SAFETY ANALYSIS OF TRANSPORTATION MODES USING DATA ANALYTICS

By

Ishita Dash

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
in
Civil Engineering
June 30, 2022
Nashville, Tennessee

Approved:

Dr. Mark Abkowitz (Chair)
Dr. Hiba Baroud
Dr. Craig Philip
Dr. Jonathan Gilligan

Copyright © 2022 Ishita Dash
All Rights Reserved

To my family…you will always be my strength.

Acknowledgment

A special thank you to my advisor Dr. Mark Abkowitz for his encouragement and kindness, without whom there would be no light at the end of the Ph.D. program tunnel, and for helping me make sense of the confusion. I would also like to thank my committee members Dr. Craig Philip, Dr. Hiba Baroud, and Dr. Jonathan Gilligan, for their suggestions, guidance, and support.

I am also thankful to Dr. Dan Work and Dr. Douglas Adams for the wonderful experience during my teaching assistantships at Vanderbilt.

The research reported in this dissertation was funded by the U.S. Department of Transportation, Pipeline and Hazardous Materials Safety Administration (PHMSA), and by a Congestion Mitigation and Air Quality (CMAQ) grant administered by the U.S. Federal Highway Administration and awarded by the Tennessee Department of Transportation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Transportation and Data Analytics

Transportation is a critical infrastructure system, whose importance is underscored by the need to provide adequate mobility for people and goods to meet society's demands. An essential element in satisfying this requirement is to provide an acceptable level of safety performance across the transportation modes upon which the public relies. This is especially challenging given the constant change in land use, demographics, technology innovation, availability of transportation alternatives, and other factors.

Against this landscape, however, are opportunities to make improvements in safety methods and practices to keep pace with these challenges and provide an enhanced level of safety. This becomes possible, in part, due to advances in information technology that produce a plethora of exposure and accident data, as well as analytical tools that utilize these data to develop explanatory and predictive tools.

1.2 Motivation of Study

Regardless of the type of transportation mode, the goal is the same, i.e., to enhance transportation safety by reducing crashes resulting in severe outcomes. Accidents involving motor vehicles (including crashes with pedestrians and bicyclists) account for the vast majority of the transportation fatalities, with railroads also contributing to this unfortunate situation (USDOT, 2021).

From Table 1.1 on fatalities in transportation by modes from the year 2017 to 2021, we observe that highway fatalities account for 95 percent of all the fatalities, with approximately 17 percent of pedestrians and pedalcyclists (i.e., bicyclists). Railway fatalities are the next highest category. An interesting trend is seen for the year 2020, where the total fatalities increased by six percent, including pedestrians and bicyclists in spite of a 13 percent decrease in vehicle miles of travel. Pedestrian, bicyclist, and railway fatalities increased by 13, five, and 20 percent, respectively, from 2020 to 2021(NHTSA, 2022; BTS, 2021). Based on these figures and facts the research focuses on pedestrian, bicycle, and railway modes where various data analytics techniques have been applied to improve the safety of these transportation modes.

Table 1.1: Fatality in Transportation Modes (2017-2021) (NHTSA, 2022; BTS, 2021)

Mode	Fatalities - 2017	Fatalities - 2018	Fatalities - 2019	Fatalities - 2020	Fatalities - 2021
Air	347	395	452	349	-
Highway	37,473	36,835	36,096	38,680	42,915
- **Pedestrian and Pedalcyclist**	6,881	7,245	7,051	7,127	8,327
Railway	817	805	863	752	902
Transit	249	260	268	289	290
Water	709	682	707	852	-
Pipeline	7	7	11	15	13

Of particular concern to railway safety are the many thousands of hazmat shipments that occur daily in the U.S., traveling across a vast network. While the industry has amassed an impressive safety and security track record, incidents continue to occur, posing risks to the health and safety of hazmat responders, inspectors, carriers, shippers, other transportation stakeholders, and communities at large. This accentuates the need for research that can advance the development of innovative techniques to minimize the risks of transporting hazardous materials. One promising area for achieving this goal is deploying intelligent detection systems and related communication technology to improve the accuracy, timeliness, and breadth of communication among stakeholders, focusing on both incident prevention and consequence mitigation.

Providing the shipper with these capabilities is arguably key to influencing the entire hazmat transportation supply chain. The hazmat shipper loads the product knows its material properties, often owns the fleet equipment, and never wholly relinquishes its custodial role (even when in the hands of a carrier) until the product is successfully delivered to the customer. Providing the shipper with these capabilities is key to influencing the entire hazmat transportation supply chain. The hazmat shipper loads the product, knows its material properties, often owns the fleet equipment, and never wholly relinquishes its custodial role (even when in the hands of a carrier) until the product is successfully delivered to the customer. In this respect, it is the shipper who is at the helm of implementing the technology solution that collects vital monitoring/detection information, assesses it, and communicates what other stakeholders need to know in a timely fashion to ensure that the safety and security of the hazmat shipment are highly coordinated and effectively managed.

Regarding pedestrian and bicyclist safety, walking and bicycling play an essential role as significant non-motorized travel modes in many urban areas. While increasingly serving as a vital part of an integrated transportation demand management system and a sustainable mobility option, interest in walking and biking as an active transportation mode has been unfortunately accompanied by an increased crash count, many with incapacitating injuries or fatal outcomes. Pedestrians and bicyclists are among the most vulnerable road users when a motor vehicle is involved. To support planning a scalable and streamlined long-term transportation goal of zero accidents, it is thus essential to improve walking and bicycling safety. With the crash data recorded and maintained by the Tennessee Department of Transportation (TDOT) and machine learning algorithms, we can identify and understand the critical factors that influence severe crash outcomes, understand their interactions, and identify and prioritize policies and actions to mitigate these risks, which benefits both modes.

1.3 Dissertation Overview

This dissertation focuses on how information technology and data analytics can be utilized to improve our understanding and implement policies that enhance the safety of passenger and freight transportation. This is investigated through three targeted applications: 1) rail transport of hazardous materials, 2) pedestrian transport, and 3) bicycle travel.

The dissertation is organized as follows. Chapter 2 describes a study on how hazmat shippers can leverage the integration of several technologies to achieve enhanced hazmat transport safety and security across various freight surface transportation modes, using rail freight transport as the research focus. A conceptual design is developed for deploying this system for use by rail hazmat shippers This includes schematics showing individual system components (hardware and software) and their interoperability and a narrative that describes how each component is utilized and how information is transmitted to the rail hazmat shipper and subsequently merged into an integrated database. Guidance is also provided to set thresholds, trigger alerts/notifications, and communicate these alerts/notifications to appropriate hazmat transportation stakeholders. Chapters 3 and 4 describe the development of classification models for determining factors influencing severe crash outcomes involving pedestrians and bicyclists, respectively, and comparing their predictive results on a highly imbalanced dataset using three data balancing methods. Chapter 5 builds on the results of the previous two chapters to demonstrate the importance of not only severe crash frequency, but also severe crash rate, through estimation of corresponding pedestrian and bicycle exposure at locations of interest. Chapter 6 contains concluding remarks and future research recommendations.

As Chapters from 2 to 5 consist of either published or soon-to-be-published journal articles, they are presented in the form of the manuscripts prepared for this purpose. For this reason, these chapters are intended to be read separately, and some content may overlap.

Chapter 2

Design and Implementation of an Integrated Technology System for Rail Shipper Safety & Security

This chapter presents an overview of a system design for addressing critical hazardous materials risks and how key technological components can operate together to improve rail safety and security. It includes a description of the design and application of an existing integrated technology system to make it available to any rail hazmat shipper for implementation consideration. Schematics are presented showing individual system components (hardware and software) and their interoperability and a narrative that describes how each component is utilized and how information is transmitted to the hazmat shipper and subsequently merged into an integrated database. Guidance is also provided to set thresholds, trigger alerts/notifications, and communicate these alerts/notifications to appropriate hazmat transportation stakeholders. A case study is subsequently presented that illustrates how system outputs are used in making improved risk-informed decisions.

2.1 System Design Overview

The technology solution utilizes as its foundation an integrated safety/security system for rail shipments of high-hazard cargo that begins inside the shipper's fence line, continues while a shipment is in transit, and does not end until the cargo is successfully delivered to the customer (Figure 2.1).

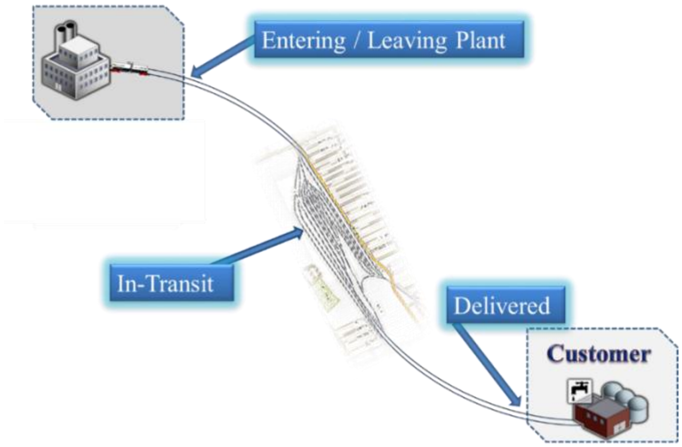

Figure 2.1: Railcar Shipment Overview

Risk domains of safety and security interest are the health and handling of the railcar, its contents, and the railroad infrastructure. A schematic of the developed rail hazmat shipper safety and security system is displayed in Figure 2.2. A variety of operational and spatial information is collected from multiple sources. Natural hazard information is streamed from source data collection agencies to identify extreme weather and earthquake events that may impact a hazmat shipment's current or future location. Sensors installed on the cargo container enable the shipper to monitor any developing problems, recognize tampering, identify an accident, or perform damage assessment. Sensors installed by the railroad (along the track) provide information on equipment and track health. Finally, the shipment is monitored according to its location within a prescribed geo-fence to ensure that it stays on the desired path.

Figure 2.3 shows a schematic of the preferred location of the GPS device mounted on the cargo container. If the unit contains a camera (which is not wide-angle), 14 feet from the platform on the top of the ladder is the desired installation site to capture the image of a seven-foot-tall man standing on the platform. If the device does not contain a camera, the four-foot indicator is the suggested location for a GPS unit install.

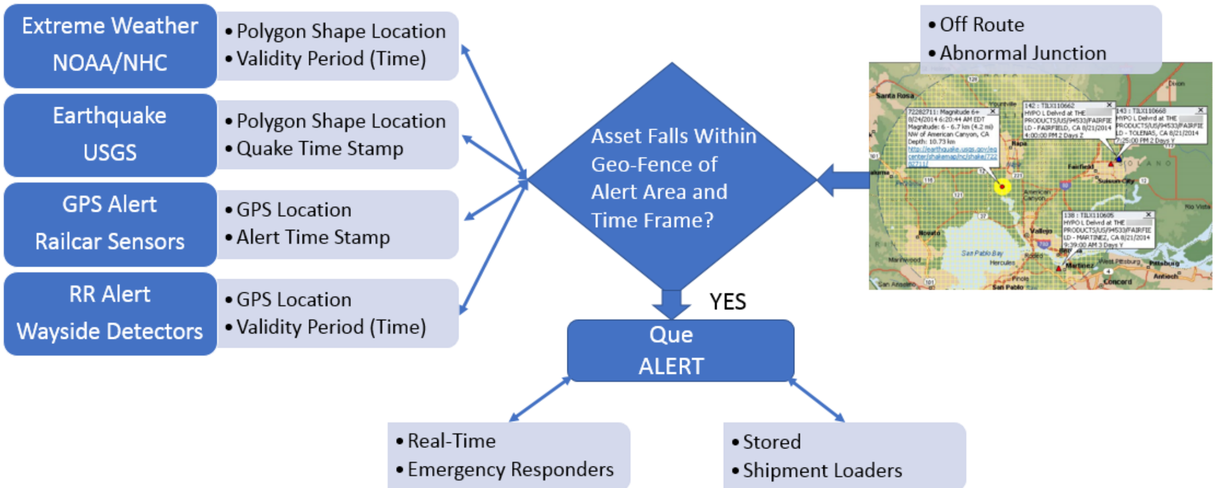

Figure 2.2: Rail Hazmat Shipper Safety and Security Information System

This information is transmitted to the shipper via GPS vendor application programming interface (API) or read from Internet sources and merged into an integrated database for each railcar. This information has three interdependent dimensions - time, location, and event criticality – which are benchmarked against safety/security performance metrics and corresponding safety/security exceedance threshold values. When a threshold is exceeded, an alert or notification is transmitted to the affected parties (Figure 2.4). Alerts are sent in real-time when a critical event has been reported, or the potential for a time-sensitive, critical event is identified. Otherwise, this information is stored as a notification to be acted upon at a more convenient location without jeopardizing immediate safety or security, such as when a part needs to be replaced the next time the railcar arrives at the shipper's facility. In both instances, the relevant information is transmitted to the person(s) with a need to know, formatted to be interoperable with the recipient's device (e.g., smartphone, laptop, etc.).

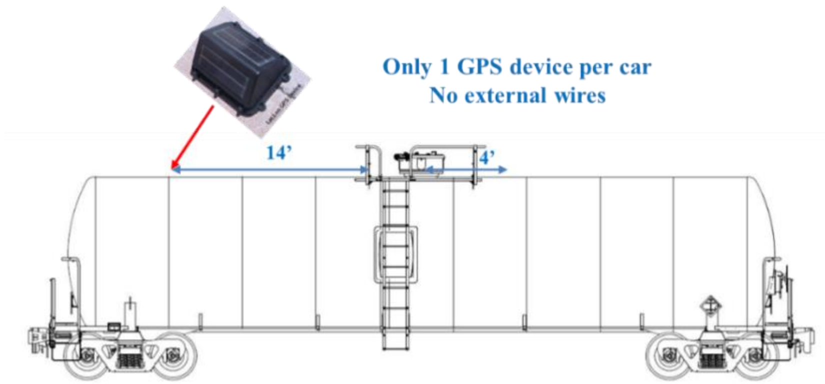

Figure 2.3: Tank Car Equipped GPS Device

Figure 2.4: Alert and Notification Protocol

2.2 System Operations

Specific system operating characteristics can be best described and understood as they map to a typical hazardous material shipment, beginning at the shipment origin.

2.2.1 Shipper Facility System Features

When a railcar enters the shipper's facility, it is scanned in multiple ways (Figure 2.5):

1. The unique railcar automatic equipment identification (AEI) tag is read. This electronic tag allows railroads, car owners, and shippers to track information related to each specific railcar.

Data-tag readers are also usually tied to a railroad's car location message (CLM) system, where the railroad can opt to share this information with its customers. This system provides a trigger to automate the delivery of the alert queue. Alerts gathered during transit are supplied to targeted employees who can act on them.

2. Cameras installed on both sides of the car enable high-definition (HD) image capture. This creates a visual record of the physical status of the railcar upon entering the facility. If any car damage is observed by the technician reviewing the image, this is documented (Figure 2.6) and reported along with the equipment health management (EHM) alerts.
3. Any scheduled maintenance due on the railcar is also identified according to the time/distance it has been in service (Figure 2.7).

This information is collected to facility personnel for appropriate action (Figure 2.8). The system also integrates with the facility entrance security system to determine which personnel are working that day to avoid the delivery of alert queues unnecessarily.

While inside the shipper facility, cars equipped with GPS can generate and provide alerts to facility employees hand-held devices, reporting and reminding them of an unresolved/unacknowledged alert based upon the employee's proximity to the car.

Before departing the shipper facility, maintenance and repair records are checked to ensure that no railcar leaves without work having been completed.

