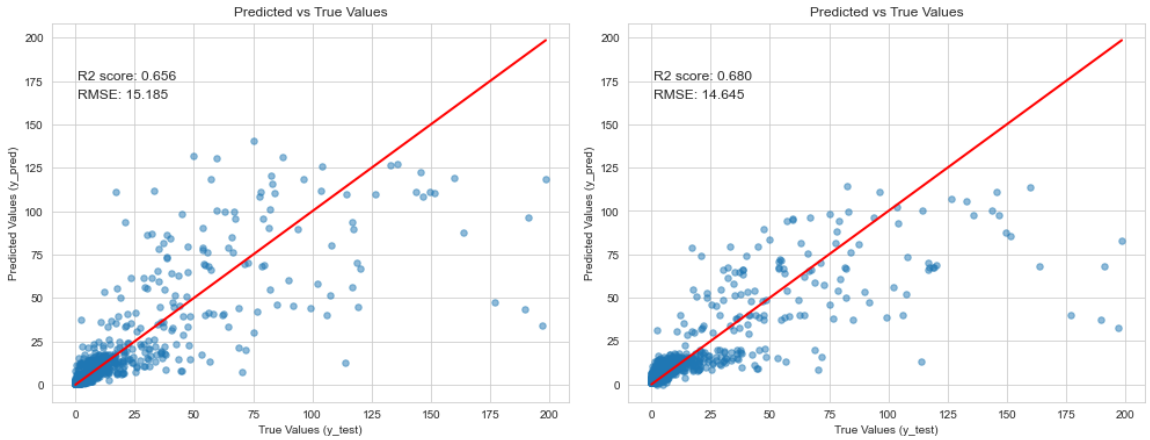


Figure 5.8 Performance Evaluation of the RFR and the Proposed DNN

5.4.4 Parametric Analysis

In addition to accuracy and generalization, interpretability plays a crucial role in assessing the effectiveness of AI/ML algorithms. To gain a deeper understanding of the influence of individual features on crash frequency prediction, a parametric analysis was conducted in this study. The top five ranked features in the proposed DNN model were

selected for this analysis, including mean AADT, truck AADT, directional distribution factor, median width, and ride score. Figure 5.9 illustrates the feature importance obtained from the DNN model, providing insights into the relative contributions of each feature.

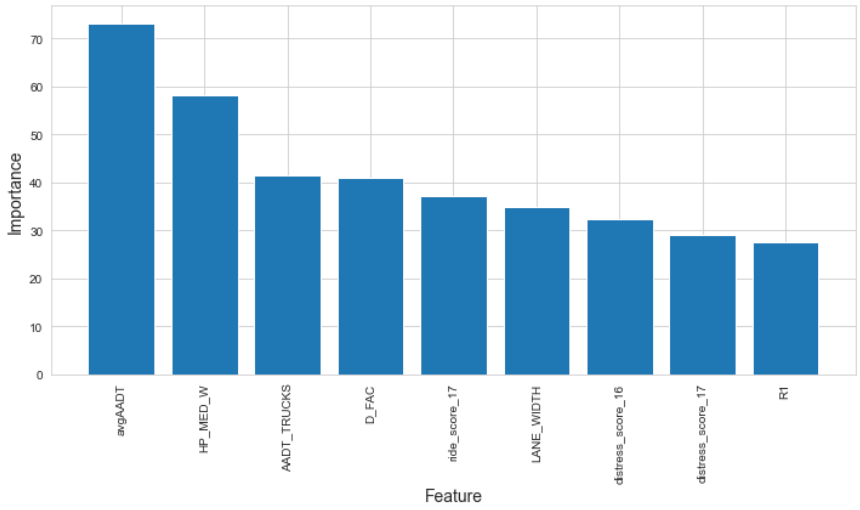


Figure 5.9 Feature Importance in the Best DNN Model

Considering the highly skewed distributions and heavy tails observed in all five features of interest (Figure 5.6), a range was carefully selected for each feature, taking into account their respective distributions. This range selection aimed to enhance the effectiveness of the model predictions. The specific details of the feature ranges can be found in Table 5.8.

Table 5.8 Feature Range Configuration for Parametric Analysis

Feature	**Start Value**	**End Value**	**Lower Threshold**	**Upper Threshold**
AADT	2,648	100,000	min.	103,310 (p98)
Truck AADT	2,500	24,000	2,490 (p2)	24,696 (p98)
Ride Score	2.8	5	2.79 (p2)	max.
Median Width	3	115	min.	115 (p98)
D_FAC	50	68	min.	68 (p98)

Note: p98 = 98th percentile, which means 98% of all data are less than p98; p2 = 2nd percentile, which means 2% of all data are less than p2.

The parametric analysis involved a systematic examination of each feature individually, focusing on the five features of interest listed in Table 5.8. During the analysis, the feature of interest varied gradually from the start value to the end value, with 50 intervals in between. Throughout this process, the remaining features were held constant utilizing the mean obtained from the original dataset, as presented in Table 5.9.

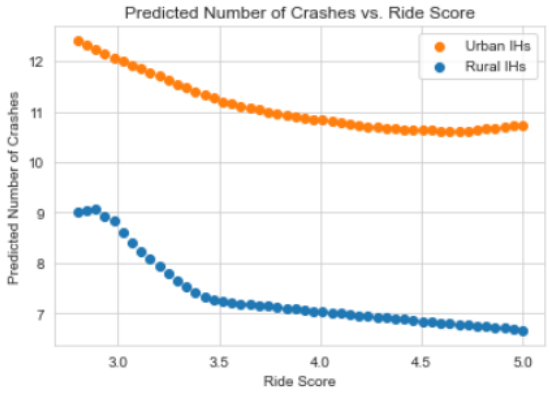
Table 5.9 Configuration for Remaining Features in Parametric Analysis

Features	**Mean**
Mean AADT	26,352
Truck AADT	10,106
Ride Score	4.17
Median Width	50
D_FAC	56
Distress Score 17	92
Distress Score 16	93
Lane Width	12

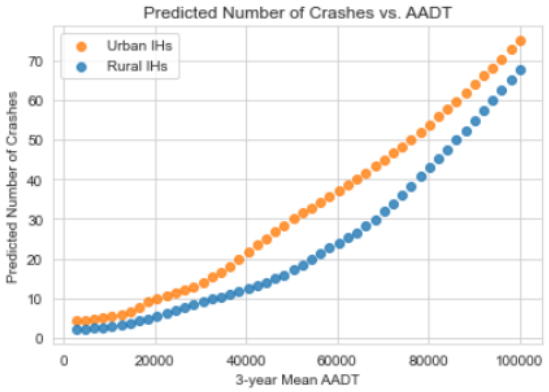
As a nonlinear method, DNN algorithms learn through a stochastic training process. This characteristic makes the output of the model heavily rely on the specific training data, initial conditions (e.g., initial weights), and other random factors that occurred in training. Consequently, even with the same model configuration and training dataset, the model generates distinct predictions. In order to address this issue, the study implemented a model averaging approach, which involved the following steps:

- Training multiple DNN models,
- Identifying the three models with the highest R^2 scores, and
- Averaging the predictions of these selected models.

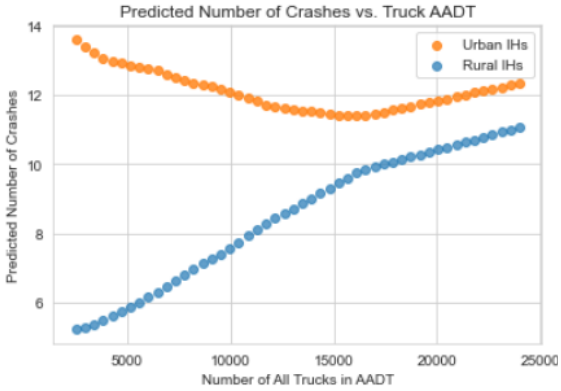
The results of the parametric analysis for all five features are presented in Figure 5.10 and discussed in the subsequent sections.