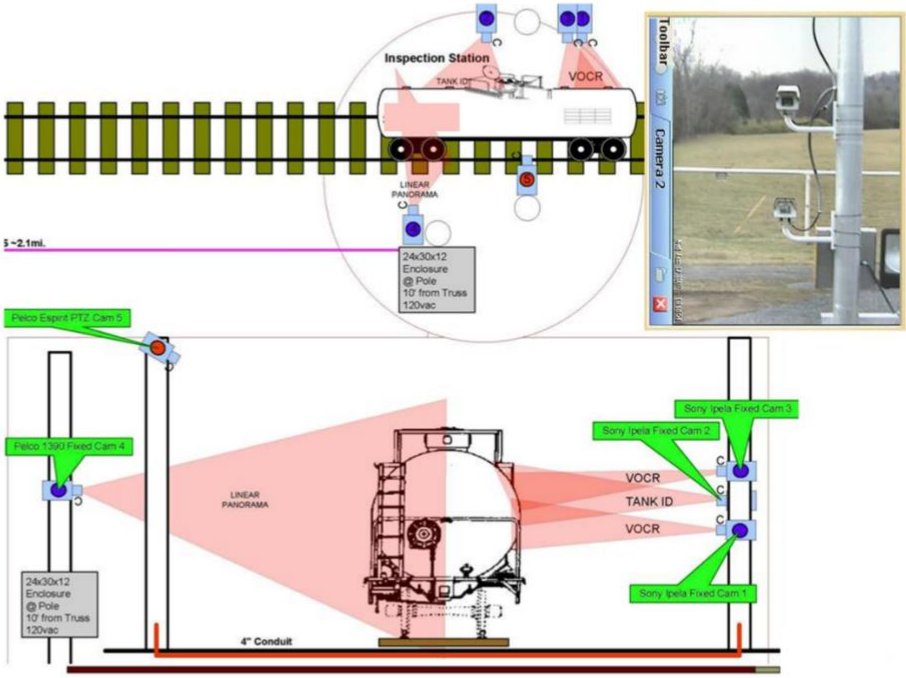

Figure 2.5: Cameras and AEI Readers at the Shipper Facility

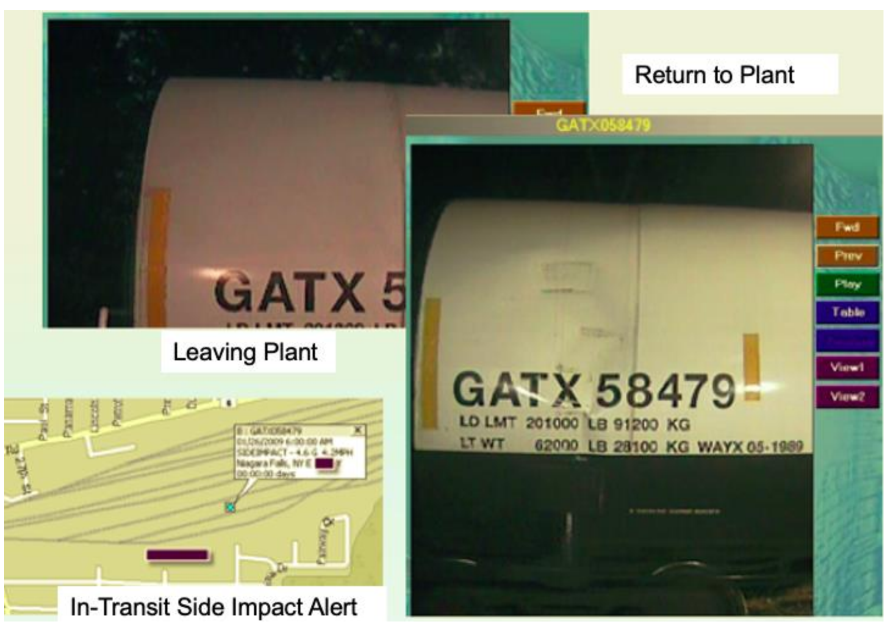

Figure 2.6: Car Damage Assessment

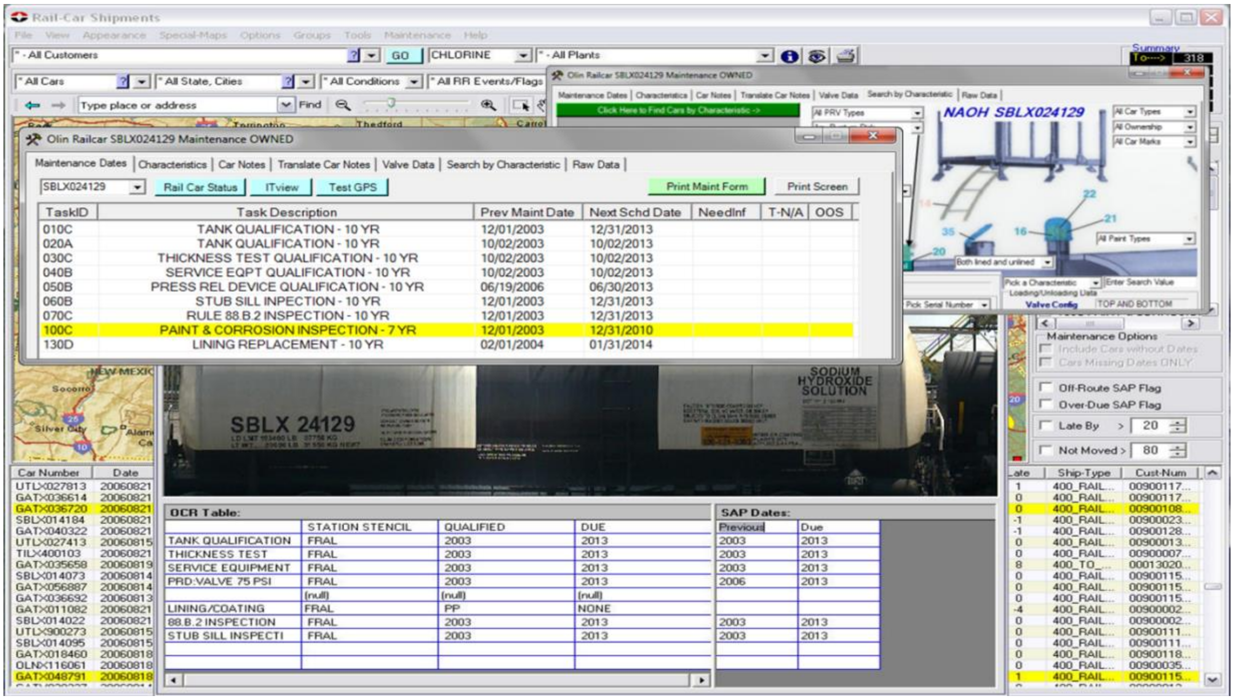

Figure 2.7: Maintenance and Repair Records

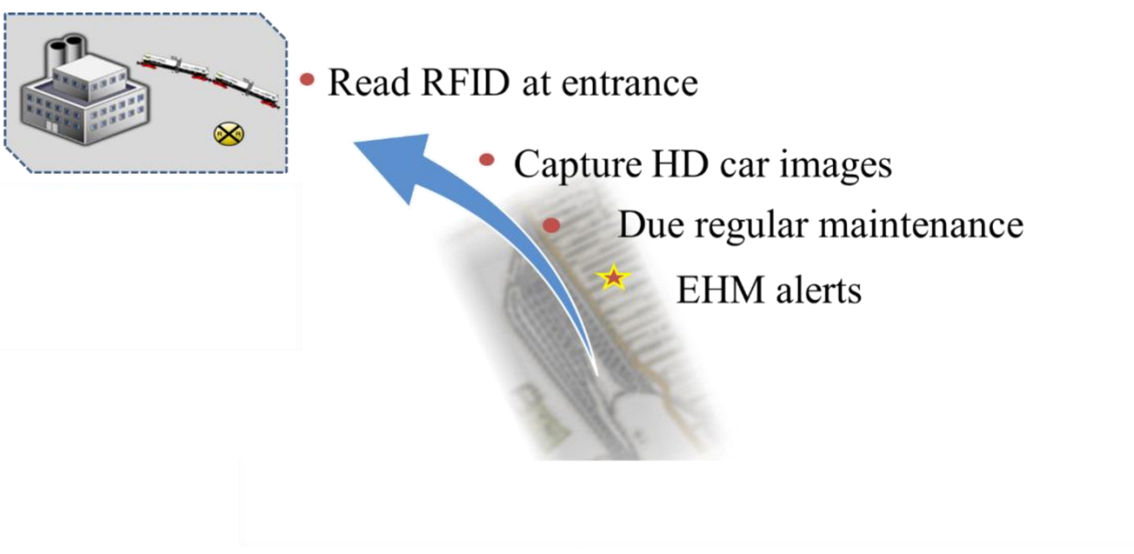

Figure 2.8: Actionable Facility Information

2.2.2 In-Transit

Once cars are in transit and outside the confines of the shipper's facility, the focus shifts to monitoring the shipment, maintaining vigilance that the cargo is not exposed to unwanted safety or security risk.

2.2.2.1 GPS Data Collection

Car safety and security are monitored via the GPS unit installed on the top of the railcar. With this technology, the tank car's dome position can be monitored for tampering, ride quality measured by collecting impact and deceleration data, and images captured when any unexpected motion or movement is experienced (Figure 2.9). The criteria are programmable according to the shipper's specifications.

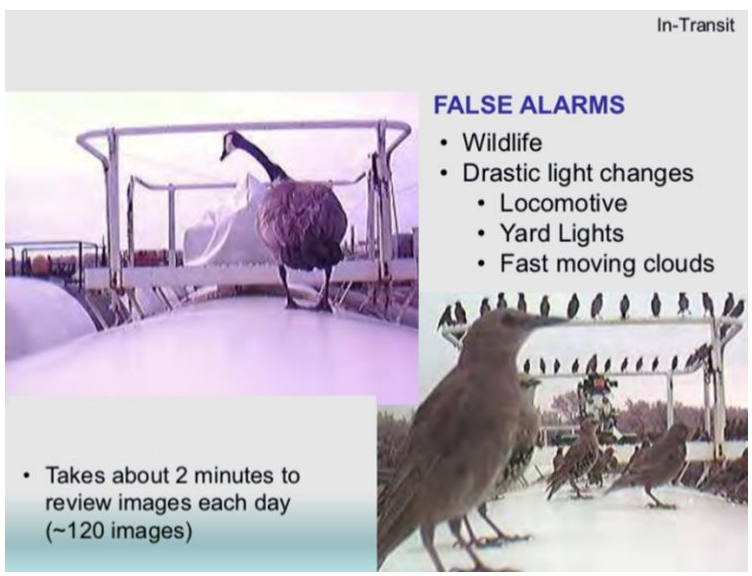

Figure 2.9: Motion Image Capture

Several techniques are available to determine tampering: 1) motion detection sensors, 2) vibration monitoring (differentiates between normal rail car movement vibrations and footsteps on the car), and 3) a wireless transmitter underneath protective dome housings with an accelerometer that detects opening of the dome lid. Regardless of the technique used, an event will trigger image capture. Data is stored by the GPS unit until it is sent to the shipper via cellular message or offloaded via wireless transmission when the car re-enters the shipper's facility.

The GPS devices also serve a valuable purpose in ensuring that the shipment travels on its intended route. Route boundaries can be defined by establishing geofences and spatial coordinates that, if intersected, infer that the railcar is off course. If such an event is detected, the GPS device triggers an alarm indicating an off-route situation.

2.2.2.2 Natural Hazard Information

Natural hazards also pose a potential in-transit threat to the safety of hazmat shipments. Among the data sources regarding natural hazard threats or events, occurrences are available through the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS). This information can be sourced through the web in real-time and associated with railcar location and movement.

Real-time satellite imagery and associated information provided by the National Oceanic and Atmospheric Administration (NOAA), when superimposed on the shipper's railcar locations, offers current information on extreme weather events (e.g., storms, wildfires), such that potentially dangerous areas can be avoided until safe passage can be restored (Figure 2.10). The USGS provides real-time earthquake information, helpful in ascertaining whether a rail car may have incurred damage from being in the impact zone (Figure 2.11) and, if not, whether the railcar's shipment route has been impeded, such that the car must be halted and stored in a safe haven or re-routed.

In the event of a release, existing weather conditions at the site can predict how a release will propagate over time and what emergency response measures would be appropriate (Figure 2.12). Toxic release modeling can be performed using publicly available tools, such as the Areal Location of Hazardous Atmospheres (ALOHA) model, developed by the U.S. Environmental Protection Agency, and the Wireless Information System for Emergency Responders (WISER), developed by the National Institutes of Health. Predictive information provided by these tools helps support timely and effective emergency response.

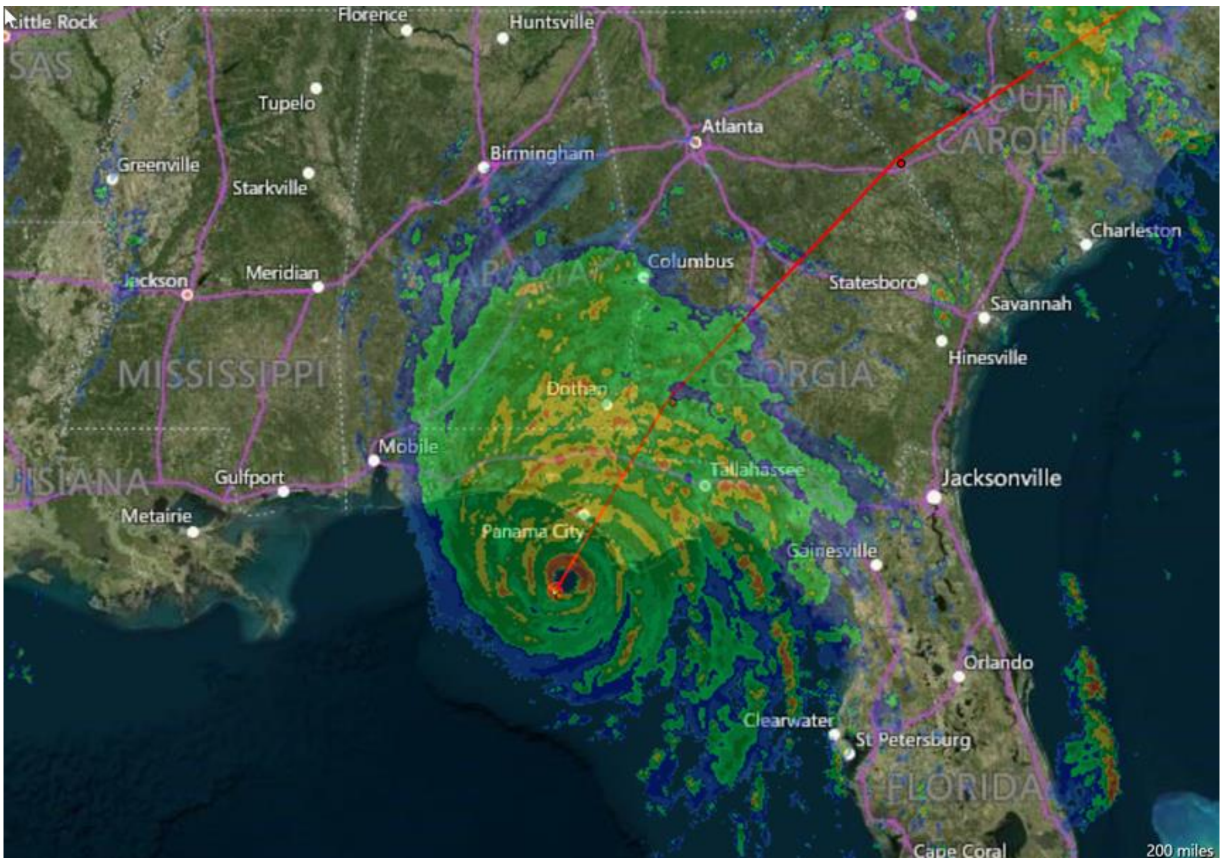

Figure 2.10: Severe Weather Event

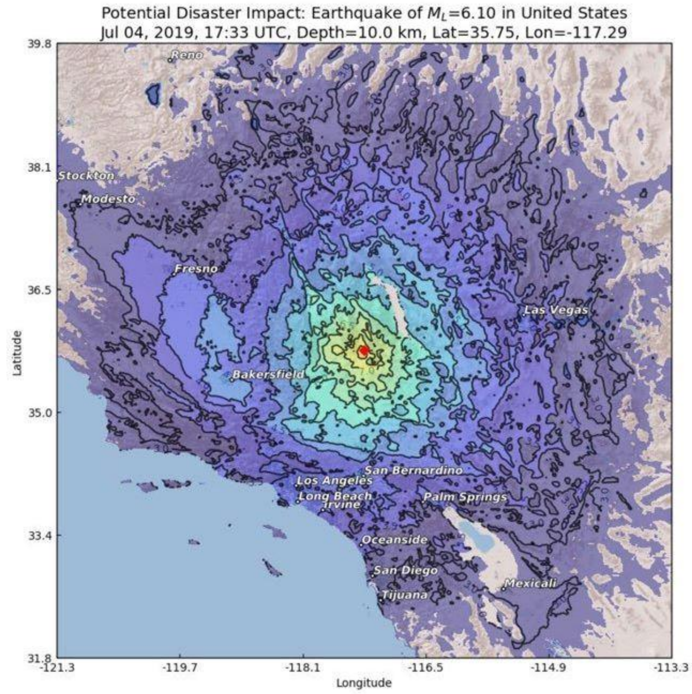

Figure 2.11: Earthquake Impact Zone

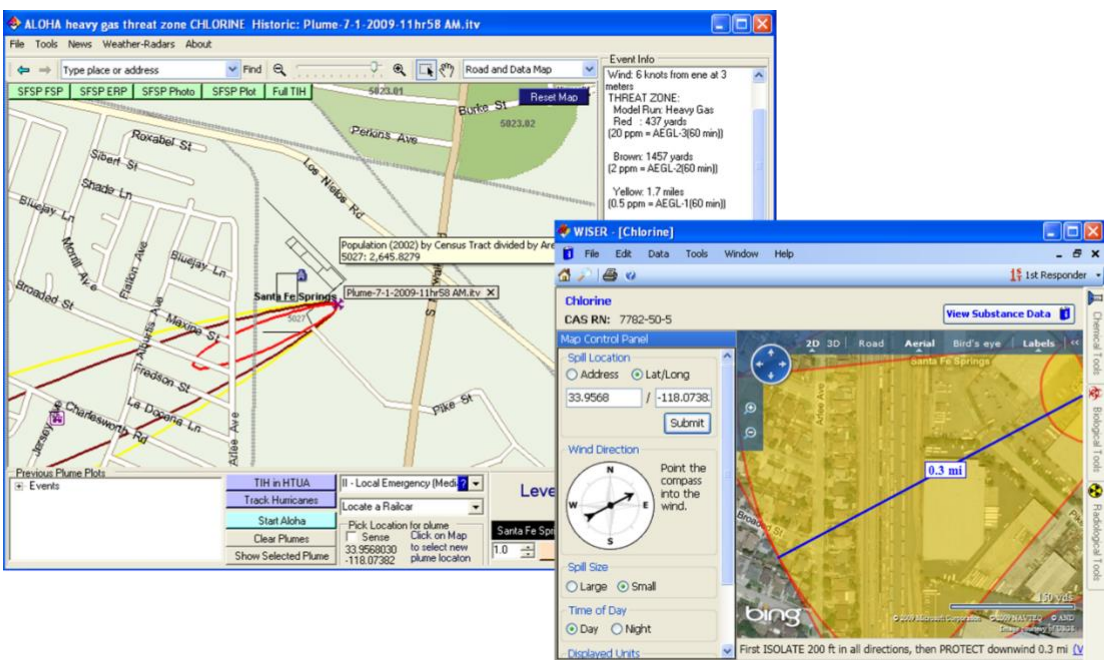

Figure 2.12: Release Modeling

2.2.2.3 Railinc

Another vital information source is Railinc, a subsidiary of the Association of American Railroads (AAR), whose mission is to provide rail data and messaging services to the North American freight railroad industry. Two services provided by Railinc, equipment health management system (EHMS) and car location messaging (CLM), can be integrated as part of a shipper's safety and security information system.

EHMS is a web-based application that communicates the condition of railroad equipment to rail carriers, car owners, and other interested parties, including alerts. The system receives and manages alert data from the following wayside detection devices located at regular intervals along rail track: 1) wheel impact load detectors (WILD), 2) truck hunting detectors (THD), 3) acoustic bearing detectors (ABD), and 4) truck performance detectors (TPD). These detectors are designed to reduce risk in railroad operations by identifying poorly performing equipment before accidents occur. Additionally, line of road failure (LORF) notifications are provided, representing a new statistical alert based on data summaries associated with component failure rates.

WILD measures vertical wheel forces via rail-mounted accelerometers or strain gages, searching for defective wheels. THD measures oscillation of the wheelset due to lateral movement in the track gauge, where exceedance above a certain threshold (i.e., dynamic instability) can cause wheel flanges to impact the rails, potentially causing damage to both and increasing the likelihood of a derailment. ABD identifies bearing flaws in railcars by recording audio from a train as it passes by and using acoustic technology to detect wheel bearing defects before failure. TPD evaluates the suspension performance of trucks[1] by measuring the vertical and lateral forces generated by the wheels as a car moves over the detectors that are placed along the instrumented track. The following defects can be flagged with TPD: 1) worn friction wedges, 2) broken suspension springs, 3) twisted car bodies, 4) mismatched side frames, 5) hollow/worn wheels, and 6) tight side bearings.

[1]In the rail industry, the term "truck" refers to the structure underneath the railcar to which axles (and wheels) are attached.

CLM delivers complete information on the car location and its shipment. CLM has many event codes which help summarize and provide data on the car and its cargo. Examples include departure and arrival times, current location and estimated arrival time, interchanges between railroad carriers, intermodal transfers, and equipment in storage or deemed currently defective.

2.2.3 Customer Delivery

At the destination, once the shipment has arrived inside the customer's fence line, the shipper remains concerned that no accident occurred at this location. The previously described technologies and information system remain active during this time and until the shipper's equipment leave the customer's facility. It is also possible that the technologies described in Section 2.2.1 could be installed at the receiving facility as well.

2.2.4 Information Transmission and Database Management

Data collected by the GPS units can be transmitted using cellular service or via long-range wireless communication. The data is received and stored at a central server maintained at the shipper's facility where cars enter or leave. The data is further enriched by combining it with the information provided by Railinc and natural hazard web services, resulting in comprehensive health, maintenance, inspection, and damage assessment of each rail car entering the facility for further action. The data collected from various sources can be analyzed based on the thresholds and decision criteria established by the shipper, consistent with any Federal Railroad Administration (FRA) requirements and AAR standards/guidance.

The process can be aided considerably by utilizing Early Warning, a web-based application/service provided by Railinc that acts as a hub for communications about rail equipment from which maintenance advisories and early warning notices can be issued. Early Warning enables railroads, equipment owners, and repair shops to have visibility into defective equipment and components, identify when tests are past due, and report when a car has been inspected or repaired so that equipment can be removed from notices.

Every source of data has a relevance period, defined as the time between the ship date and the delivery date. All GPS readings and alerts that occur between these dates belong to that shipment. Events that occur during transit can also have a relevance period and sometimes a geo-coordinate shape, such as a tornado warning. All GPS readings from a car inside the geo-coordinate shape that occurred within the event relevance period can be flagged for inspection once the car arrives back at the shipper facility. Events such as earthquakes may have a short relevance period but, depending upon the magnitude, and regional location, can have a large geofence radius. All GPS readings reported within the geofenced area and timing within one update cycle would be flagged for inspection.

2.3 Case Study: The Olin Experience

Olin Corporation is a large multinational petrochemical company with a substantial North American presence as one of the largest producers of chlorine, industrial bleach, and on-purpose hydrochloric acid. As a shipper of hazardous materials, Olin is committed to ensuring the safety and security of these shipments through active monitoring and active engagement with carriers, incident responders, and other transportation stakeholders as an essential part of its risk management program.

Olin developed and has implemented the aforementioned integrated technology system, equipping a portion of its railcar fleet used for shipping high-hazard cargo with GPS devices and establishing criteria for triggering alerts and notifications. In the discussion to follow, the Olin experience is described in terms of how the system is operated and utilized. Hereafter, we refer to this system using the acronym SHRIS (**S**hipper **H**azmat **R**isk **I**nformation **S**ystem).

2.3.1 Olin System Data Sources

As described in the prior discussion and shown in Figure 2.13, Olin sources data via GPS, the Internet, and Railinc to form a comprehensive information system from which risk-informed safety and security decisions can be made. The GPS unit is utilized for dome open/close detection, motion detection image capture, and collection of impact and deceleration data. Web services are sourced for recognizing natural hazard threats and events, whereas Railinc provides EHM and CLM data.

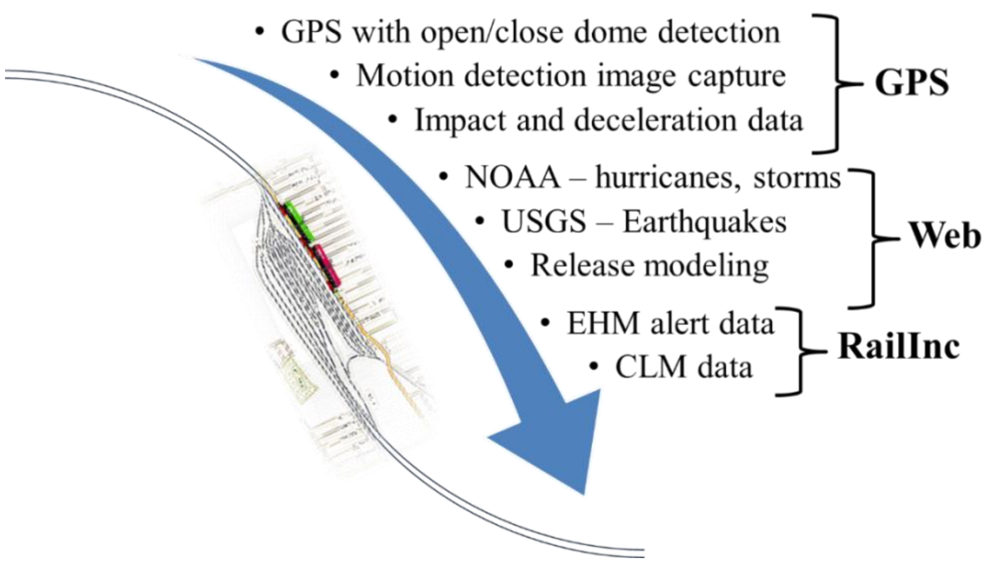

Figure 2.13: Olin System Data Sources

2.3.2 Anomaly Alerts

As shown in Figure 2.14, collectively, this information enables Olin to assess a multitude of system performance considerations and evaluate whether alerts are warranted. Any alerts are recorded and transmitted to the plant rail crew for maintenance and repair (Figure 2.15)[2]. These notifications are managed within an alert management system, wherein the plant crew has complete visibility over each railcar and its corresponding safety and security needs (Figure 2.16).

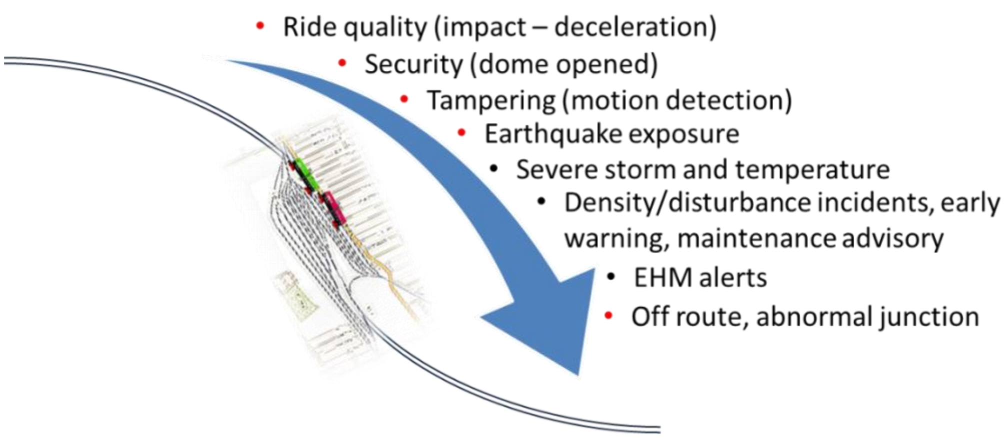

Figure 2.14: Olin System Alerts

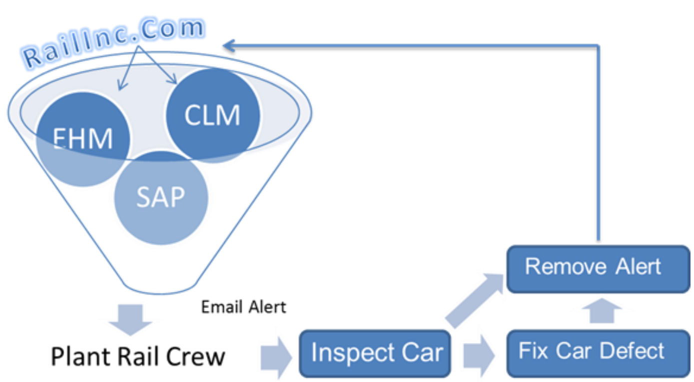

Figure 2.15: Alerts Directed to Olin Rail Crew

[2]The reference to SAP in Figure 2.15 represents Olin internal digital platform.

- Do Not Load
- Bad Order
- Do Not Use
- RR DDCT
- RR EHM (Wild,etc)
- RR EW (ABT,etc)
- GPS Alerts
  - Impacts
  - Speed Drops
- Tracking Alerts
  - Weather
  - Quakes
  - Floods

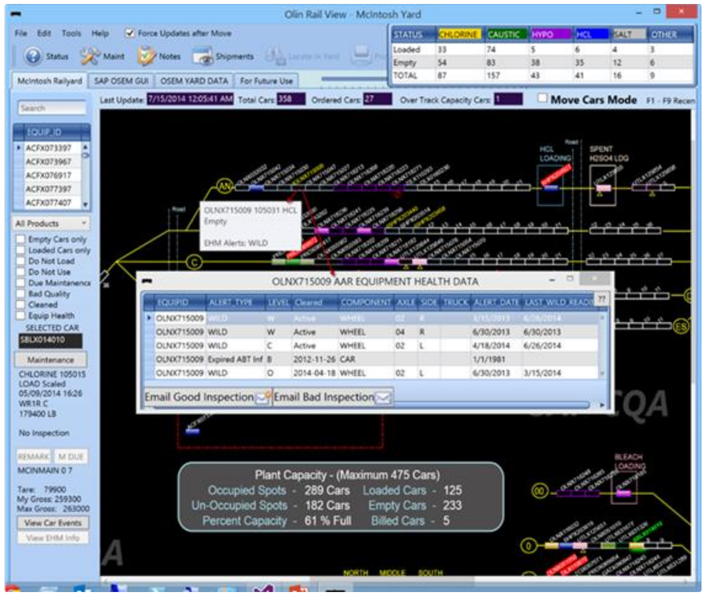

Figure 2.16: Alert Visibility

Another element of the system focuses primarily on rail security. In addition to any immediate emails directed to Olin's security coordinator if an undesirable event is detected, the security coordinator receives a daily summary of each event, which includes images captured from the GPS unit that can be used to verify if the event had any impact on shipment integrity (Figure 2.17). Image capture from the GPS unit when in transit helps eliminate false alarms by providing visual context around the event.

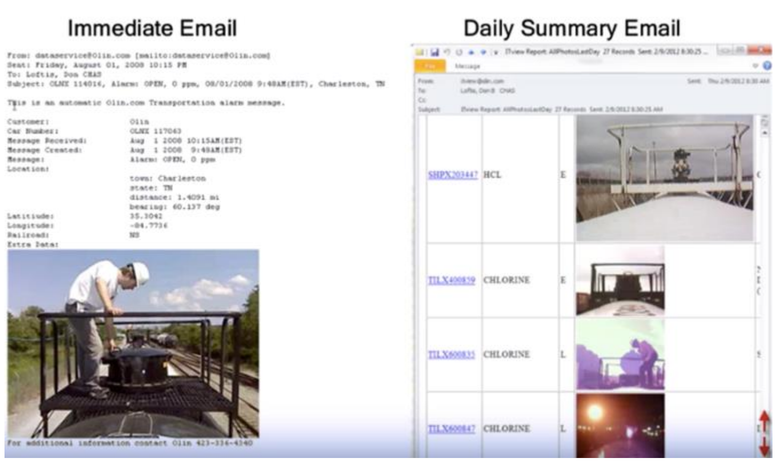

Figure 2.17: Olin Daily Summary Report with GPS Image Capture

2.3.3 Deceleration Threshold Determination

As previously discussed, alerts are sent to plant operations when railcars entering Olin plants have experienced a deceleration event in excess of a target threshold. In an effort to determine where to set the deceleration alert threshold that would trigger visual and mechanical inspection of coupler components (Figure 2.18), Olin conducted a study to determine damage encountered as a function of deceleration speed. The study involved an analysis of 200 tank cars that Olin owns that were tracked and inspected over a 12-month period, using data generated by the GPS-enabled fleet when being moved by Class I and regional railroads and while located in shipping and receiving yards. This amounted to a sample size of 2,660 car bills. All cars were loaded to within 3% of their maximum allowable weights, and decelerations were monitored during the loaded legs of the trip because the higher mass would result in higher forces and more damage. Olin operation inspections and repairs were combined with rail car repair billing data from Railinc (supplying in-transit repair information) to identify any coupler damage incurred by these tank cars during the study period.