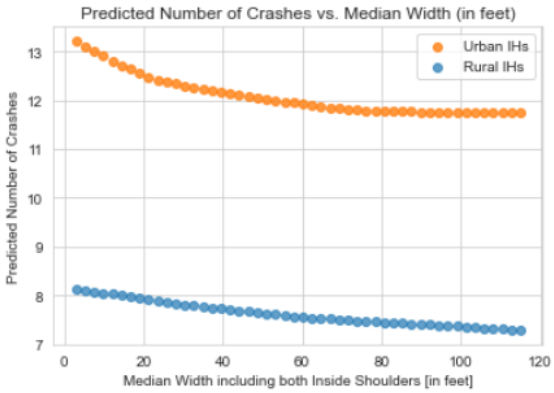
(a) Ride Score



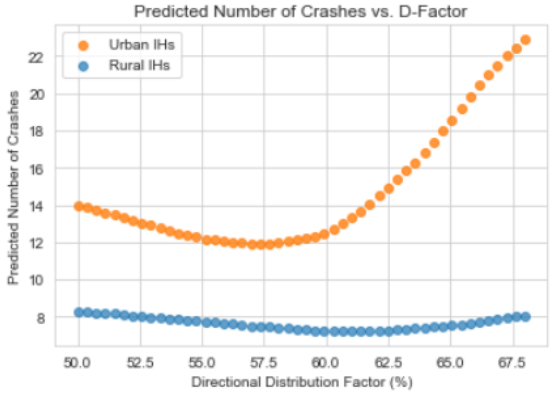
(b) Mean AADT



(c) Truck AADT



(d) Median Width



(e) Directional Distribution Factor

Cross-Validation Mean Scores:
R^2 = 0.740 (0.008)
RMSE: 14.023 (1.195)

Figure 5.10 Relationship between Selected Features and Predicted Number of Crashes

At the bottom of Figure 5.10, the mean and standard deviation of performance measures are provided, representing the accuracy and generalization of the DNN model after applying model averaging techniques.

Ride Score

Figure 5.10. (a) depicts the comparison between Urban IH routes (upper curve) and Rural IH routes (lower curve) concerning the relationship between ride score and the predicted number of crashes. The predictions for urban IH routes exhibit a gradual and consistent decline. In contrast, the curve for rural IH routes displays some fluctuations. Notably, there is a substantial decrease in the predicted number of crashes as the ride score increases from 3.0 to 3.5, followed by a continued steady decline beyond 3.5.

The findings of this study provide valuable insights by suggesting that road segments with higher ride scores are potentially associated with a lower probability of crashes while controlling for other pertinent factors. However, it is important to exercise caution when interpreting this observed correlation, as it does not establish a causal relationship between ride score and crash frequency. It is possible that a lower ride score serves as an indicative factor rather than a direct cause of increased accident occurrence. For instance, a road segment with a poor ride score may imply a history of insufficient maintenance, resulting in inadequate lane markings or other factors that can significantly impact safety. Therefore, additional empirical investigations are necessary to thoroughly examine the underlying factors that contribute to road segments with increased crash risk. These research efforts will assist in identifying the specific root causes and provide valuable insights for implementing targeted interventions to effectively mitigate crash risks.

Mean AADT

AADT is a paramount indicator of exposure and plays a crucial role in crash prediction. Safety performance functions (SPFs), which are regression equations used to estimate the average crash frequency for a specific site, are formulated as functions primarily dependent on AADT (AASHTO 2010).

In this study, AADT was also identified as the most influential feature in the proposed DNN model. The relationship between the three-year average AADT and the predicted number of crashes is illustrated in Figure 5.10.(b). From the predictions, it is evident that the predicted number of crashes exhibits a significant increase as the three-year average AADT rises. Furthermore, the curve representing urban IH routes exceeds the other one that represents rural IH routes, indicating that an IH segment in an urban area is more prone to experiencing a higher number of crashes compared to a rural IH segment with the same level of exposure and assuming other factors remain identical.

Upon comparing the vertical axis of AADT with the other features, it becomes apparent that the predicted number of crashes exhibits a significantly larger range compared to that of the other features. For instance, in Figure 5.10, when AADT hovers around 3,000, the predicted number of crashes for urban IH routes is below 10. As AADT reaches the maximum value of 100,000, the predicted number of crashes for urban IH routes is approximately 75. This represents a range of over 70 in predictions. In contrast, most other features (e.g., Ride Score, Truck AADT, and Median Width) exhibit a much narrower range, with the predicted crash frequency for urban IH routes varying by less than 5. These findings indicate that the predicted crash frequencies are primarily influenced by the mean AADT, establishing it as the dominant factor among other features. This observation aligns with intuitive expectations.

Truck AADT

For rural IH routes, it is evident that the predicted number of crashes gradually increases as the Truck AADT rises. However, the plot of predicted crash frequencies for urban IH routes reveals an opposite trend. In the plot, as shown in Figure 5.10.(c), it can be observed that the predicted number of crashes experiences a consistent decrease as the Truck AADT increases from 2,500 to 15,000. This decline is followed by a period of stable predictions between 15,000 and 24,000. This intriguing finding might suggest that urban IH segments with higher Truck AADT are more likely to have lower crash probabilities. One possible explanation for this observation is that – in urban areas – road segments with more trucks tend to have lower speed limits. Moreover, it should be noted that the magnitude of the change in predicted crash frequency is relatively small. Further investigations are required to uncover the underlying causes of this relationship.

Median Width

The plot depicting median width (Figure 5.10.(d)) reveals that, for both rural and urban IH segments, there is a slight decrease in the predicted number of crashes as the median width increases. This finding indicates that road segments within IH routes that have wider medians are more inclined to have fewer accidents. However, it is important to note that the influence of median width on the predicted crash frequency is quite limited. As the width increases from 3 to over 100, the predicted number of crashes decreases from 13 to 12 for urban IH routes, and from 8 to 7 for rural IH routes. This indicates that while there is a downward trend in the predicted crash frequency with increasing median width, the magnitude of this impact is considerably small.

Directional Distribution Factor

The plot representing the discretional distribution factor (Figure 5.10.(e)) indicates that the predicted number of crashes for rural IH routes remains unaffected by this factor. In contrast, for urban IH routes, the curve depicting the predicted crash frequency shows a gradual decline as the discretional distribution factor increases from 50 to 55. Subsequently, it remains stable between 55 and 60. However, beyond a discretional distribution factor of 60, there is a notable increase in the predicted number of crashes. This observation suggests a correlation between the discretional distribution factor and the predicted crash frequency for urban IH routes, wherein an increase in the factor leads to a corresponding change in the number of predicted crashes.

5.5 SUMMARY

This chapter presents an implementation of AI-oriented techniques to investigate the connection between pavement surface conditions and crash frequency. The study used 5,632 road segments from interstate highway routes in Texas to develop a DNN model for predicting crash frequency from various input features. These features included pavement condition scores, traffic volume, and roadway geometric characteristics. A two-phase hybrid hyperparameter optimization approach was employed to identify the optimal configurations of hyperparameters for the DNN model. The best model identified by the search algorithms was a network comprised of three hidden layers with 311, 103, and 362 nodes in each layer. The performance of the best DNN model was evaluated using customized testing data, achieving an RMSE of 14.645 and a R^2 of 0.680. A comparison study was conducted, and the result indicated that the proposed DNN model outperformed the comparative model based on a RFR in terms of both RMSE and R^2 scores using the customized testing dataset.

Using the best DNN model, a parametric analysis was conducted to gain a comprehensive understanding of the impact of individual variables on crash frequency. The analysis focused on the top five features that exhibited the highest influence on crash predictions, namely mean AADT, truck AADT, median width, directional distribution factor, and ride score. During the parametric analysis, the feature of interest was systematically modified one by one while the remaining features were kept constant. By doing so, it enables a detailed investigation into the isolated effects of each variable on crash frequency. This analysis contributes to the interpretability of the results by providing a thorough assessment of the individual contributions of these variables to crash frequency.

Overall, this study not only demonstrates the effectiveness of AI-oriented methods in establishing the link between pavement conditions and crash frequency but also provides valuable insights into the specific variables that significantly influence crash predictions. By improving the understanding of the relationship between pavement conditions and crash frequency, this research contributes to the development of accurate safety preventions and paves the way for enhanced pavement M&R programming strategies.