Using the maximum coupling speed obtained during each car bill trip, the probability of a damaged component was determined by dividing the number of maximum coupling speeds at each level by the number of bad components found (Figure 2.19). Noting that damage classified is cumulative left to right, one can observe that at a collision velocity of greater than 8 mph, coupling damage becomes a much more frequent consequence. Hence, Olin set its alert threshold at 7+ mph, meaning that railcar alerts are sent to plant operations when railcars entering the plant have experienced a deceleration event in excess of the reporting threshold.

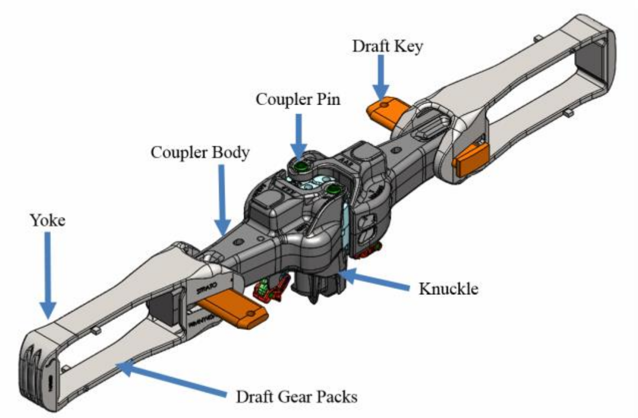

Figure 2.18: Railcar Coupler Configuration

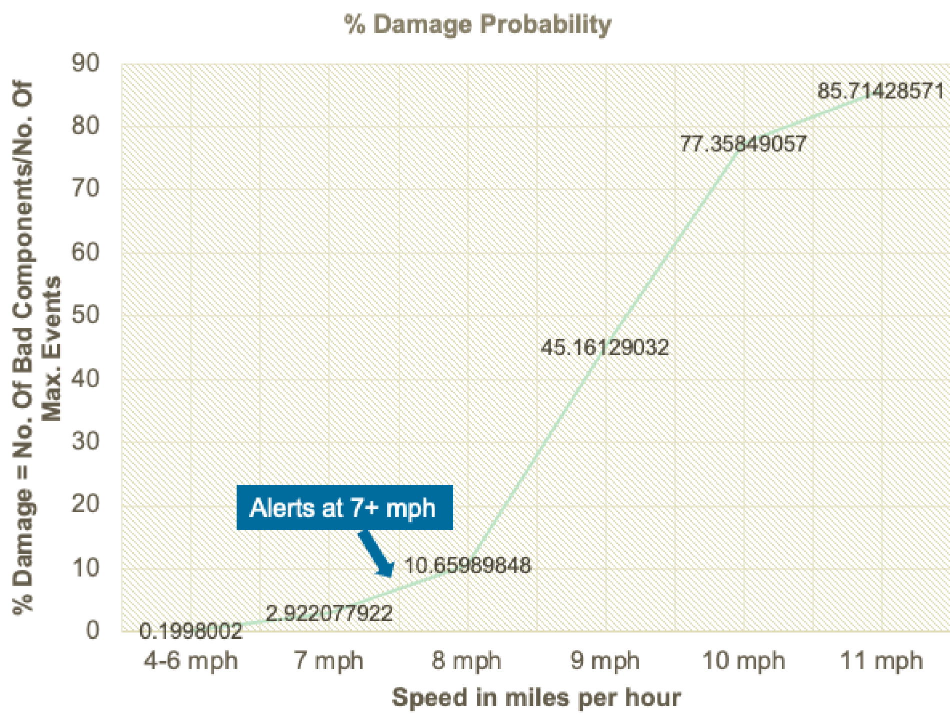

Figure 2.19: Coupler Damage Probability Versus Maximum Deceleration Speed

To adequately capture deceleration speeds, the location of the GPS unit on top of the railcar is an important consideration, as forces are concentrated, especially in an undamped system, and they may not propagate fully down to the GPS location. Since a car can be struck on either end, ideally having the GPS on the center of the car would be best to ensure impact is detected equally from both ends. In consultation with Lat-Lon, Olin's GPS vendor, a decision was made to position the unit off-center, enabling the device camera to obtain a wide angle of view such that the platform and ladder can be clearly seen. Placing the unit outside the platform area also reduces the potential for device damage and tampering.

As the GPS units contain an accelerometer that is capable of generating 10 Hz and 100 Hz filtered values, this provides an additional method for determining (and verifying) deceleration speed.[3] The accelerometer-based approach is capable of detecting short duration, high impact events that last less than 50 milliseconds. This shock pulse type event is measured by subtracting the 10Hz reading from the 100Hz reading, producing a G-force lateral impact. To accommodate this consideration, Olin's GPS devices are programmed to report any impact forces of larger than 2G that are accompanied by deceleration events in excess of 3 mph. An important consideration, however, is recognition of false alarms. For that reason, the GPS unit is prompted to take a picture whenever a deceleration event greater than 5 mph or an impact over 5G is recorded.

2.3.4 Monitoring Derailment Incidents

Derailments are another Olin trigger event requiring an alert notification. Figure 2.20 shows a derailment event on a train carrying an Olin shipment, which prompted an alert to be generated within three minutes of the event occurrence. Note that the derailment location is identified along with an image captured by the GPS unit. This information has been supplemented by data recorded on the change in velocity leading up to the incident. From the captured image, it can be seen that the trees are horizontal, so the car is definitely on its side. In Figure 2.21, images are shown

[3]Filtering is used to remove structural vibrations or ringing from the signal.

following re-positioning of the railcar on the track. The image shows that the dome lid is open, but also indicates that no product is leaking (opening the dome lid provides access to the valves, but does not expose the product unless the valves are compromised). Significant damage to the railcar skin is also visible.

Figure 2.20: Olin Derailment Event

Impact Picture

After Re-Track

Figure 2.21: Dome Lid Open

2.3.5 Natural Hazard Event Tracking

Olin's Crisis Management System includes a real-time feed that provides up-to-date information on severe weather as well as seismic activity. Figure 2.22 displays information the company utilized after Hurricane Matthew had made landfall. By overlaying the hurricane footprint on Olin railcar assets (dots show railcar locations; colors represent different hazmat products), a determination could be made as to how to manage the safety of these assets. Seismic activity potentially impacting Olin's shipments are shown in Figure 2.23 for the La Habra earthquake that struck southern California. The location of railcars in vicinity of the earthquake was identified, from which a decision could be made as to whether any damage inspection was warranted.

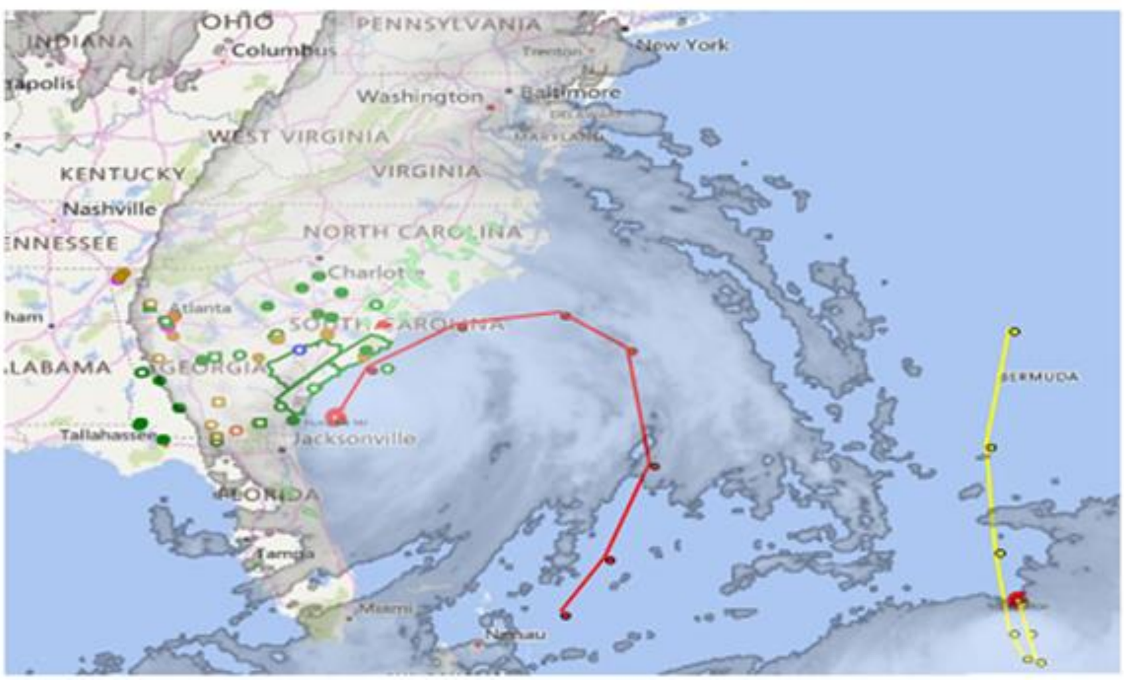

Figure 2.22: Severe Weather Tracking: Hurricane Matthew

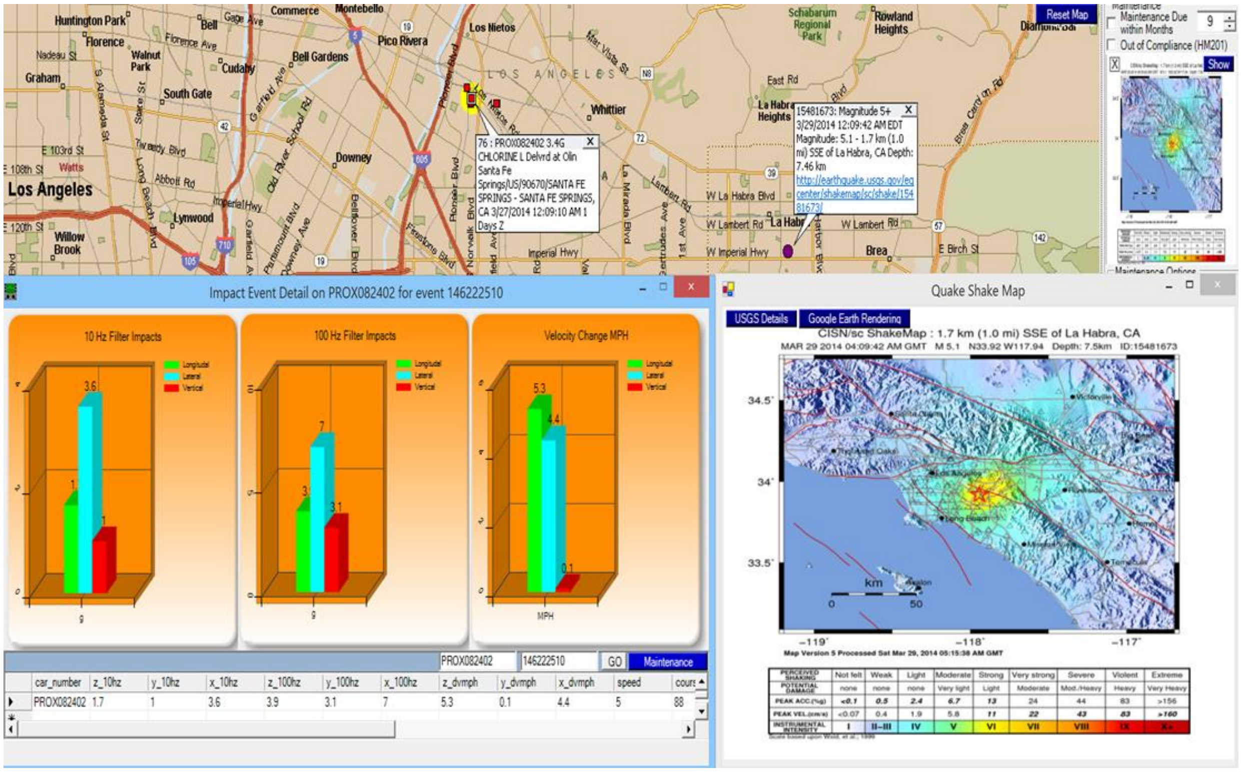

Figure 2.23: La Habra Earthquake

2.3.6 Google Earth Flight Verification

An additional activity that Olin deploys is to utilize Google Earth flights to generate a movie that provides a bird's eye view of the trip made by a particular railcar (Figure 2.24). This includes demarcating the location of any alerts that were issued during the trip wh, ich, when combined with other relevant data, can create a more comprehensive profile of the event in question (Figure 2.25).

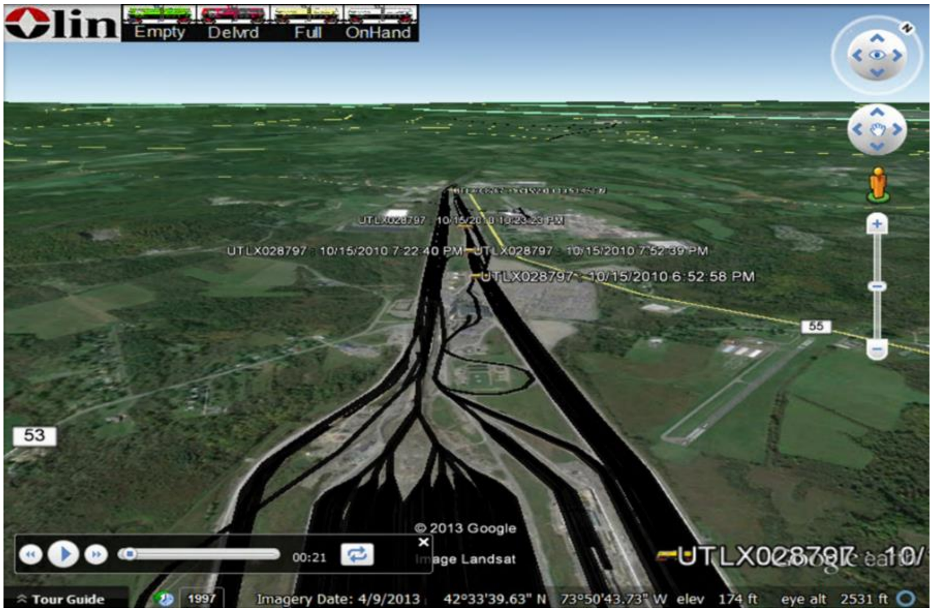

Figure 2.24: Trip Imagery via Google Earth Flight

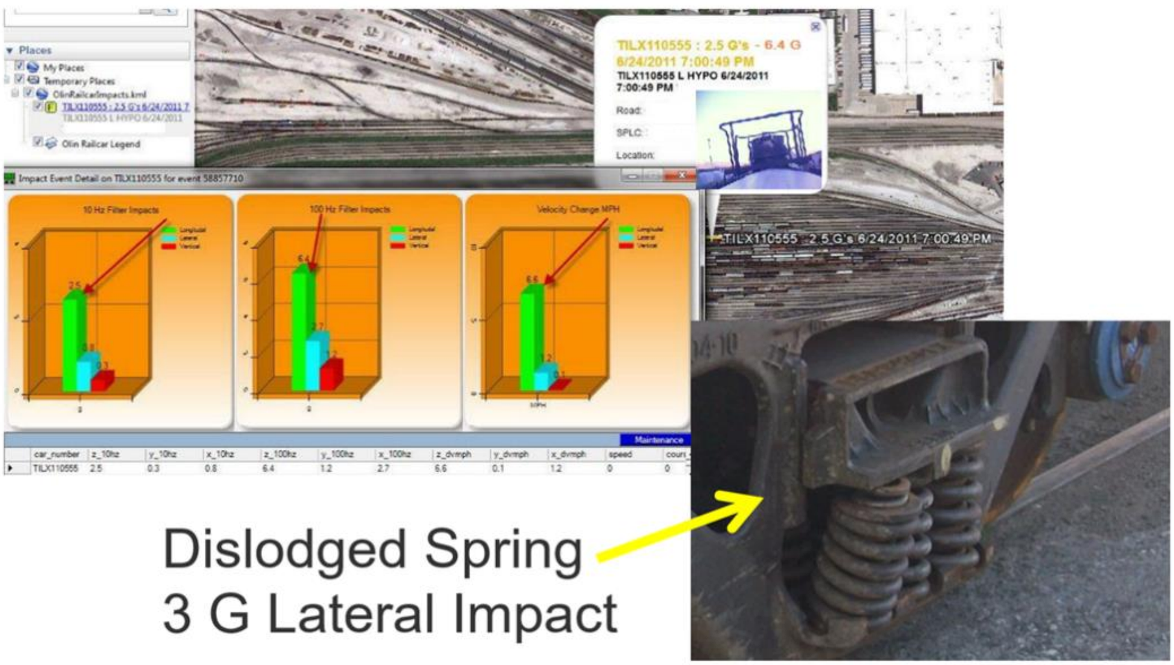

Figure 2.25: Google Earth Flight Events and Images

2.3.7 System Impacts

Since implementing the Integrated Technology System for Rail Shipper Safety & Security, Olin has experienced a dramatic reduction in the risk of transporting hazardous materials and realizing cost savings. The discussion below provides examples of this achievement.

2.3.7.1 Condemnable Wheels

Figure 2.26 shows the impact on the weekly average of condemnable wheels during the past several years, coinciding with when the integrated rail safety and security system was in place. Although there are aberrations in week-to-week performance, there is a clear general trend of a substantial reduction in the number of condemnable wheels related to Olin railcars. This significant improvement in fleet health has several important implications: 1) reduces the costs associated with installing new wheelsets, 2) helps to flatten the maintenance budget, 3) provides the ability to focus better and streamline inspections, and 4) finds and repairs dangerous trucks before they can cause a severe incident.

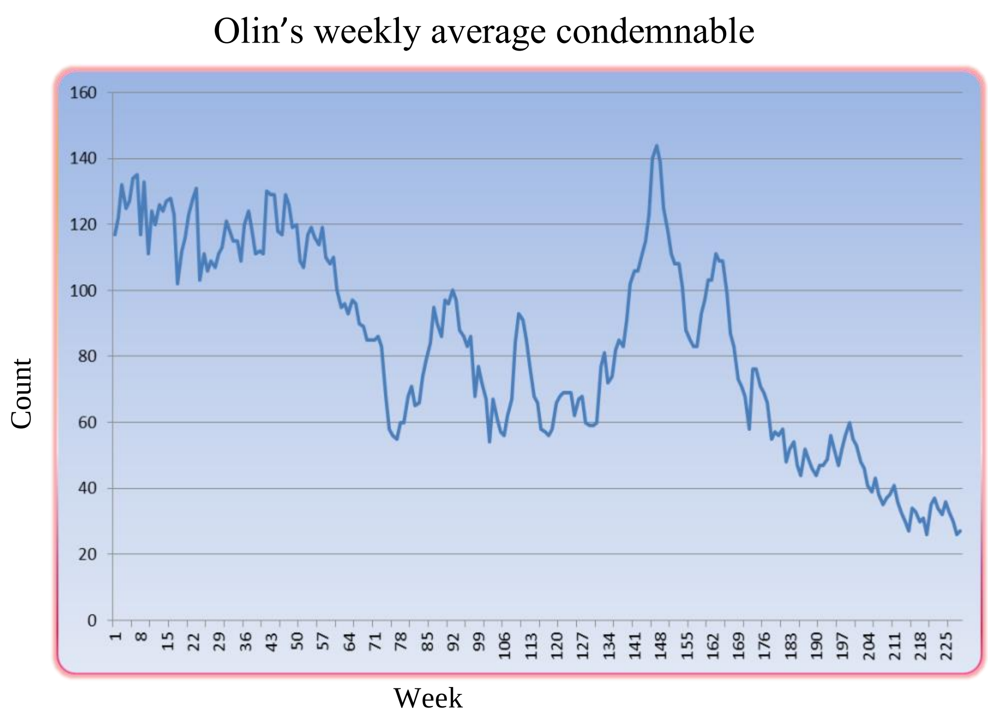

Figure 2.26: Olin Weekly Average Condemnable Wheels

2.3.7.2 Other Considerations

The ability to collect and archive such a vast array of railcar performance data, augmented by data analytics to investigate specific rail safety and security issues, has enabled Olin to utilize technology adoption proactively. For example, the company could compare safety records at various classification yards and notify corresponding railroads where sub-standard handling of Olin railcars is being observed. A related consideration is evaluating flat yard versus hump yard safety performance, intending to work with the rail carriers to route Olin shipments to reduce the risk of railcar damage based on analysis results.

2.4 Conclusion

The adoption of intelligent detection systems and related communication technology affords an opportunity to further improve the safety and security of rail shipments of hazardous materials, with an eye on both incident prevention and consequence mitigation. The SHRIS system described herein leverages these technology advancements, with the hazmat shipper as the driver of this process. Providing the shipper with these capabilities is key to influencing the entire hazmat transportation supply chain, given their knowledge of the product and their relationship with rail carriers and customers.

A distinct project objective is to transfer system knowledge to enable other hazmat shippers who use the rail mode to leverage its availability. A proven and affordable system can enable hazmat rail shippers of all sizes to leverage this capability, not just a select few within the industry. Such widespread adoption benefits not only each shipper but the industry as a whole. While developing a SHRIS-type system for hazmat movements by either the truck or barge mode is likely to offer safety and security benefits, a greater need and opportunity appear to exist for developing and deploying such a system in the barge domain. This rationale rests with the communication technology gap found in the maritime industry created by the dependence on a paper-based system kept on board each vessel. Further, the significantly larger cargo volumes per barge shipment create the potential for a more consequential impact in the event of a material release.

Chapter 3

Factors Impacting Bike Crash Severity in Urban Areas

3.1 Introduction

Bicycling represents a relatively small portion of the total commuting activity in the United States (US). Still, this non-motorized travel mode plays a vital role in many of the nation's urban areas. Biking is a relevant part of many emerging integrated transportation demand management systems. It offers a sustainable mobility option with a lower carbon footprint should commuters choose to switch from modes that rely on traditional fuel sources. This has prompted several state and local agencies to promote biking by employing strategies such as sidewalk modifications and the construction of dedicated bike lanes. In recent years, many cities with bicycle sharing programs have also increased dramatically. However, these developments have resulted in an increase in bike crashes, many with incapacitating injuries or fatal outcomes. Therefore, it is essential to improve our understanding of the critical factors impacting bike crashes in urban areas, aiming to develop risk mitigation strategies to curb this trend. This chapter discusses an analysis performed with this intent.

The study objective is to determine bike road safety in select urban areas within the State of Tennessee using detailed crash data to investigate the factors affecting bike crashes with incapacitating and fatal outcomes and subsequently develop a classification model for fatal or incapacitating events. It concludes with a policy discussion directed at enhancements to transportation infrastructure and operations with bicycle safety in mind.

3.2 Background

Commuting on a bicycle is the third most utilized US transportation mode and is quickly gaining popularity as a commuting option. The number of commuters biking to work has increased by 65%

nationwide from 2000 to 2019 (U.S. Census Bureau, 2014 & 2021). Unfortunately, with increased usage, there are also alarming trends involving fatal and incapacitating bicycle crashes. Traffic hazards for bicyclists include poorly designed roads, high motor vehicle speeds, and lack of responsibility exhibited by other road users (Furth et al., 2016; Jacobsen and Rutter, 2017). In 2019, bikers accounted for 0.5% of 156 million commuters; however, of all traffic crashes, bikers account for 1.78% of injury crashes and 2.3% of fatal and serious (incapacitating) injuries for the entire nation. Fatal and serious bike injuries have seen a 36% increase since 2010 (NHTSA, DOT HS 813 197: Traffic Safety Facts 2019), indicating bicyclists were among the most vulnerable users being disproportionately impacted (Jacobsen and Rutter, 2017; Smart Growth America, 2020).

The disturbingly high number of crashes involving bicycles resulting in fatal or incapacitating injury outcomes leads one to question whether the transportation infrastructure and operations lack accessible and safe facilities for bikers which can be problematic when bikers must share roads with other users, particularly motor vehicles. Bicyclists are considered among the most vulnerable participants in mixed traffic because of the kinetic energy produced upon crashes between two differential masses. One is traveling at a higher velocity and mass (Jacobsen and Rutter, 2017). In the case of an automobile colliding with a cyclist, speeds above 20 miles per hour increase the risk of severe road injury or fatality (Jacobsen and Rutter, 2017; Chris Jurewicz et al., 2016). Therefore, heavily utilized urban corridors impose a potentially significant danger to cyclists if not provided with adequate safety measures (NTSB, 2019).

3.3 Literature Review

Bicycle crashes have been studied by researchers worldwide. Many of these efforts have been directed at individual areas or regions to identify and rectify safety issues within the bicycle infrastructure and operations. The most common modeling techniques have included the use of the Poisson distribution, negative binomial models, linear regression models, logit models, ordered probit models, and multivariable logistic regression. Table 3.1 lists the results of significant factors found in previous studies, organized according to field type and variable. Table 3.2 summarizes relevant study methodologies.

About Table 3.2, note that the use of random forest modeling is not included. Studies modeling bicycle injury prediction using random forest are currently in their infancy, such as one examining bicyclist-only crashes in Victoria, BC, Canada; however, the dataset consists of only 111 crashes and 234 near misses and was collected via surveys rather than from official crash records.

Table 3.1: Significant Bicycle Crash Factors from Prior Studies

Variable Field	**Variable Analyzed**	**Relevant Studies**
Environmental	Lighting Weather Intersection Type Speed Limit Traffic Control Device Number of Lanes Road Curvature Traffic Volume (AADT) Land Use (urban, rural, residential, industry, farmland, institutional, commercial)	Zangenehpour et al., 2016 Yan et al., 2011 Klop et al., 1999 Allen-Munley et al., 2004 Strauss et al., 2015 Reynolds et al., 2009 Turner et al., 2011 Lee and Abdel-Aty, 2005 Petritsch et al., 2006 Pai, 2011 Schepers and den Brinker, 2011 Dixon et al., 2012 Kim et al., 2007 Eluru et al., 2008 Oh et al., 2008 Vandenbulcke et al., 2014
Crash Specific	Crash Type Severity	Wang et al., 2015 Klop et al., 1999 Allen-Munley et al., 2004
Time	Year Month Day Hour	Wang et al., 2015

Table 3.2: Previous Methodologies and Modeling Techniques

Modeling Technique	**Author**	**Study Focus**	**Variables Analyzed**
Poisson Distribution	Oh et al., 2008	Bicycle Crash at Urban Signalized Intersections	Average daily traffic volume, presence of bus stops, sidewalk widths, number of driveways, presence of speed restrict devices, and presence of crosswalks are all statistically significant risk factors.
Negative Binomial Model	Oh et al., 2008	Bicycle Crash at Urban Signalized Intersections	Found different types of facility designs impact bicycle safety such as bike lanes, bike track, pavement markings or colors.
	Wang et al., 2004	Bicycle - Motor Vehicle Crashes at a Signalized Intersection	Intersection design impacts on bicycle safety in multiple ways.
Linear Regression	Dixon et al., 2012	State Highways	For intersection and network movement, hazardous crossings, right hook, left sneak and complicated interactions are potentially dangerous to bicyclists. Intersection safety influenced by vehicle volume, vehicle speed, percentage of heavy vehicles, among others.
Logit Model	Eluru et al., 2008	Road Segments	Crashes on curved/non-flat roadways tend to result in more severe injuries.
	Kim et al., 2007	Bicycle-Motor Vehicle Crashes	Curved rounds significantly increase the chance that a fatal or incapacitating injury will occur during a vehicle-bicycle crash.
	Pai, 2011	Road Segments	Horizontal and vertical curves can contribute to bicycle crashes.
	Schepers & Brinker, 2011	Road Segments	Bicyclists colliding with a bollard, road narrowing or riding off a curve found to occur more than when bicyclists hit an obstacle. More crashes were observed where the bicycle had the right-of-way on a through movement at intersections with two-way bicycle tracks that are well marked and are reddish in color. Fewer crashes occurred when there are raised bicycle crossings (speed humps) or other speed reduction measures.
	Abdel-Aty and Keller, 2005	Signalized Intersections	The division of a minor road, as well as a higher speed limit on the minor road lowered the expected injury level, while a median on the minor road may prevent more head-on crashes, which were found to be more severe crashes.
	Haleem and Abdel-Aty, 2010	Unsignalized Intersections	Traffic volume on the major approach, number of through lanes on the minor approach, upstream and downstream distance to the nearest signalized intersection, left and right shoulder width, number of left-turn movements on the minor approach, and number of right- and left-turn lanes on the major approach are significant factors influencing bicycle risk.
Decision Tree	Rahman, 2018	Pedestrian & Bicycle Crashes	Highlighted the most significant predictor variables for pedestrian and bicycle crash count in terms of three broad categories: traffic, roadway, and socio demographic characteristics

Modeling Technique	Author	Study Focus	Variables Analyzed
Bayesian Model	Vandenbulcke et al., 2014	Selected Controlled Sites or Bikeable Road Network	Right-of-way intersections equipped with bicycle lanes tend to have higher crash risk for cyclists, due to vehicles not respecting the right-of-way (i.e., right-hook crashes). Cyclists riding on marked bicycle lanes in roundabouts and signalized intersections with marked cycle lanes had higher crash risk, attributed to bicyclists being in drivers' blind spots. Additionally, complex intersections (high number of road legs, road users, high number of signs, dense traffic crossings, etc.), and therefore complex traffic situations, increase bicycle risk.
Safety Analyst and Clustering Algorithm	Dolatsara, 2014	Roadway Segments in Michigan	Exposure, the presence of bicycle lanes and bus stops, and the number of left-turn lanes at intersections are positively associated with bicycle crashes.

3.4 Data Analysis

The bicycle crash data utilized in this analysis was obtained from the Tennessee Department of Transportation (TDOT) for the period of January 1, 2017 through December 31, 2020, covering the entire state. In Tennessee, a crash is reported when a driver of a vehicle is involved in a crash resulting in injury, death or property damage exceeding $50 (Tennessee Code Title 55. Motor and Other Vehicles § 55-10-106). A crash is also reported when a vehicle collides with an unattended vehicle (Tennessee Code Title 55. Motor and Other Vehicles § 55-10-104), such as one located in a parking lot. Crash data obtained from TDOT and used for this study consists of only bicyclist-motor vehicle crashes. Attributes associated with each crash record are listed in Appendix I.

During this period, 5,347 bike crashes were recorded for which there was complete information (see Table 3.3), distributed across the state, as shown in Figure 3.1. Of the ninety-five counties in Tennessee (TN), Shelby County and Davidson County recorded the highest bike crashes, collectively accounting for 2,942 incidents, more than one-half of the overall state total. This is to be expected since these two counties are densely populated and include the cities of Memphis and Nashville, respectively. As a result, these two locations subsequently became the focus of the modeling effort.