Chapter 6: Conclusions and Future Work

This dissertation endeavors to shed light on the effectiveness of leveraging data-driven methods to inform decision-making for roadway safety analysis and pavement management. This chapter highlights the key findings and major contributions of this research. Also provided are limitations of the proposed studies and recommendations for future improvements in applying data-driven methods in safety analysis and pavement management.

6.1 CONTRIBUTIONS

Key findings from and major contributions of this research are summarized below:

- This research makes a significant contribution by proposing an automated method for evaluating data consistency in motor vehicle crash databases (Chapter 2). Through the analysis of a large dataset comprising 1,136,551 on-system crashes, the study successfully categorized these data into six distinct misclassification types. The use of ArcGIS Python libraries enabled the automatic computation of the percentage of each misclassification type, revealing that approximately 27.1% of the crash data exhibited curve-related misclassification issues. The findings of this study highlight the applicability and effectiveness of the proposed methodological procedure. By implementing this procedure, transportation agencies can enhance the accuracy and reliability of safety analysis based on crash data. Moreover, the flexibility of the proposed procedure allows for customization to meet the specific needs of different transportation agencies. Overall, this study provides a replicable method and valuable insights for improving the identification of misclassifications in crash data, ultimately leading to more effective

interventions and investments in reducing motor vehicle-related fatalities and injuries.

- Chapter 3 of this research makes a valuable contribution by introducing a GIS-based methodological framework for identifying crash hot spots and ranking high-risk horizontal curves within a network. To achieve this, the study proposes a composite performance measure (CSI) and a novel safety index (SI) for horizontal curves. A comprehensive case study is provided to demonstrate the effectiveness of using these measures to screen sites with high crash risk at the network level. The case study showcases the practical application of the CSI measure in generating a crash hot spot map using spatial autocorrelation analysis. Additionally, the study presents the computation of the SI measures from obtained CSI values, which enables the quantification and prioritization of safety performance for horizontal curves. The outcomes from the case study emphasize the significance of the proposed method in effectively identifying hazardous locations, which can serve as targets for further investigation. This highlights the potential of the methodological framework to assist local transportation agencies in screening roadway networks and identifying areas with critical safety needs. The contribution of this study lies in providing a practical and effective approach that enables agencies to allocate resources and interventions efficiently, leading to enhanced safety performance for all road users.
- The contribution of Chapter 4 lies in its extensive coverage of relevant studies in applying AI-oriented algorithms to various pavement management activities, its comprehensive evaluation of achievements and limitations, and its provision of valuable perspectives for future research. By synthesizing and presenting this

knowledge, the study serves as a catalyst for the further development and implementation of AI-oriented solutions in pavement management.

- Chapter 5 presents the implementation of AI-oriented techniques to examine the correlation between pavement surface conditions and crash frequency. The study utilized 5,632 road segments from Texas interstate highways to develop a deep learning model which incorporates various input features such as pavement condition scores, traffic volume, and roadway geometric characteristics. Results of the comparison study revealed that the proposed DNN model outperformed the RFR model in terms of both RMSE and R^2 scores. Additionally, using the best DNN model, a parametric analysis was conducted to assess the impact of individual variables on crash frequency. This analysis enhances result interpretability by thoroughly evaluating the contribution of each variable to crash predictions. Overall, this study not only demonstrates the effectiveness of AI-oriented methods in establishing the link between pavement conditions and safety analysis but also provides valuable insights into the specific variables that significantly influence crash predictions. By improving the understanding of the relationship between pavement conditions and crash frequency, this research contributes to the development of accurate safety interventions and paves the way for enhanced pavement M&R programming strategies.

6.2 Limitations and Future Work

6.2.1 Improving Data Consistency in Motor Vehicle Crash Databases

It should be acknowledged that the potential causes presented in Chapter 2 are based on a limited sample of law enforcement crash reports. Consequently, it is imperative to conduct further research to validate the accuracy and reliability of these findings. The

exploration of potential causes for data inconsistencies in motor vehicle crash databases is an independent and underdeveloped area of research that requires additional attention and investigation.

It is also worth noting that the accuracy of the reference dataset has a significant impact on the outcome of the analysis. In the case study, the Texas Highway Curves GIS Layer was used as a reliable roadway geometry source for checking curve-related crash misclassification. Although the majority of horizontal curves along on-system routes appear on the Highway Curves GIS layer, there is a certain level of missing and incomplete curves in the Highway Curves GIS layer. This could affect the accuracy of the result to some degree. For future studies, it would be beneficial to explore other candidate references for horizontal curves (e.g., a more comprehensive and reliable curve dataset).

Last but not least, future studies can also explore the impacts of certain factors on curve-related crash misclassifications. For example, it would be interesting to investigate if certain types of roads are more likely to be misclassified than other types; if crashes that occurred on rural roads are more likely to be misclassified than those on urban roads; and, if roadbed types would affect the misclassification of crashes.

6.2.2 Monitoring the Latest AI Advancements in Pavement Management

In the review presented in Chapter 4, pavement management activities are categorized into three broad categories: distress evaluation, performance modeling, and M&R (maintenance and rehabilitation) programming. This categorization is primarily based on the authors' understanding of the sequential order in which these activities typically occur in real-world pavement management processes. The categorization chosen by the authors serves solely as a framework to organize the various aspects of pavement management for this study. The goal is to provide a structured presentation that facilitates

a comprehensive understanding of advancements in AI applications in pavement management. It is important to note that different divisions can be employed, and there is no definitive right or wrong approach with respect to how these activities are classified. Fundamentally, all these activities, regardless of how they are grouped, are integral parts of pavement management.

It is also important to acknowledge that the subject of the review, which focuses on AI-oriented techniques in pavement management, is a rapidly evolving field characterized by unprecedented high-speed advancements. The conclusions and recommendations are drawn from a comprehensive analysis of journal papers published between 2015 and 2020. It should be noted that the studies discussed in these papers were likely conducted even earlier than this timeframe. Considering the time-sensitive nature of research in this specific domain, it is not unexpected to find that certain gaps identified in the original study may have already been addressed by recent research conducted in the past few years.

The rapid pace of advancements in AI techniques demands continuous monitoring and updates to stay abreast of the latest developments. Consequently, it is recommended that future studies in this area periodically reassess the literature to incorporate the latest achievements and fill any potential gaps that may arise from the rapidly evolving nature of AI-oriented techniques. By doing so, researchers can ensure the continued relevance and applicability of their findings, ultimately contributing to improved decisions and enhanced effectiveness in pavement management.

6.2.3 Investigating the Relationship between Pavement Conditions and Road Safety

The findings presented in Chapter 5 provide valuable insights regarding the relationship between ride scores and crash frequency on road segments. The results suggest that road segments with higher ride scores may be associated with a lower likelihood of

crashes, which is consistent with previous research on the link between pavement roughness and crash frequency. However, it is important to interpret this correlation with caution, as it does not establish a causal relationship between ride score and crash frequency. The observed correlation may indicate that a lower ride score serves as an indicative factor rather than a direct cause of increased accident occurrence. For example, a road segment with a poor ride score may signal a history of insufficient maintenance, leading to inadequate lane markings or other factors that significantly impact safety. Therefore, there is a need for further empirical investigations to explore a comprehensive range of variables and identify the underlying factors contributing to locations with high crash risks. These additional research endeavors will facilitate a better understanding of the specific root causes behind crash risks and yield valuable insights for implementing targeted interventions. By identifying and addressing these factors, transportation agencies can effectively mitigate crash risks and enhance overall road safety.

References

Abiodun, O. I., *et al.* 2018. State-of-the-art in artificial neural network applications: A survey. *Heliyon,* 4(11), e00938.

Ai, D., *et al.* 2018. Automatic pixel-level pavement crack detection using information of multi-scale neighborhoods. *IEEE Access,* 6, 24452-24463.

Akiba, T., *et al.*, Optuna: A next-generation hyperparameter optimization framework. ed. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, 2623-2631.

Albin, R., *et al.*, 2016. *Low-cost treatments for horizontal curve safety 2016.* United States. Federal Highway Administration. Office of Safety. https://safety.fhwa.dot.gov/roadway_dept/countermeasures/horicurves/fhwasa15084/

Alhasan, A., *et al.* 2018. Impact of pavement surface condition on roadway departure crash risk in Iowa. *Infrastructures,* 3(2), 14.

Ali, J., *et al.*, Random Forests and Decision Trees. ed., 2012.

Alin, A. 2010. Multicollinearity. *Wiley interdisciplinary reviews: computational statistics,* 2(3), 370-374.

Al-Omari, B. and Darter, M. I. 1994. Relationships between international roughness index and present serviceability rating. *Transportation Research Record,* (1435).

American Association of State Highway and Transportation Officials (AASHTO). 2010. *Highway safety manual.* AASHTO.

American Association of State Highway and Transportation Officials (AASHTO). 2021. *Linear referencing system.* Retrieved from http://aii.transportation.org/Pages/LinearReferencingSystem.aspx

Anderson, T. K. 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention,* 41(3), 359-364.

ASTM, D. 2009. 6433-07 Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys. *ASTM International, West Conshohocken, Pennsylvania*.

Attoh-Okine, N. and Adarkwa, O. 2013. Pavement condition surveys–overview of current practices. *Delaware Center for Transportation, University of Delaware: Newark, DE, USA*.

Avila, M., *et al.*, 2D image based road pavement crack detection by calculating minimal paths and dynamic programming. ed. *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, 783-787.

Awad, M. and Khanna, R., 2015. Support vector regression. *Efficient learning machines.* Springer, 67-80.

Baddeley, A., *et al.* 2021. Analysing point patterns on networks—A review. *Spatial Statistics,* 42, 100435.

Bang, S., *et al.* 2019. Encoder–decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civil and Infrastructure Engineering,* 34(8), 713-727.

Barua, L., *et al.* 2020. A gradient boosting approach to understanding airport runway and taxiway pavement deterioration. *International journal of pavement engineering*, 1-15.

Bauer, K. M. and Harwood, D. W. 2013. Safety effects of horizontal curve and grade combinations on rural two-lane highways. *Transportation Research Record,* 2398(1), 37-49.

Belgiu, M. and Drăguţ, L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing,* 114, 24-31.

Bello, I., *et al.* 2016. Neural combinatorial optimization with reinforcement learning. *arXiv preprint arXiv:1611.09940*.

Bergstra, J. and Bengio, Y. 2012. Random search for hyper-parameter optimization. *Journal of machine learning research,* 13(2).

Berrar, D., 2019. Cross-Validation. https://www.researchgate.net/publication/324701535_Cross-Validation

Blincoe, L., *et al.*, 2015. *The economic and societal impact of motor vehicle crashes, 2010* (Revised. Report No. DOT HS 812 013). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013

Bogenreif, C., Souleyrette, R.R. and Hans, Z. 2012. Identifying and measuring horizontal curves and related effects on highway safety. Journal of Transportation Safety & Security, 4(3), pp.179-192.
Brownlee, J. 2014. Discover feature engineering, how to engineer features and how to get good at it. *Machine Learning Process*.
Brownlee, J. 2016. A gentle introduction to xgboost for applied machine learning. *Machine Learning Mastery*. https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/
Brownlee, J. 2019. How to use data scaling improve deep learning model stability and performance. *Machine Learning Mastery: Vermont, Australia*.
Cafiso, S., *et al.* 2021. Crash modification functions for pavement surface condition and geometric design indicators. *Accident Analysis & Prevention*, 149, 105887.
Cafiso, S., La Cava, G. and Montella, A. 2007. Safety index for evaluation of two-lane rural highways. *Transportation Research Record*, 2019(1), 136-145.
Campbell, J. L., 2012. *Human factors guidelines for road systems.* Transportation Research Board.
Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 679-698.
Cao, Z., *et al.* 2020. Using reinforcement learning to minimize the probability of delay occurrence in transportation. *IEEE Transactions on Vehicular Technology*, 69(3), 2424-2436.
Carvalho, D. V., Pereira, E. M. and Cardoso, J. S. 2019. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8), 832.
Centers for Disease Control and Prevention (CDC). 2020. Transportation Safety. Accessed on November 4, 2022. Retrieved from https://www.cdc.gov/transportationsafety/index.html
Ceylan, H., Bayrak, M. B. and Gopalakrishnan, K. 2014. Neural networks applications in pavement engineering: A recent survey. *International Journal of Pavement Research & Technology*, 7(6).
Cha, Y. J., Choi, W. and Büyüköztürk, O. 2017. Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering,* 32(5), 361-378.
Chan, C. Y., *et al.* 2010. Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS). *Journal of advanced transportation*, 44(3), 150-161.
Chen, S., *et al.* 2019. The global macroeconomic burden of road injuries: estimates and projections for 166 countries. *The Lancet Planetary Health*, 3(9), e390-e398.
Chen, T. and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. ed. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.
Chi, S., *et al.* 2013. Development of network-level project screening methods supporting the 4-year pavement management plan in Texas. *Journal of Management in Engineering*, 29(4), 482-494.
Chicco, D., Warrens, M. J. and Jurman, G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623.
Coenen, T. B. and Golroo, A. 2017. A review on automated pavement distress detection methods. *Cogent Engineering*, 4(1), 1374822.
Cord, A. and Chambon, S. 2012. Automatic road defect detection by textural pattern recognition based on AdaBoost. *Computer-Aided Civil and Infrastructure Engineering,* 27(4), 244-259.
Cubero-Fernandez, A., *et al.* 2017. Efficient pavement crack detection and classification. *EURASIP Journal on Image and Video Processing*, 2017(1), 1-11.
Daoud, J. I. 2017. Multicollinearity and regression analysis. ed. *Journal of Physics: Conference Series*, 012009.
Davies, R. and Sorenson, J. 2000. PAVEMENT PRESERVATION: PRESERVING OUR INVESTMENT IN HIGHWAYS. *Public Roads*, 63(4).
de Mello, R. F. and Moacir, A., 2018. A Practical Approach on the Statistical Learning Theory. Springer, New York, USA.

Domitrović, J., *et al.* 2018. Application of an artificial neural network in pavement management system. *Tehnički vjesnik,* 25(Supplement 2), 466-473.
Dong, Y., *et al.* 2019. Forecasting pavement performance with a feature fusion LSTM-BPNN model. ed. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1953-1962.
Donnell, E. T., *et al.* 2019. *Reducing Roadway Departure Crashes at Horizontal Curve Sections on Two-Lane Rural Highways.* United States. Federal Highway Administration. Office of Safety.
Doshi-Velez, F. and Kim, B. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
Du, Y., *et al.* 2020. Pavement distress detection and classification based on YOLO network. *International journal of pavement engineering*, 1-14.
Eisenbach, M., *et al.*, 2017. How to get pavement distress detection ready for deep learning? A systematic approach. ed. *2017 international joint conference on neural networks (IJCNN)*, 2039-2047.
Elbagalati, O., *et al.* 2018. Development of an enhanced decision-making tool for pavement management using a neural network pattern-recognition algorithm. *Journal of Transportation Engineering, Part B: Pavements*, 144(2), 04018018.
Elhadidy, A. A., El-Badawy, S. M. and Elbeltagi, E. E. 2019. A simplified pavement condition index regression model for pavement evaluation. *International journal of pavement engineering*, 1-10.
Environmental Systems Research Institute (ESRI). 2022a. *Notebooks in ArcGIS Pro*. Retrieved from https://pro.arcgis.com/en/pro-app/2.7/arcpy/get-started/pro-notebooks.htm
Environmental Systems Research Institute (ESRI). 2022b. *How Hot Spot Analysis (Getis-Ord Gi*) works*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm
Environmental Systems Research Institute (ESRI). 2022c. *Modeling spatial relationships – Best practices for selecting a conceptualization of spatial relationships*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/modeling-spatial-relationships.htm
Environmental Systems Research Institute (ESRI). 2022d. *Modeling spatial relationships – K nearest neighbors*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/modeling-spatial-relationships.htm
Erdogan, S., *et al.* 2008. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accident Analysis & Prevention,* 40(1), 174-181.
Farid, D. M., *et al.* 2014. Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert systems with applications,* 41(4), 1937-1946.
Federal Highway Administration (FHWA). 2016. Roadway Departure: Pavement Friction. Accessed on June 17, 2023. Retrieved from https://safety.fhwa.dot.gov/roadway_dept/pavement_friction/
Federal Highway Administration (FHWA). 2021. *Roadside Design Improvements at Curves*. Publication No. FHWA-SA-21-029. Retrieved from https://safety.fhwa.dot.gov/provencountermeasures/roadside_design.cfm
Federal Highway Administration (FHWA). 2022. *How Do Weather Events Impact Roads?* Retrieved from https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm
Fei, Y., *et al.* 2019. Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V. *IEEE Transactions on Intelligent Transportation Systems*, 21(1), 273-284.
Fernández-Delgado, M., *et al.* 2014. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research,* 15(1), 3133-3181.
Findley, D. J., *et al.* 2012. Modeling the impact of spatial relationships on horizontal curve safety. *Accident Analysis & Prevention,* 45, 296-304.
Findley, D.J., Zegeer, C.V., Sundstrom, C.A., Hummer, J.E., Rasdorf, W. and Fowler, T.J., 2012. Finding and measuring horizontal curves in a large highway network: a GIS approach. Public Works Management & Policy, 17(2), pp.189-211.
Flahaut, B. t., *et al.* 2003. The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. *Accident Analysis & Prevention,* 35(6), 991-1004.