Table 3.3: TN Bike Crashes by Year

Year	Total Bike Crashes
2017	1,384
2018	1,299
2019	1,432
2020	1,232

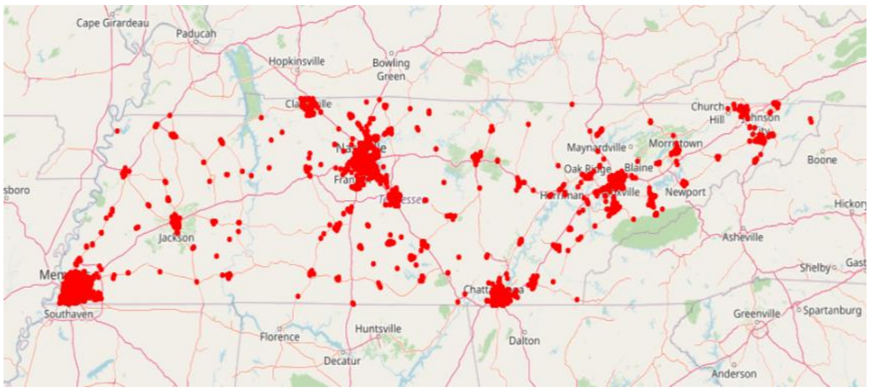

Figure 3.1: Bicycle Crashes Cluster in TN (2017-2020)

Bike crash severity results for the two counties are shown in Figure 3.2. TDOT crash data includes an injury severity attribute according to whether there was no injury, non-serious injury, serious (incapacitating) injury or fatality. An incapacitating injury is one which results in one or more of the following: 1) severe laceration resulting in exposure of underlying tissues/muscle/organs or resulting in significant loss of blood, 2) broken or distorted extremity (arm or leg), 3) crush injuries, 4) suspected skull, chest or abdominal injury other than bruises or minor lacerations, 5) significant burns (second and third degree burns over 10% or more of the body), 6) unconsciousness when taken from the crash scene, and 7) paralysis. Particularly notable is that fatal and incapacitating injury collisions account for 27% of 2,942 recorded crashes.

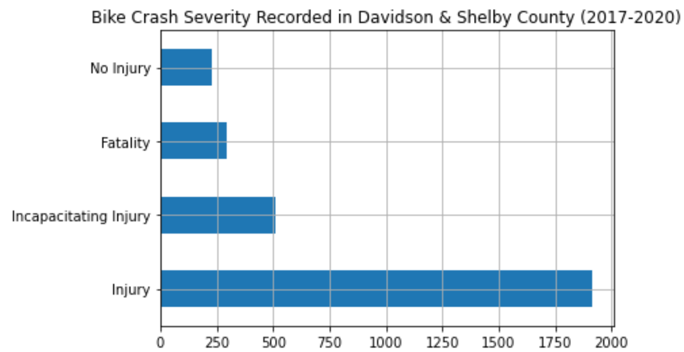

Figure 3.2: Bike Crash Severity in Davidson and Shelby County

Figure 3.3 shows the distribution of bike crashes by time of day, where times have been grouped into the following categories: 1) midnight-5:00 am, 2) 5:00-9:00 am, 3) 9:00 am-1:00 pm, 4) 1:00-5:00 pm, 5) 5:00-9:00 pm, and 6) 9:00 pm-midnight. Note that while the frequency of bicycle crashes tends to increase as the day goes along, the percentage of those that result in incapacitating and fatal injuries are highest during the earlier part of the day.

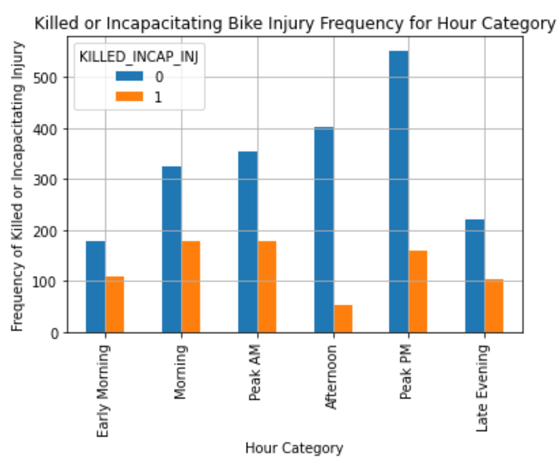

Figure 3.3: Bike Crashes by Time of Day in Davidson & Shelby County

As displayed in Figure 3.4, the largest number of bike crashes in general as well as those resulting in a fatality or incapacitating injury occurred on four lane roads, two lanes in each direction. It was also observed that four lane roads experience a high number of bicycle injuries on medians and in turn lanes.

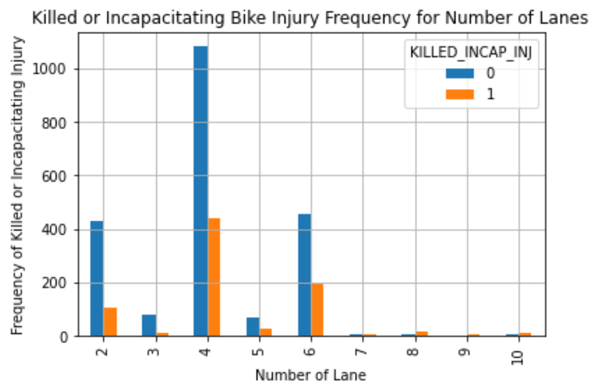

Figure 3.4: Bike Crashes by Lane Configuration

As seen in Figure 3.5, roads with speed limits from 30 mph to 45 mph experience a significant number of bicycle crashes, with the proportion of those resulting in a fatality or incapacitating injury increasing at higher speeds. This observation is consistent with prior studies (Irene Isaksson-Hellman et al, 2019).

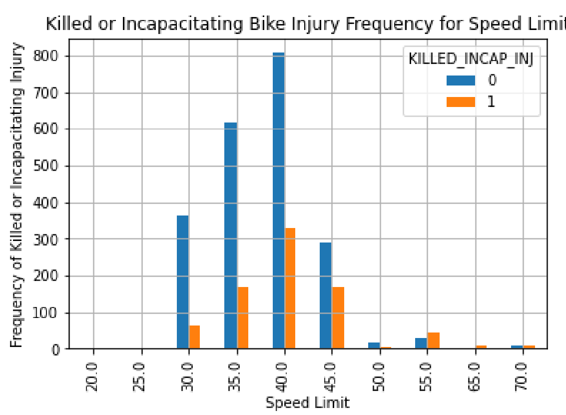

Figure 3.5: Bike Crashes by Road Speed Limit

3.5 Modeling Approach

While the literature review cited a variety of modeling approaches that have been developed for predicting bicycle crash severity, none have considered a comparison of various classification predictive modeling techniques with different balancing methods along with a rigorous feature selection process. An overview of the process used in developing a predictive model of bicycle crash severity is shown in Figure 3.6. Model estimation was performed using Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF). A classification technique is applied to the models such that one can predict the categorical outcome of a killed or/and incapacitating injury (Class 1) and no killed and/or incapacitating injury (Class 0). LR is a supervised machine learning algorithm that uses a logistic function to model the outcome and serves as a baseline for our binary classification problem. It represents a widely used method to study risk factors impacting injury severity. The advantage of LR is its easy implementation and quick description of the relationship between the input variables and the output variable with no scaling of features. The drawback with LR is that it can only construct linear boundaries, it assumes no correlation between input variables and the output variables, input variables are correlated to each other, and a constant need to set the threshold (from the baseline of 0.5) on which classification is based such that we reduce the false prediction of the output variable.

DT is another frequently used supervised learning classification algorithm for understanding and interpreting data, where the top node is the root node, representing the best feature that divides the data. Each internal node is a feature and branches indicate the decision, with the class label being represented by a leaf node. A DT consists of nested if-else statements where successive conditions are checked unless a conclusion is reached (i.e., a decision is made if the output will be a class 1 or class 0 only if it satisfies certain criteria for each of the features), which can then be shown graphically in the form of a decision tree or a flow chart. DT outperforms an LR, especially when the relationship between the input and out variables is complex and non-linear. DT also helps build easy-to-understand models for visualization; however, DTs tend to overfit. DT serves as a foundation for RF, which is yet another supervised machine learning algorithm.

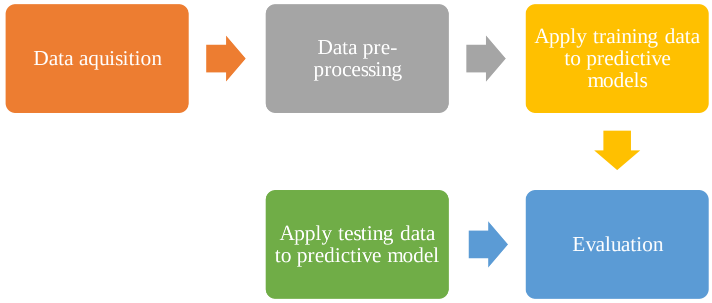

Figure 3.6: Predictive Model Development Framework

Although RF has not been extensively used as a classification algorithm for analyzing bicycle crashes, it was included because RF has been shown to improve modeling performance relative to a single tree classifier (e.g., DT) and LR. RF enables multiple uncorrelated DTs to grow, thus creating a forest. RF uses a technique called feature bagging, where features are selected randomly for individual DTs, which is similar to bagging procedure. With feature bagging, the correlation between each DT is reduced but the overall accuracy of the model increases. RF performs better compared to LR and DT as it is more robust to noise, and able to capture the non-linear tendencies by putting all the weak learners in an ensemble that is used to make the prediction. It also avoids overfitting because those individual learners are weak, so it is not one massive model that could lead to overfitting the data (A.C. Muller et.al.,2017).

In this study, we elected to use LR followed by DT and RF to observe the model prediction outcome. It is not necessary to use models that build on the previous ones; however, this was done to tune the classifier and improve model performance.

The dependent variable was defined as a numerical Boolean variable, with a value of 1 indicating a fatal or incapacitating injury outcome, and 0 otherwise (i.e., minor injury or no injury). Prior to conducting model estimation, data pre-processing was performed to remove records with missing data, following which exploratory data analysis was performed. This resulted in the selection of the following candidate crash factors (attributes) to be considered as independent variables in model estimation: location, functional class, number of lanes, speed limit, average annual daily traffic (AADT), impaired driver, weather, lighting and weekend. Categorical values for location (roadway, intersection, bridge, ramp), functional class (urban, rural), impaired driver (yes, no), weather (clear, cloudy, rain, fog, snow, severe cross wind, sleet, hail), lighting (dark, dawn, daylight, dusk) and weekend (yes, no) were converted to numerical Boolean variables (0 or 1). AADT, speed limit and number of lanes, were scaled to help decrease the magnitude as per a fixed ratio; this process assists with reducing fluctuations in model performance.

The data set was divided where 80% of the observations were used for training and the remaining 20% for testing. We attempted to balance the training data before model insertion. Note that as shown in Figure 3.7, the dependent variable is unevenly distributed in the training dataset, with 27% of bike crashes resulting in a fatality and/or incapacitating injury (i.e., minority class), and 73% of bike crashes resulting in no fatality and/or incapacitating injury (i.e., majority class). There are several techniques to handle imbalanced datasets, but broadly-speaking, data can be balanced by decreasing the majority class sample size (under-sampling) or increasing the minority class sample size (oversampling). We will consider two widely used algorithms for under and oversampling (i.e., Near-Miss and SMOTE). Additionally, we look at misclassification costs as a way to address imbalanced classification.

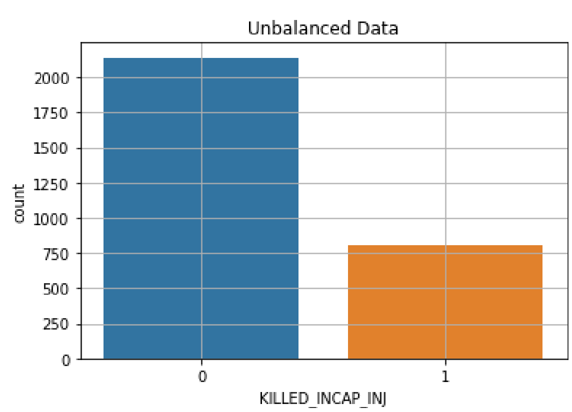

Figure 3.7: Unbalanced Data for Dependent Variable

Three sampling techniques were applied to training data as part of the modeling effort to gauge which method gave the best prediction capability. We used the NearMiss algorithm for under-sampling to prevent the problem of information loss in most traditional under-sampling techniques. Here, the majority class is reduced to the total number of the minority class. In a near-miss algorithm, distance between all the points in the majority and minority class is calculated; for all instances where this distance is the shortest, this group of points in the majority class is selected for elimination. For each example in the minority class, a given number of the closest majority class is selected. This method guarantees that every minority point is surrounded by some majority samples. Synthetic Minority Oversampling Technique (SMOTE) was applied for oversampling, a technique where synthetic samples are generated for the minority class that helps to overcome the overfitting problem posed by traditional random oversampling techniques. By linear interpolation of the minority class, synthetically more training data is generated by randomly selecting one or more of the k-nearest neighbors for each minority class by calculating the Euclidean distance between a point and every other sample point in the minority class. Another balancing approach involves the use of cost-sensitive learning (CSL), whereby a larger weight is assigned to the minority class and a smaller weight is applied to the majority class. Since the data points for a killed and/or incapacitating injury are way smaller compared to injured and/or no injured cases, and this chapter focuses on being able to detect killed and/or incapacitating injury, CSL can be used in especially such cases. In CSL, each class is given a misclassification cost when training a model, where the aim is to minimize the total misclassification cost. When the class weight is set according to the imbalanced ratio, it implies a modification in the loss function, thus improving the training model by pushing the decision boundary that allows improvement in the minority class.

Even though RF is insensitive to many features, we applied the following feature selection methods to improve the classification performance of LR and DT so that we can examine and compare the models unbiasedly. As the dataset consists of both numerical and categorical inputs, three methods of feature selection were applied sequentially: 1) Correlation coefficient, 2) DT feature importance, and 3) Recursive Feature Elimination (RFE). Feature selection using correlation is a filter approach that is based on the feature-to-feature correlation using the standard Pearson correlation coefficient value. The goal here is to find a subset of features that are highly correlated with one another and drop them as they may influence the performance of the model performance. A correlation coefficient threshold of ±0.7 was applied to eliminate highly correlated features (Hulse, J.V. et al., 2012). The second step of feature selection involved the use of DT after dropping the highly correlated features. By using all the features in a DT, we can quickly observe the portion of the features DT uses for the full classification. We drop the features which do not contribute to the classification. The final filter selection method used was RFE, which works by starting with all the features in the training dataset and subsequently removing the undesired features until a subset of the desired features remains. The RFE starting point was the set of features filtered using DT in the training dataset. The core of the model used here is DT, where features are ranked by importance, the least important features are discarded, and the model is refitted. This process is repeated until only the desired features remain by performing a cross-validation evaluation of the different number of features and selecting the number of features with the best mean score.

Finally, to understand and explain the output for a killed and/or incapacitating bicycle injury (class 1) for the selected model, Shapley additive explanations (SHAP) is used. SHAP helps interpret the predictions by measuring each feature's contribution (known as Shapley value) to the output (class

1). Shapley values are a concept adopted from the game theory field, whose objective is to measure each player's contribution to the game (Shapley, 1953). We used Kernel Shap to calculate the Shapley value as it can interpret any ML model regardless of its nature. Kernel Shap is based on weighted linear regression where the coefficients of the solution are the Shapley value (Lundberg et. al., 2017).

Some limitations were identified in the crash data. The data used for this study consists of only bicyclist-motor vehicle crashes. All the data is recorded at the scene of crash by law enforcement officers. Once this police report is filed, it is then entered into the data platform. Hence, this data can suffer from human error when reported, collected and processed at the various stages. TDOT does not include near misses and unreported bike incidents. The data does not provide information on the cause of crash, which party involved in the crash was injured (although we assume that if any injuries are reported, it at a minimum involves a bicyclist), nor any details on how or in which direction the involved parties where moving (i.e., circumstances prior to crash).

3.6 Model Results

The overall model performance measure is the extent to which the model can accurately predict whether a bike crash results in a fatality or incapacitating injury. Appendix II provides a list of relevant metrics and their corresponding definitions for indicators considered in evaluating the efficacy of model performance.

Table 4 summarizes the performance metrics for various models that were estimated using the features that emerged from the aforementioned elimination process: lighting (dark), number of lanes, speed limit, AADT, weekend, and location type (roadway).

As shown in Table 3.4, three models (oversampled LR, weighted CSL applied to both LR and RF) perform well. However, weighted CSL applied to RF performs slightly better, due to its higher true negativity rate (0.63) and true positivity rate (0.77), and with lower Type I and Type II errors. Moreover, RF with weighted CSL has the highest value of G-mean (0.7), and weighted accuracy (0.7).