France-Mensah, J. and O'Brien, W. J. 2018. Budget allocation models for pavement maintenance and rehabilitation: Comparative case study. *Journal of Management in Engineering,* 34(2), 05018002.
Freund, Y. and Schapire, R. E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences,* 55(1), 119-139.
Gao, L. and Zhang, Z., 2010. *Optimal infrastructure maintenance scheduling problem under budget uncertainty.* Southwest Region University Transportation Center (US).
Gao, L., *et al.* 2012. Network-level road pavement maintenance and rehabilitation scheduling for optimal performance improvement and budget utilization. *Computer-Aided Civil and Infrastructure Engineering,* 27(4), 278-287.
Gareth, J., et al., 2013. *An introduction to statistical learning: with applications in R*. Spinger.
Geedipally, S. R., *et al.*, 2022. *Calibrating the Highway Safety Manual Predictive Methods for Texas Highways: Technical Report.* Texas Department of Transportation. Research and Technology Implementation Office.
Geedipally, S. R., Pratt, M. P. and Lord, D. 2019. Effects of geometry and pavement friction on horizontal curve crash frequency. *Journal of Transportation Safety & Security,* 11(2), 167-188.
Georgiou, P., Plati, C. and Loizos, A. 2018. Soft computing models to predict pavement roughness: a comparative study. *Advances in Civil Engineering,* 2018.
Getis, A. and Ord, J. K. 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical analysis,* 24(3), 189-206.
Getis, A., 2010. Spatial autocorrelation. *Handbook of applied spatial analysis.* Springer, 255-278.
Glennon, J. C., Neuman, T. R., and Leisch, J. E. 1985. *Safety and operational considerations for design of rural highway curves*. Publication FHWA-RD-86-035. Federal Highway Administration (FHWA), 1985. https://www.mchenrysoftware.com/Leisch%20Curve%20Study.pdf
Gong, H., *et al.* 2018a. Improving accuracy of rutting prediction for mechanistic-empirical pavement design guide with deep neural networks. *Construction and Building Materials,* 190, 710-718.
Gong, H., *et al.* 2018b. Use of random forests regression for predicting IRI of asphalt pavements. *Construction and Building Materials,* 189, 890-897.
Gong, H., Sun, Y. and Huang, B. 2019. Gradient boosted models for enhancing fatigue cracking prediction in mechanistic-empirical pavement design guide. *Journal of Transportation Engineering, Part B: Pavements,* 145(2), 04019014.
Goodfellow, I., et al. 2016. *Deep learning*. MIT press Cambridge.
Gopalakrishnan, K. 2018. Deep learning in data-driven pavement image analysis and automated distress detection: A review. *Data,* 3(3), 28.
Gopalakrishnan, K., *et al.* 2017. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials,* 157, 322-330.
Hadjidemetriou, G. M., Vela, P. A. and Christodoulou, S. E. 2018. Automated pavement patch detection and quantification using support vector machines. *Journal of Computing in Civil Engineering,* 32(1), 04017073.
Hafez, M., Ksaibati, K. and Atadero, R. A. 2019. Optimizing expert-based decision-making of pavement maintenance using artificial neural networks with pattern-recognition algorithms. *Transportation Research Record,* 2673(11), 90-100.
Han, C., *et al.* 2020. Intelligent decision model of road maintenance based on improved weight random forest algorithm. *International journal of pavement engineering*, 1-13.
Harmon, T., Bahar, G. B. and Gross, F. B., 2018. *Crash costs for highway safety analysis.* United States. Federal Highway Administration. Office of Safety.
Hausman, J., Roff, T., & Clarke, J., 2014. All road network of linear referenced data reference manual. Federal Highway Administration (FHWA), USDOT Contract No. GS-35F-0001P. https://www.fhwa.dot.gov/policyinformation/hpms/documents/arnold_reference_manual_2014.pdf

Hazaymeh, K., Almagbile, A. and Alomari, A. H. 2022. Spatiotemporal Analysis of Traffic Accidents Hotspots Based on Geospatial Techniques. *ISPRS International Journal of Geo-Information*, 11(4), 260.
Hinton, G. E., Osindero, S. and Teh, Y.-W. 2006. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
Hoang, N.-D. and Nguyen, Q.-L. 2018. Automatic recognition of asphalt pavement cracks based on image processing and machine learning approaches: a comparative study on classifier performance. *Mathematical Problems in Engineering*, 2018.
Hoang, N.-D. and Nguyen, Q.-L. 2019. A novel method for asphalt pavement crack classification based on image processing and machine learning. *Engineering with Computers*, 35(2), 487-498.
Hooker, S., *et al.* 2019. A benchmark for interpretability methods in deep neural networks. *Advances in neural information processing systems*, 32.
Hossain, M., Gopisetti, L. and Miah, M. 2019. International roughness index prediction of flexible pavements using neural networks. *Journal of Transportation Engineering, Part B: Pavements*, 145(1), 04018058.
Hossain, M., Gopisetti, L. S. P. and Miah, M. S. 2020. Artificial neural network modelling to predict international roughness index of rigid pavements. *International Journal of Pavement Research and Technology*, 1-11.
Howard, J. and Bowles, M. 2012. The two most important algorithms in predictive modeling today. ed. *Strata Conference presentation, February*.
Hsieh, Y.-A. and Tsai, Y. J. 2020. Machine learning for crack detection: review and model performance comparison. *Journal of Computing in Civil Engineering*, 34(5), 04020038.
Huang, J., *et al.*, Speed/accuracy trade-offs for modern convolutional object detectors. ed. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, 7310-7311.
Hummer, J. E., *et al.* 2010. Curve collisions: road and collision characteristics and countermeasures. *Journal of Transportation Safety & Security*, 2(3), 203-220.
IJ, H. 2018. Statistics versus machine learning. *Nature methods*, 15(4), 233.
Insurance Institute for Highway Safety (IIHS). 2022. *Fatality Facts 2020 State by State - Fatal crash totals*. Retrieved from https://www.iihs.org/topics/fatality-statistics/detail/state-by-state
Intini, P., *et al.* 2019. Geometric and operational features of horizontal curves with specific regard to skidding proneness. *Infrastructures*, 5(1), 3.
Jafari Anarkooli, A., Nemtsov, I. and Persaud, B. 2021. Safety effects of maintenance treatments to improve pavement condition on two-lane rural roads—insights for pavement management. *Canadian Journal of Civil Engineering*, 48(10), 1287-1294.
Janani, L., *et al.* 2020. Prioritisation of pavement maintenance sections deploying functional characteristics of pavements. *International journal of pavement engineering*, 21(14), 1815-1822.
Justo-Silva, R., Ferreira, A. and Flintsch, G. 2021. Review on Machine Learning Techniques for Developing Pavement Performance Prediction Models. *Sustainability*, 13(9), 5248.
Karballaeezadeh, N., *et al.* 2019. Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan–Firuzkuh road). *Engineering Applications of Computational Fluid Mechanics*, 13(1), 188-198.
Karballaeezadeh, N., *et al.* 2020. Intelligent road inspection with advanced machine learning; hybrid prediction models for smart mobility and transportation maintenance systems. *Energies*, 13(7), 1718.
Kargah-Ostadi, N. and Stoffels, S. M. 2015. Framework for development and comprehensive comparison of empirical pavement performance models. *Journal of transportation engineering*, 141(8), 04015012.
Khan, G., *et al.* 2013. Safety evaluation of horizontal curves on rural undivided roads. *Transportation Research Record*, 2386(1), 147-157.
Khurana, U., Samulowitz, H. and Turaga, D. 2017. Feature engineering for predictive modeling using reinforcement learning. *arXiv preprint arXiv:1709.07150*.

Koch, C., *et al.* 2015. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics,* 29(2), 196-210.
Kohavi, R., A study of cross-validation and bootstrap for accuracy estimation and model selection. ed. *Ijcai*, 1995, 1137-1145.
Le, K. G., Liu, P. and Lin, L.-T. 2022. Traffic accident hotspot identification by integrating kernel density estimation and spatial autocorrelation analysis: a case study. *International journal of crashworthiness,* 27(2), 543-553.
Lei, X., *et al.* 2020. Automated Pavement Distress Detection and Deterioration Analysis Using Street View Map. *IEEE Access,* 8, 76163-76172.
Li, B., *et al.* 2019. Automatic segmentation and enhancement of pavement cracks based on 3D pavement images. *Journal of Advanced Transportation,* 2019.
Li, B., *et al.* 2020. Automatic classification of pavement crack using deep convolutional neural network. *International journal of pavement engineering,* 21(4), 457-463.
Li, H., *et al.* 2018. Automatic pavement crack detection by multi-scale image fusion. *IEEE Transactions on Intelligent Transportation Systems,* 20(6), 2025-2036.
Li, J. Q. and Yu, W. 2021. Enhanced safety performance function for highway segments in Oklahoma. *Journal of Infrastructure Systems,* 27(3), 04021018.
Li, L., *et al.* 2014. Automatic pavement crack recognition based on BP neural network. *PROMET-Traffic&Transportation,* 26(1), 11-22.
Li, L., Lv, Y. and Wang, F.-Y. 2016. Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica,* 3(3), 247-254.
Li, Y., Liu, C. and Ding, L., 2013. Impact of pavement conditions on crash severity. Accident Analysis & Prevention, 59, pp.399-406.
Liashchynskyi, P. and Liashchynskyi, P. 2019. Grid search, random search, genetic algorithm: a big comparison for NAS. *arXiv preprint arXiv:1912.06059*.
Liu, C., Chen, C. L. and Utter, D., 2005a. *Trend and pattern analysis of highway crash fatality by month and day.* (Report No. DOT HS 809 855). Washington, DC: National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/809855
Liu, C., Chen, C. L., & Utter, D., 2005b. Speeding-related crash fatalities by month, day, and selected holiday periods. (Traffic Safety Facts Crash Stats. Report No. DOT HS 809 890). Washington, DC: National Highway Traffic Safety Administration.
Liu, C., Chen, C.L., Subramanian, R. and Utter, D. 2005c. Analysis of speeding-related fatal motor vehicle traffic crashes. No. HS-809 839.
Liu, P., *et al.* 2017. SVM or deep learning? A comparative study on remote sensing image classification. *Soft Computing,* 21(23), 7053-7065.
Long, K., Wu, H., Zhang, Z., & Murphy, M. 2014. Quantitative relationship between crash risks and pavement skid resistance (No. FHWA/TX-13/0-6713-1). Texas. Dept. of Transportation. Research and Technology Implementation Office.
Lyon, C. and Persaud, B. 2008. Safety effects of targeted program to improve skid resistance. *Transportation Research Record,* 2068(1), 135-140.
Lyon, C., *et al.* 2020. Empirical Bayes before-after study to develop crash modification factors and functions for high friction surface treatments on curves and ramps. *Transportation Research Record,* 2674(12), 505-514.
Lyon, C., Persaud, B. and Merritt, D. 2018. Quantifying the safety effects of pavement friction improvements–results from a large-scale study. *International journal of pavement engineering,* 19(2), 145-152.
Maeda, H., *et al.* 2018. Road damage detection and classification using deep neural networks with smartphone images. *Computer-Aided Civil and Infrastructure Engineering,* 33(12), 1127-1141.
Mahmood, M., *et al.*, A unified artificial neural network model for asphalt pavement distress prediction. ed. *Proceedings of the Institution of Civil Engineers-Transport*, 2020, 1-23.

Majidifard, H., *et al.* 2020. Pavement Image Datasets: A New Benchmark Dataset to Classify and Densify Pavement Distresses. *Transportation Research Record*, 2674(2), 328-339.
Mao, H., *et al.*, Resource management with deep reinforcement learning. ed. *Proceedings of the 15th ACM workshop on hot topics in networks*, 2016, 50-56.
Mathavan, S., Kamal, K. and Rahman, M. 2015. A review of three-dimensional imaging technologies for pavement distress detection and measurements. *IEEE Transactions on Intelligent Transportation Systems*, 16(5), 2353-2362.
Mayora, J. M. P. and Piña, R. J. 2009. An assessment of the skid resistance effect on traffic safety under wet-pavement conditions. *Accident Analysis & Prevention*, 41(4), 881-886.
Mazari, M. and Rodriguez, D. D. 2016. Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(5), 448-455.
Merritt, D. K., Lyon, C. and Persaud, B., 2015. *Evaluation of pavement safety performance.* United States. Federal Highway Administration.
Milad, A., *et al.*, Using an Azure Machine Learning Approach for Flexible Pavement Maintenance. ed. *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, 2020, 146-150.
Miller, J. S. and Bellinger, W. Y., 2014. *Distress Identification Manual for the Long-Term Pavement Performance Program (Fifth Revised Edition).* United States. Office of Infrastructure Research and Development.
Mohaymany, A. S., Shahri, M. and Mirbagheri, B. 2013. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. Geo-spatial Information Science, 16(2), 113-119.
Mokhtari, S., Wu, L. and Yun, H.-B. 2016. Comparison of supervised classification techniques for vision-based pavement crack detection. *Transportation Research Record*, 2595(1), 119-127.
Morales, F. J., *et al.* 2020. A machine learning methodology to predict alerts and maintenance interventions in roads. *Road Materials and Pavement Design*, 1-22.
Mousa, M., Elseifi, M. A. and Abdel-Khalek, A. 2019. Development of Tree-Based Algorithm for Prediction of Field Performance of Asphalt Concrete Overlays. *Journal of Transportation Engineering, Part B: Pavements*, 145(2), 04019011.
Múčka, P. 2017. International Roughness Index specifications around the world. *Road Materials and Pavement Design*, 18(4), 929-965.
Murphy, M., *et al.* 2018. Economic and Safety Considerations: Motor Vehicle Safety Inspections for Passenger Vehicles in Texas. *Center for Transportation Research at the University of Texas at Austin*. https://library.ctr.utexas.edu/ctr-publications/iac/sb2076.pdf
Natekin, A. and Knoll, A. 2013. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7, 21.
National Highway Traffic Safety Administration (NHTSA), 2017. MMUCC Guideline: Model Minimum Uniform Crash Criteria Fifth Edition. (Report No. DOT HS 812 433).
National Highway Traffic Safety Administration (NHTSA). 2019. *Estimate of motor vehicle traffic crash fatalities for the holiday periods of 2019*. (Traffic Safety Facts Research Note. Report No. DOT HS 812 823). National Center for Statistics and Analysis. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812823
National Highway Traffic Safety Administration (NHTSA). 2022a. *Newly Released Estimates Show Traffic Fatalities Reached a 16-Year High in 2021*. Retrieved from https://www.nhtsa.gov/press-releases/early-estimate-2021-traffic-fatalities
National Highway Traffic Safety Administration (NHTSA). 2022b. *Alcohol-impaired driving: 2020 data* (Traffic Safety Facts. Report No. DOT HS 813 294). National Center for Statistics and Analysis. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813294
National Highway Traffic Safety Administration (NHTSA). 2022c. *Fatality Analysis Reporting System*. Retrieved from https://www.nhtsa.gov/crash-data-systems/fatality-analysis-reporting-system

National Safety Council (NSC). 2022. *Distracted Driving*. Retrieved from https://injuryfacts.nsc.org/motor-vehicle/motor-vehicle-safety-issues/distracted-driving/

Neuman, T., *et al.*, 2003. *A guide for addressing run-off-road collisions. Vol. 6 of NCHRP report 500: Guidance for implementation of the AASHTO Strategic Highway Safety Plan.* Washington, DC: Transportation Research Board of the National Academies.