The Receiver Operating Characteristic (ROC) curve value for RF with weighted CSL is also high for the testing data (0.7) and varies the least (0.01) from the training data (0.71). This curve plots two parameters: true positive rate (TPR) and false positive rate (FPR). This measure is derived from a curve plotted on a graph showing the performance of a classification model at different classification thresholds The ROC curve can help identify the threshold by balancing the TPR and FPR than manually checking which threshold works best. A cut-off point of 0.5 is taken for ROC, which means that below this value the model is unable to distinguish between class 1 and class 0. In RF, you obtain the probability of the prediction belonging to a class when you aggregate the indication functions from its decision trees. When you do the inference on the train and test dataset, you get a distribution, and the ROC curve represents precision of the chosen point of the corresponding probability space. The ROC measures the area under the curve; when the ROC is closer to 1 but greater than 0.5, it indicates a strong model.

Table 3.4: Performance Metrics for Various Model Estimation Techniques

Performance Metrics	True Negative Rate	True Positive Rate	False Negative Rate	False Positive Rate	Geometric -Mean	Weighted Accuracy	Receiver Operating Characteristics - Train	Receiver Operating Characteristics - Test
LR – Unbalanced	0.18	0.94	0.062	0.82	0.41	0.56	0.58	0.56
LR – Undersample	0.57	0.68	0.32	0.43	0.62	0.625	0.64	0.63

Performance Metrics	True Negative Rate	True Positive Rate	False Negative Rate	False Positive Rate	Geometric -Mean	Weighted Accuracy	Receiver Operating Characteristics - Train	Receiver Operating Characteristics - Test
LR – Oversample	0.71	0.68	0.32	0.29	0.69	0.695	0.67	0.7
LR – Weighted CSL	0.68	0.69	0.31	0.32	0.68	0.685	0.68	0.68
DT– Unbalanced	0.31	0.89	0.11	0.69	0.53	0.6	0.68	0.6
DT– Undersample	0.62	0.66	0.34	0.38	0.64	0.64	0.69	0.64
DT – Oversample	0.5	0.78	0.22	0.5	0.62	0.64	0.75	0.64
DT – Weighted CSL	0.66	0.7	0.3	0.34	0.68	0.68	0.71	0.68
RF – Unbalanced	0.2	0.95	0.055	0.8	0.44	0.575	0.6	0.57
RF – Undersample	0.63	0.62	0.38	0.37	0.62	0.625	0.7	0.63
RF – Oversample	0.56	0.76	0.24	0.44	0.65	0.66	0.77	0.66
RF – Weighted CSL	0.63	0.77	0.23	0.37	0.70	0.7	0.71	0.70

Shapley additive explanations (SHAP), as shown in Figure 3.8, measure the contribution of a feature in model prediction (Apley and Zhu, 2020). Note that both classes use the same feature equally (i.e., all features have equal impact on model prediction). Among these features, dark lighting and roadway crash location are the most important factors affecting bike crash severity, while roads with higher motor vehicle speed limits, heavy traffic, multilane roads and weekend travel are also significant contributors.

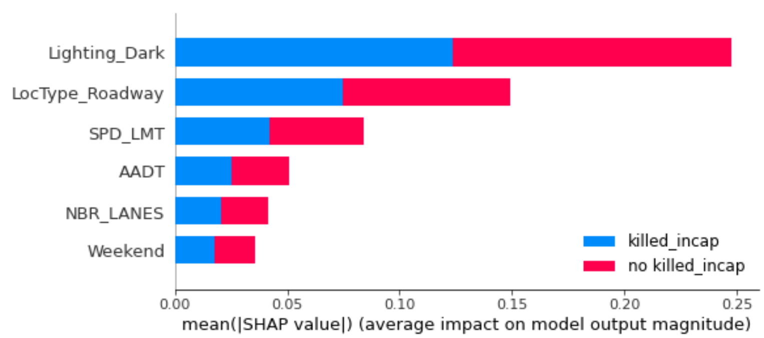

Figure 3.8: Summary Plot Displaying SHAP Values for Model Features

Figure 3.9 displays a bee swarm plot for the study data. This plot helps one understand how a variable may influence model prediction. In this plot, every record in the database is shown as a dot on each row. The color of the dot represents the value of that feature for the event, with red indicating a high value and blue a low value. Here, one can observe that for Class 1 (killed and/or incapacitating bike injury), when the lighting condition is inadequate and location type is roadway, it is more likely to result in a killed and/or incapacitating bike injury.

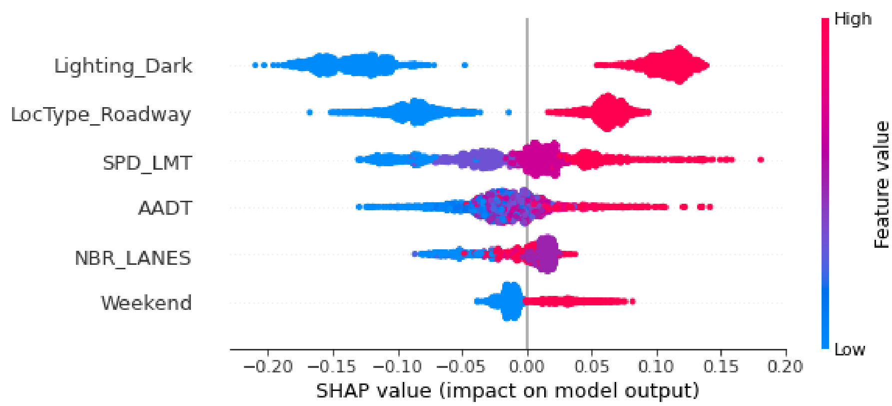

Figure 3.9: Summary Plot Combining Feature Importance with Feature Effect for Class 1 (Killed and/or Incapacitating Bike Injury)

Understanding prediction for individual instances can provide meaningful information, as it explains how individual predictions are reached in terms of feature contribution. To illustrate, we selected this information for two bicycle crash records, one which resulted in a killed or incapacitating injury (Event 419), and another where the outcome was not a killed or incapacitating injury (Event 422). Using the feature inputs for Event 419, the model predicts a killed and/or incapacitating bike injury with 0.71 probability. This compares with when we do not know any features for a specific event, in which case the average model output over the training dataset is 0.4995 (base value). Doing the same for Event 422, In the case of Event 422, the model predicts a no killed and/or incapacitating bike injury with 0.74 probability, compared to a base value of 0.5005.

Figure 3.10 helps identify groups of similar instances by using hierarchical agglomerative clustering to order the instances. Each position on the x-axis is an event in the database, where red plots increase the model prediction and blue decreases it. A cluster is observed towards the right of the curve with high prediction of killed and/or incapacitating bike injury.

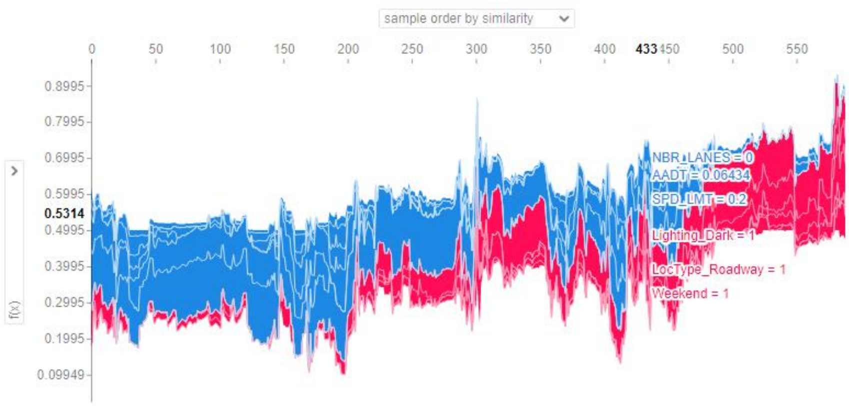

Figure 3.10: Clustering Based on Features for Class 1 (Killed and/or Incapacitating Bike Injury)

The heavy influence of inadequate lighting conditions on bike crash severity is a finding consistent with prior studies and is the largest factor influencing bicycle injury severity (Asgarzadeh et al., 2018), with Kim et al. (2007) concluding that the probability of a fatal bike injury doubles in the absence of streetlighting. The magnitude of this factor in the model results suggests that risk mitigation strategies should seriously consider improvements to lighting infrastructure.

Many previous studies have focused on crashes along the intersections since they have the highest conflict points. However, within our study database, more than one-half of the bike crashes occurred on non-intersection segments, one reason why this feature emerged as a significant explanatory factor for serious injury outcomes. Asgarzadeh et al. (2017) similarly found these locations to be important, reporting that crashes on non-intersection segments are more likely to result in 1.31 times higher injury severity.

An increase in speed limit is also positively associated with a killed and/or incapacitating biker injury outcome. Chimba et al. (2012) noted a similar relationship when comparing crashes on roads with speed limits of 30 mph to those with a 35-45 mph speed limit. Fridman et al. (2020) describe several case studies which significantly reduce the likelihood of a killed and/or incapacitating bike (and pedestrian) injury by lowering road speed to 20 mph.

As a larger number of motor vehicles (AADT) travel across a road segment, it creates greater opportunity for crash exposure. Therefore, it is not surprising that biker injury frequency would increase; however, it is less clear based on model results that AADT alone accounts for more severe injury outcomes. This may be explained by the presence of other related factors such as vehicle speeds and number of lanes.

The same can be said for the significance of number of roadway lanes as an explanatory factor in predicting a serious biker injury outcome. In fact, the combination of multilane roads with higher speed limits being associated with higher risk of fatal or incapacitating injuries is one of the most consistent findings across the literature (Chen, 2015; Siddiqui et al.,2012; Huang et al.,2010; Lee et al., 2015; Noland and Quddus, 2004; Quddus, 2015; Wier et al., 2009; Yu and Zhu, 2015).

Finally, the weekend effect, albeit small, influences the likelihood of a bike crash causing a killed and/or incapacitating injury. Similar findings were observed in research performed by Shubo Wu et al., (2021).

In reviewing these findings, it is important to acknowledge the potential differences between factor correlation and causation, particularly absent any information on traffic volume to normalize the results (Von Stülpnagel, et. al., 2021). Consequently, one must be careful in interpreting how to associate these results with potential risk mitigation considerations. It is entirely possible, for example, that bicyclist fatalities and incapacitating injuries are actually occurring more often in well-lit places, not because the individual safety risk is greater, but because the volume of bicyclists is so much higher that it confounds this relationship. This issue is addressed in Chapter 5.

3.7 Policy Implications

The feature importance associated with the selected model (see Figure 3.8) provides insights into key factors that most influence serious biker safety outcomes as well as their relative contribution to those impacts. The following discussion provides a general description of policies that may be cost-effective in reducing serious biker safety risk based on the model results. However, the extent to which a particular strategy makes sense is dependent on the site-specific conditions that are present at the location of interest. For example, implementation of a dedicated bike lane may be possible in one location that would be physically infeasible at another site or the benefit-cost may not be sufficient to justify allocation of construction resources.

It is also possible that the unintended consequences of implementing a supposed safety improvement actually creates greater risk. For example, consider a case where a dedicated bike lane is established by demarcating the lane by painting the roadway, but it is relatively narrow in width and a physical barrier is not present to divide it from the rest of the roadway. The existence of a bike lane may motivate more bicyclists to use the facility because of a perception safety has been improved, yet they may be placing themselves more in harm’s way (Ferenchak and Marshall, 2016; Ferenchak and Marshall, 2019).

Regarding lighting conditions, relatively simple risk mitigation strategies would include the placement of street lighting along popular bike routes. In addition to improving illumination by providing better street lighting, it would also help if the bikers made their presence known on a roadway by wearing reflective materials and installing blinking lights on their bikes (Abdur et al, 2021). The latter is required by the traffic law in Tennessee (TN), especially at nighttime. It goes without saying that personal protective wear, which includes helmets, should always be worn by the bicyclist.

While the relationship between roadways and serious bike crash outcomes is clear, the particular built environment and usage may influence exposure; hence, the reason why higher AADT's and number of lanes also contribute to the problem. Controlling for bicyclist exposure, Kaplan and Prato (2015) concluded that separated bicycle facilities reduce both bicyclist injury crashes and fatal crashes, whereas on-street bike lanes do not. This suggests that efforts to create dedicated bikeways which are physically separated from the roadway would be a more effective, albeit a more expensive, risk management strategy. In the absence of resources to provide these means, creating sufficient street width for an on-street bike lane is paramount, as most bicycle lanes today are placed between the vehicular route and the curb, often at widths of no more than four feet (including the 1–2 feet gutter pan as part of the bicycle lane). This problem is compounded by motorist expectations that bicyclists will remain in their dedicated lane, even when physically unable to do so. It is further exacerbated by the presence of "mixing zones", which are placed in advance of right-turn lanes to allow vehicles to cross the bicycle lane to enter the right-turn lane. When combined with adequate signage and other demarcations, these intervention strategies should help alleviate at least some crashes and reduce the impact of others when they occur.

Regarding speed limits, we recommend reviewing all urban streets with speed limits above 30 mph to assess whether the limit should be lowered. When this is not deemed a viable strategy, signage with dynamic message boards could be placed at vulnerable locations, reminding motorists to obey speed limits. Another strategy would be to deploy speed sensors coupled with speed cameras (either mobile or fixed) at vulnerable locations that display the actual speed of a passing vehicle which flashes when the speed limit is being exceeded. Speed bumps and round abouts are other options to slow vehicular traffic speed along the roadway and at intersections.

While recommendations for improving bike safety are encoded into bicycle design guidance (American Association of State Highway and Transportation Officials, 2014; National Association of City Transportation Officials, 2014), the widespread use of bike lanes generally, and mixing zones in particular, has been cited as an example of broader professional ignorance on matters of traffic safety (Hauer, 2016). There are recent and ongoing efforts to better understand bicyclist safety, including NCHRP 17-84: Pedestrian and Bicycle Safety Performance Functions for the Highway Safety Manual, NCHRP 15-73: Design Options to Reduce Turning Motor Vehicle – Bicycle Conflicts at Controlled Intersections and NCHRP 15-74: Safety Evaluation of On-Street Bicycle Facility Design Features. While a lack of crash and exposure data continues to be a hindrance to bike safety research, it has generally been accepted that as the biker population increases, the crash rate decreases (Elvik, 2009), perhaps an indication of greater awareness on the part of motorists of the need to share the road with this travel mode.

To that end, both Sweden and the Netherlands have developed approaches to address this challenge. Starting in the early 1990s, Sweden's Vision Zero and the Netherlands' Sustainable Safety Vision have integrated motorists and vulnerable road users with the concept of shared road responsibility to create homogeneous, multimodal transportation networks (Welle, Sharpin, et al.

2018, Wegman, et al. 2006). The same concept has been recently adopted in Davidson County, TN (Vision Zero, 2020).

3.8 Conclusion

Bike safety has been a much-discussed topic, particularly of late, as interest in bicycling as a sustainable transportation alternative continues to gain popularity. Consequently, policy analysts and planners have been grappling with cost-effective methods to reduce bicycle crashes, particularly those with serious outcomes. We believe that the results of this study have shed additional light on the subject, in particular: 1) demonstrating the use of random forest modeling and select sampling techniques as having the potential to provide greater accuracy in predicting the likelihood of a fatal and serious bike injury, and 2) utilizing the feature weighting of the predictive model to prioritize the types of risk mitigation strategies that offer the greatest impact.

Chapter 4

Explanatory Analysis of Pedestrian Crash Severity in Urban Areas

4.1 Introduction

Walking is an integral part of active transportation since everyone is a pedestrian at one time or another. This is particularly true in urban areas, where walkability is a primary mode used to satisfy mobility needs, in part to avoid traffic congestion, but also as a sustainable alternative and one that can improve human wellbeing, both physically and emotionally. These benefits have prompted federal, state and local agencies to encourage walking by designing more pedestrian-friendly streets. However, pedestrian crash frequency is rising, with many crashes leading to incapacitating injuries or fatal outcomes; notably, 82% of pedestrian fatalities occur in urban areas (NHTSA, 2021).

Pedestrians are among the most vulnerable road users, especially when a motor vehicle is involved. According to one study, only 10% of pedestrians will survive if involved in a crash with a car traveling at 40 mph (Tefft, 2013). Therefore, heavily utilized urban corridors impose a potentially significant danger to pedestrians if not provided with adequate safety measures. Consequently, it is essential to improve our understanding of the critical factors impacting pedestrian crashes in urban areas, aiming to develop risk mitigation strategies to curb this trend.