Nhat-Duc, H., Nguyen, Q.-L. and Tran, V.-D. 2018. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. *Automation in Construction*, 94, 203-213.

Nie, M. and Wang, K., Pavement distress detection based on transfer learning. ed. *2018 5th International Conference on Systems and Informatics (ICSAI)*, 2018, 435-439.

Noriega, L. 2005. Multilayer perceptron tutorial. *School of Computing. Staffordshire University,* 4, 5.

akden-Rayner, L., *et al.* 2017. Precision radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Scientific reports,* 7(1), 1-13.

Okabe, A., Okunuki, K. i. and Shiode, S. 2006. SANET: a toolbox for spatial analysis on a network. *Geographical analysis,* 38(1), 57-66.

Oliveira, H. and Correia, P. L. 2009. Automatic road crack segmentation using entropy and image dynamic thresholding. ed. *2009 17th European Signal Processing Conference*, 622-626.

Oliveira, H. and Correia, P. L. 2014. CrackIT—An image processing toolbox for crack detection and characterization. ed. *2014 IEEE international conference on image processing (ICIP)*, 798-802.

Ord, J. K. and Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis,* 27(4), 286-306.

Othman, S., Thomson, R. and Lannér, G., 2012. Using naturalistic field operational test data to identify horizontal curves. Journal of transportation engineering, 138(9), pp.1151-1160.

Park, K., Thomas, N. E. and Wayne Lee, K. 2007. Applicability of the international roughness index as a predictor of asphalt pavement condition. *Journal of transportation engineering,* 133(12), 706-709.

Pauly, L., *et al.*, Deeper networks for pavement crack detection. ed. *Proceedings of the 34th ISARC*, 2017, 479-485.

Pierce, L. M. and Weitzel, N. D., 2019. *Automated Pavement Condition Surveys.*

Ping, W., *et al.* 2017. Deep voice 3: Scaling text-to-speech with convolutional sequence learning. *arXiv preprint arXiv:1710.07654*.

Piryonesi, S. M. and El-Diraby, T. E. 2020. Data analytics in asset management: Cost-effective prediction of the pavement condition index. *Journal of Infrastructure Systems,* 26(1), 04019036.

Politis, S. S., *et al.* 2020. Framework for network-level pavement condition assessment using remote sensing data mining. *Journal of Applied Remote Sensing,* 14(2), 024504.

Popescu, M.-C., *et al.* 2009. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579-588.

Quddus, M. 2013. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. *Journal of Transportation Safety & Security*, 5(1), 27-45.

Ragnoli, A., De Blasiis, M. R. and Di Benedetto, A. 2018. Pavement distress detection methods: A review. *Infrastructures,* 3(4), 58.

Ramchoun, H., *et al.* 2016. Multilayer perceptron: Architecture optimization and training.

Rodriguez-Lozano, F. J., *et al.* 2020. Benefits of ensemble models in road pavement cracking classification. *Computer-Aided Civil and Infrastructure Engineering*.

Satria, R. and Castro, M. 2016. GIS tools for analyzing accidents and road design: a review. *Transportation research procedia*, 18, 242-247.

Schapire, R. E. 2013. Explaining adaboost. *Empirical inference.* Springer, 37-52.

Schneider IV, W. H., Savolainen, P. T. and Moore, D. N. 2010. Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways. *Transportation Research Record,* 2194(1), 91-98.

Schneider, R. J., *et al.* 2021. United States fatal pedestrian crash hot spot locations and characteristics. *Journal of transport and land use,* 14(1), 1-23.

Serigos, P. A., *et al.*, 2015. *Automated distress surveys: analysis of network level data (Phase III).*
Shahin, M. Y., 2005. *Pavement management for airports, roads, and parking lots.* Springer New York.
Shahzad, M. 2020. Review of road accident analysis using GIS technique. *International journal of injury control and safety promotion*, 27(4), 472-481.
Sharma, S., Sharma, S. and Athaiya, A. 2017. Activation functions in neural networks. *Towards Data Sci*, 6(12), 310-316.
Shi, Y., *et al.* 2016. Automatic road crack detection using random structured forests. *IEEE Transactions on Intelligent Transportation Systems*, 17(12), 3434-3445.
Shipp, E. M., Trueblood, A. B., Perez, M., Ko, M., Wu, L., Stewart, C., Pant, A., and Chigoy, B. Analysis of motorcycle crashes in Texas, 2010–2017. Texas A&M Transportation Institute, 2018. https://www.looklearnlive.org/wp-content/uploads/2018/09/MotorcycleAnalysisReportFinal_Final.pdf
Sola, J. and Sevilla, J. 1997. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on nuclear science*, 44(3), 1464-1468.
Sollazzo, G., Fwa, T. and Bosurgi, G. 2017. An ANN model to correlate roughness and structural performance in asphalt pavements. *Construction and Building Materials*, 134, 684-693.
Soltani, A. and Askari, S. 2017. Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, 48(3), 637-647.
Songchitruksa, P. and Zeng, X. 2010. Getis–Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data. *Transportation Research Record*, 2165(1), 42-51.
Souleyrette, R. R., 2011. *Systematic identification of high crash locations.* Iowa State University. Center for Transportation Research and Education. https://intrans.iastate.edu/app/uploads/2018/03/hcl-1.pdf
Steudle, K., *et al.* 2012. *Pavement Management Guide, Second Edition.* AASHTO: The American Association of State Highway and Transportation Officials.
Stevenson, J. 2021. *TxDOT Pavement Manual.* http://onlinemanuals.txdot.gov/txdotmanuals/pdm/pdm.pdf
Stewart, T. 2022. Overview of motor vehicle crashes in 2020 (Report No. DOT HS 813 266). Washington, DC: National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813266
Strong, A. 2016. Applications of artificial intelligence & associated technologies. *Science [ETEBMS-2016]*, 5(6).
Subramanian, R., & Liu, C. 2004. *Fatalities related to impaired driving during the Christmas and New Year's Day holiday periods*. (Traffic Safety Facts Crash Stats. Report No. DOT HS 809 824). Washington, DC: National Highway Traffic Safety Administration.
Sun, J. and Zhang, Z. 2020. A post-disaster resource allocation framework for improving resilience of interdependent infrastructure networks. *Transportation Research Part D: Transport and Environment*, 85, 102455.
Sun, L., Kamaliardakani, M. and Zhang, Y. 2016. Weighted neighborhood pixels segmentation method for automated detection of cracks on pavement surface images. *Journal of Computing in Civil Engineering*, 30(2), 04015021.
Sun, Y., *et al.* 2020. Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE transactions on cybernetics*, 50(9), 3840-3854.
Sutton, R. S. and Barto, A. G., 2018. *Reinforcement learning: An introduction.* MIT press.
Szandała, T. 2021. Review and comparison of commonly used activation functions for deep neural networks. *Bio-inspired neurocomputing*, 203-224.
Szegedy, C., Toshev, A. and Erhan, D. 2013. Deep neural networks for object detection. ed. *Advances in neural information processing systems*, 2553-2561.
Tabatabaee, N., Ziyadi, M. and Shafahi, Y. 2013. Two-stage support vector classifier and recurrent neural network predictor for pavement performance modeling. *Journal of Infrastructure Systems*, 19(3), 266-274.
Texas Department of Transportation (TxDOT). 2018. Geospatial Roadway Inventory Database User Guide. https://ftp.txdot.gov/pub/txdot-info/tpp/GRID/GRID_User_Guide.pdf

Texas Department of Transportation (TxDOT). 2021a. Public Extract Crash Data from the Crash Records Information System (CRIS). Texas Department of Transportation (TxDOT). Accessed on November 21, 2021. Available at: https://cris.dot.state.tx.us/secure/Share/

Texas Department of Transportation (TxDOT). 2021b. Highway Curves GIS Layer (dataset) – Draft curves along state-maintained roadways. Transportation Planning and Programming Division – Data Management Section, Texas Department of Transportation (TxDOT). Accessed on September 14, 2021. Available at: http://arcg.is/1SPG8i

Texas Department of Transportation (TxDOT). 2021c. Crash data analysis and statistics – Automated crash data extract files. Accessed on November 4, 2022. Retrieved from https://www.txdot.gov/content/txdotreimagine/us/en/home/data-maps/crash-reports-records/crash-data-analysis-statistics.html

Texas Department of Transportation (TxDOT). 2021d. Condition of Texas Pavements - Pavement Management Information System (PMIS) Annual Report FY 2018-2021. https://ftp.txdot.gov/pub/txdot/mtd/pmis-annual-report.pdf

Thakali, L., Kwon, T. J. and Fu, L. 2015. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. *Journal of Modern Transportation*, 23(2), 93-106.

Timothée, P., *et al.* 2010. A network based kernel density estimator applied to Barcelona economic activities. ed. *International conference on computational science and its applications*, 2010, 32-45.

Tola, A. M., *et al.* 2021. Severity, spatial pattern and statistical analysis of road traffic crash hot spots in Ethiopia. *Applied Sciences*, 11(19), 8828.

Tong, Z., Gao, J. and Zhang, H. 2017. Recognition, location, measurement, and 3D reconstruction of concealed cracks using convolutional neural networks. *Construction and Building Materials*, 146, 775-787.

Torbic, D. J., *et al.*, 2004. *Guidance for implementation of the AASHTO strategic highway safety plan. Volume 7: A guide for reducing collisions on horizontal curves.* 0309087600. (No. Project G17-18 (3) FY'00).

Truong, L. T. and Somenahalli, S. V. 2011. Using GIS to identify pedestrian-vehicle crash hot spots and unsafe bus stops. *Journal of Public Transportation*, 14(1), 6.

Tsai, Y.-C. J., *et al.* 2021. Automatically detect and classify asphalt pavement raveling severity using 3D technology and machine learning. *International Journal of Pavement Research and Technology*, 14(4), 487-495.

Tu, J. V. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11), 1225-1231.

U.S. Transportation and Infrastructure Committee. *History*. 2023. Retrieved from: https://transportation.house.gov/about/history.htm

Uddin, W., Hudson, W. R. and Haas, R. C. 2013. *Public infrastructure asset management.* McGraw Hill Professional.

Ukhwah, E. N., Yuniarno, E. M. and Suprapto, Y. K. 2019. Asphalt Pavement Pothole Detection using Deep learning method based on YOLO Neural Network. ed. *2019 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 35-40.

United States Department of Transportation (USDOT). 2022. *National Roadway Safety Strategy*. https://www.transportation.gov/sites/dot.gov/files/2022-02/USDOT-National-Roadway-Safety-Strategy.pdf

Vandervalk, A., Jeanotte, K., Snyder, D., Bauer, J. and Systematics, C. 2016. State of the practice on data access, sharing, and integration. No. FHWA-HRT-15-072. United States. Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/operations/15072/index.cfm

Varshney, K. R. and Alemzadeh, H. 2017. On the safety of machine learning: Cyber-physical systems, decision sciences, and data products. *Big data*, 5(3), 246-255.

Verster, T. and Fourie, E. 2018. The good, the bad and the ugly of South African fatal road accidents. South African Journal of Science, 114(7-8), 63-69.

Wang, F., Zhang, Z. and Machemehl, R. B. 2003. Decision-making problem for managing pavement maintenance and rehabilitation projects. *Transportation Research Record*, 1853(1), 21-28.
Wang, J., *et al.*, Cnn-rnn: A unified framework for multi-label image classification. ed. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 2285-2294.
World Health Organization (WHO). 2022. *Road traffic injuries*. Accessed on November 4, 2022. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
Xie, Z. and Yan, J. 2008. Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems*, 32(5), 396-406.
Xie, Z. and Yan, J. 2013. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of transport geography*, 31, 64-71.
Xin, C., *et al.* 2017. Safety effects of horizontal curve design on motorcycle crash frequency on rural, two-lane, undivided highways in Florida. *Transportation Research Record*, 2637(1), 1-8.
Xu, H. and Wei, D., 2016. Improved identification and calculation of horizontal curves with geographic information system road layers. Transportation Research Record, 2595(1), pp.50-58.
Xu, Y., Han, Z., Murphy, M. and Zhang, Z. 2022. Development of an Automated Methodological Procedure to Improve the Identification of Curve-Related Crashes in the Crash Records Information System (CRIS) (No. FHWA/TX-22/0-7050-1). University of Texas at Austin. Center for Transportation Research. https://rosap.ntl.bts.gov/view/dot/64509
Xu, Z., *et al.* 2017. A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs. ed. *2017 IEEE International Conference on Communications (ICC)*, 2017, 1-6.
Yamada, I. and Thill, J.-C. 2004. Comparison of planar and network K-functions in traffic accident analysis. *Journal of transport geography*, 12(2), 149-158.
Yamany, M. S., *et al.* 2020. Characterizing the Performance of Interstate Flexible Pavements Using Artificial Neural Networks and Random Parameters Regression. *Journal of Infrastructure Systems*, 26(2), 04020010.
Yao, L., *et al.* 2020. Deep reinforcement learning for long-term pavement maintenance planning. *Computer-Aided Civil and Infrastructure Engineering*, 35(11), 1230-1245.
Yellman, M. A. 2022. Motor Vehicle Crash Deaths—United States and 28 Other High-Income Countries, 2015 and 2019. *MMWR. Morbidity and Mortality Weekly Report*, 71.
Zakeri, H., Nejad, F. M. and Fahimifar, A. 2017. Image based techniques for crack detection, classification and quantification in asphalt pavement: a review. *Archives of Computational Methods in Engineering*, 24(4), 935-977.
Zeiada, W., *et al.* 2019. Investigation and modelling of asphalt pavement performance in cold regions. *International journal of pavement engineering*, 20(8), 986-997.
Zeiada, W., *et al.* 2020. Machine learning for pavement performance modelling in warm climate regions. *Arabian Journal for Science and Engineering*, 1-19.
Zhang, A., *et al.* 2017a. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. *Computer-Aided Civil and Infrastructure Engineering,* 32(10), 805-819.
Zhang, Q., Zhou, D. and Zeng, X. 2017b. HeartID: A multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications. *IEEE Access*, 5, 11805-11816.
Zhang, A., *et al.* 2018a. Deep learning–based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet. *Journal of Computing in Civil Engineering*, 32(5), 04018041.
Zhang, K., Cheng, H. and Zhang, B. 2018b. Unified approach to pavement crack and sealed crack detection using preclassification based on transfer learning. *Journal of Computing in Civil Engineering*, 32(2), 04018001.
Zhang, A., *et al.* 2019. Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network. *Computer-Aided Civil and Infrastructure Engineering,* 34(3), 213-229.
Zhang, L., *et al.* 2016. Road crack detection using deep convolutional neural network. ed. *2016 IEEE international conference on image processing (ICIP)*, 3708-3712.

Zhou, Z., *et al.* 2019. Optimization of molecules via deep reinforcement learning. *Scientific reports*, 9(1), 1-10.
Ziari, H., *et al.* 2016. Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods. *International journal of pavement engineering*, 17(9), 776-788.
Zou, Q., *et al.* 2012. CrackTree: Automatic crack detection from pavement images. *Pattern Recognition Letters*, 33(3), 227-238.