This chapter discusses a study performed to improve our understanding of pedestrian crash severity in urban areas by investigating the factors affecting pedestrian crashes with incapacitating and fatal outcomes, leading to the development of an explanatory model. This was accomplished using detailed pedestrian crash records from Nashville and Memphis, the two largest urban areas in the State of Tennessee. It concludes with a policy discussion directed at enhancements to transportation infrastructure and operations with pedestrian safety in mind. It also addresses the extent to which policy changes directed at pedestrian safety may offer co-benefits for bicyclist safety.

4.2 Literature Review

As pedestrian safety in urban areas is a global problem, the international research community has been actively engaged in providing adequate protection within pedestrian infrastructure and operations. Table 4.1 lists the results of significant factors found in previous studies, organized according to field type and variable.

Table 4.1: Significant Pedestrian Crash Factors from Prior Studies

Variable Field	Variable Analyzed	Relevant Studies
Environmental	Lighting Weather Intersection Type Speed Limit Traffic Control Device Number of Lanes Road Curvature Traffic Volume (AADT) Land Use	Mujalli et al., 2019 Li et al. 2017 Haleem et al., 2015 Samerei et al., 2021 Almasi et al., 2021 Verzosa and Miles, 2016 Khattak and Tung, 2015 Wanvik, 2009
Crash Specific	Crash Type Severity	Theofilatos and Efthymiou, 2012
Time	Time of day Day of week	Mokhtarimousavi et al., 2020 Mokhtarimousavi, 2019

Table 4.2 summarizes relevant study methodologies. The most common modeling techniques have included the Multinomial Logit Model, Mixed Logit Model, and Ordered Logit Model. Note, however, that some of the more innovative modeling techniques, such as support vector machines and weighted cost-sensitive learning, have not been applied to this problem.

Table 4.2: Previous Methodologies and Modeling Techniques

Modeling Technique	**Author**	**Study Focus**	**Variables Analyzed**
Multinomial logit model (MNL)	Chen and Fan, 2019	Pedestrian – vehicle crash injury severities into five categories (no injury, possible injury, evident injury, disabling injury and fatality) in North Carolina	Probability of fatalities and disabling injuries are increased due to driver's physical condition (bad condition), vehicle type (motorcycle and heavy truck), pedestrian age (26–65 and over 65), weekend, light condition (dawn, dusk and dark), roadway characteristics (curve), roadway surface (water), roadway class (NC route) and speed limit (35–50 mph and above 50 mph)
Partial proportional odds logit model	Li and Fan, 2019	Pedestrian – vehicle crash injury severities in North Carolina	Models have a better performance of developing separate injury severity models for each age group compared with estimating a single model utilizing all data.
Support vector machine and MNL	Mokhtarimousavi, 2019	Pedestrian crash - time of day analysis in California	Pedestrian action, type of vehicle, roadway type, weather condition and crash type are the top variables which affect pedestrian fatal and severe injuries for daytime and nighttime.
Binary logistic regression and tree-based models	Hu et al., 2020	Pedestrian causality in Changsha City, China	Several clusters of pedestrian crashes were identified in urban areas, which are related to the population, road network, regional functional zoning and social and economic characteristics. However, the severity of pedestrian casualties has strong relationships with darkness, lighting conditions, road isolation facilities and pedestrian age and behavior. Casualties are more severe at night than during the day, and school-age children and elderly pedestrians tend to suffer more.
Classification and regression tree with random forest approach	Li et al., 2017	Impact of weather conditions on injury severity in Great Britain	Under severe weather conditions pedestrian age, vehicle maneuver and speed limit are the important features affecting pedestrian severity
Extracted rules from Bayesian networks	Mujalli et al., 2019	Urban and suburban areas of Jordan	Roadway type, number of lanes, speed limit, lighting, and adverse weather conditions increase the risk of fatal and severe injuries.
Mixed logit model	Haleem et al., 2015	Signalized and non-signalized locations in Florida	For both location types higher AADT, speed limit, and percentage of vehicle type; at-fault pedestrians; pedestrians age; rainy weather; and dark lighting condition were associated with higher pedestrian severity risk.
Mixed logit model	Kim et al., 2010	Pedestrian-injury severity in pedestrian-vehicle crashes of North Carolina	Dark lighting conditions, vehicle size, freeway, high speed, impaired driving and old age of pedestrians lead to high probability of fatal injuries.
Random-parameter (mixed) logit	Aziz et al., 2013	Pedestrian injury severity levels of New York City	Number of lanes, grade, light condition, road surface, presence of signal control, type of vehicle, parking facilities, commercial and industrial land use are found to be statistically significant.

Modeling Technique	Author	Study Focus	Variables Analyzed
Artificial neural network and random parameter ordered response models	Mokhtarimousavi et al., 2020	injury severity of pedestrian crashes by time-of-week in California	age, alcohol consumption, pedestrian presence location, time of day, light, and surface conditions significantly impact injury severity
Latent class clustering and MNL	Sun et al., 2019	Pedestrian crashes in Louisiana	Pedestrian crossing and entering roads, crash hours between midnight to 6 pm, dark-unlighted conditions, dark-lighted conditions, and non-intersection locations are identified as significant variables
Latent class with ordered probit method, k-means with MNL	Mohamed et al., 2013	Pedestrian injury severity for New York City, US and the City of Montreal, Canada	Pedestrian age, location type, driver age, vehicle type, driver alcohol involvement, lighting conditions, and several built environment characteristics influence the likelihood of fatal crashes
Latent-class logit and mixed logit models.	Behnood & Mannering, 2016	Pedestrian injury severity in three distinct economic time periods in Chicago, Illinois	Variables potentially affecting injury severities were considered, including time, location, and severity of crashes and data on the roadway and environmental conditions, pedestrian characteristics, and crash characteristics. Significant temporal instability was seen, which likely results from a combination of the economic recession and the long-term evolution of the influence of factors that affect pedestrian-injury severity.

4.3 Data Analysis

Data associated with pedestrians involved in motor vehicle crashes were obtained from the Tennessee Department of Transportation (TDOT) for the period from January 1, 2017, through December 31, 2020, covering the entire state. Attributes associated with each crash record are listed in Appendix I.

During this period, 5,494 pedestrian crashes were recorded for which there was complete information (see Table 4.3), distributed across the state as shown in Figure 4.1. Of the 95 counties in TN, Shelby County and Davidson County recorded the highest number of pedestrian crashes, collectively accounting for 56% of the total crashes (56% of 5,494 crashes), more than one-half of the overall state total. This is to be expected since these two counties are densely populated and include the cities of Memphis and Nashville, respectively. These two locations subsequently became the focus of the modeling effort.

Table 4.3: TN Pedestrian Crashes by Year

Year	Total Pedestrian Crashes
2017	1,400
2018	1,394
2019	1,473
2020	1,227

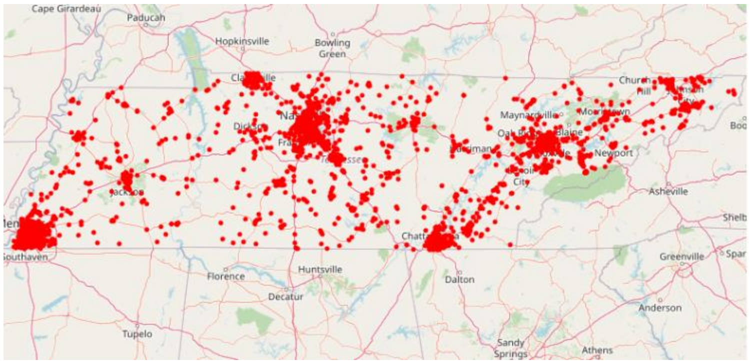

Figure 4.1: Pedestrian Crash Cluster in TN (2017-2020)

Pedestrian crash severity results for the two counties are shown in Figure 4.2. The pedestrian crash injuries were classified as resulting in either no injury, injury, serious (incapacitating) injury and fatal injury. Particularly noteworthy is that fatal and/or incapacitating injury (KII) crashes account for 28% of 3,055 recorded crashes for Davidson and Shelby County.

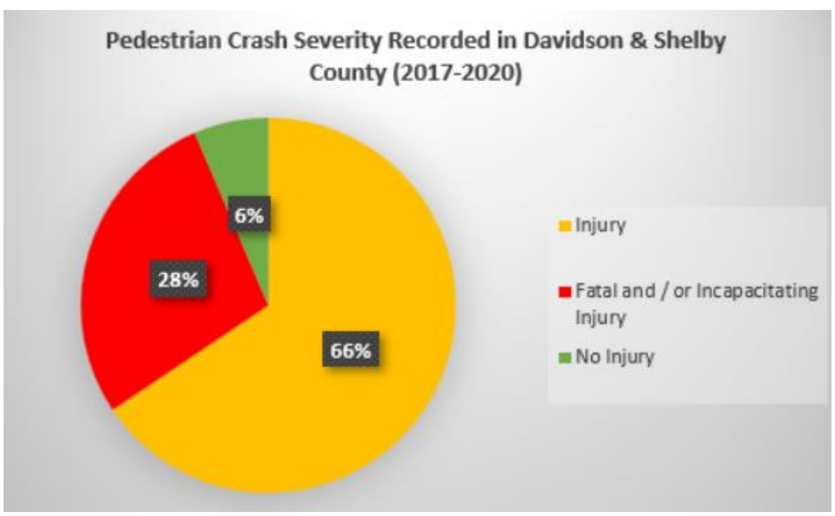

Figure 4.2: Pedestrian Crash Severity in Davidson and Shelby County

Figure 4.3 shows the distribution of pedestrian crashes based on location type. Pedestrian KII crashes are highest on roadways, followed by intersections and ramps.

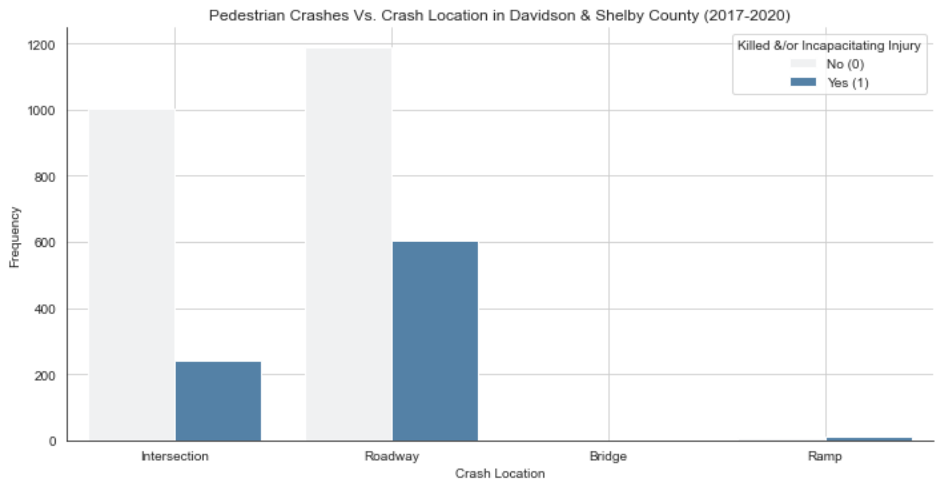

Figure 4.3: Pedestrian Crashes by Location Type in Davidson & Shelby County

As displayed in Figure 4.4, the most significant number of pedestrian crashes in general and those resulting KII occurred on four-lane roads (two lanes in each direction). It can also be seen that four-lane roads experience a high number of pedestrian crashes on medians and in turn lanes.

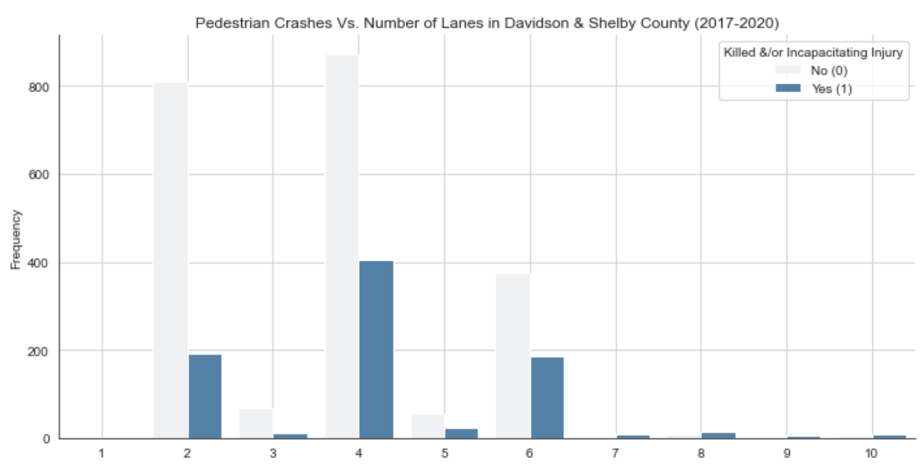

Figure 4.4: Pedestrian Crashes by Lane Configuration

As seen in Figure 4.5, roads with speed limits from 30 mph to 45 mph experience high pedestrian crashes, with the proportion resulting in a KII increasing at higher speeds.

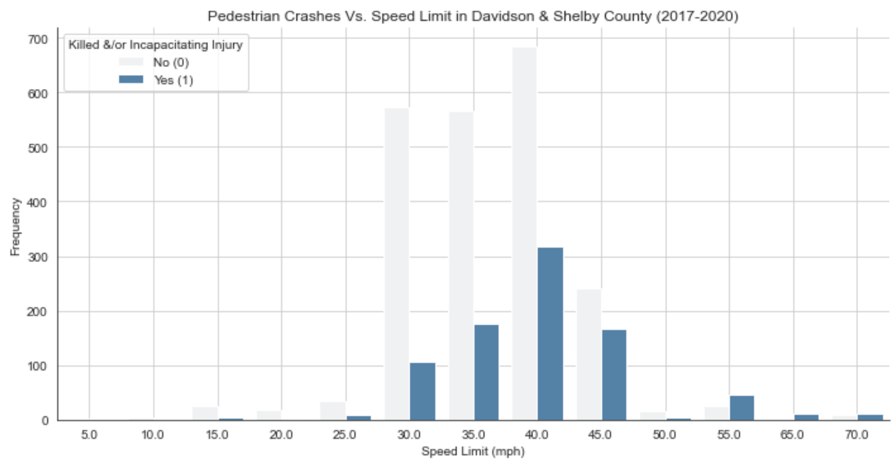

Figure 4.5: Pedestrian Crashes by Road Speed Limit

4.4 Modeling Approach

While the literature review cited a variety of modeling approaches that have been developed for predicting pedestrian crash severity, none have considered a comparison of various classification predictive modeling techniques with different balancing methods along with a rigorous feature selection process. In this study, model estimation was performed using Logistic Regression (LR) and Decision Tree (DT) techniques while also considering Random Forest (RF) and Support Vector Machine (SVM). A classification technique is applied to the models such that one can predict the categorical outcome of a killed and/or incapacitating injury (Class 1) and no killed and/or incapacitating injury (Class 0). LR is a supervised machine learning algorithm that uses a logistic function to model the outcome and serves as a baseline for our binary classification problem. It represents a widely used method to study risk factors impacting injury severity. The advantage of LR is its easy implementation and quick description of the relationship between the input variables and the output variable with no scaling of features. The drawback with LR is that it can only construct linear boundaries, it assumes no correlation between input variables and the output variables, input variables are correlated to each other, and a constant need to set the threshold (from the baseline of 0.5) on which classification is based such that we reduce the false prediction of the output variable.

DT is another frequently used supervised learning classification algorithm for understanding and interpreting data, where the top node is the root node, representing the best feature that divides the data. Each internal node is a feature and branches indicate the decision, with the class label being represented by a leaf node. A DT consists of nested if-else statements where successive conditions are checked unless a conclusion is reached (i.e., a decision is made if the output will be a class 1 or class 0 only if it satisfies certain criteria for each of the features), which can then be shown graphically in the form of a decision tree or a flow chart. DT outperforms an LR, especially when the relationship between the input and out variables is complex and non-linear. DT also helps build easy-to-understand models for visualization; however, DTs tend to overfit. DT serves as a foundation for RF, which is yet another supervised machine learning algorithm.

Although RF has not been extensively used as a classification algorithm for analyzing bicycle crashes, it was included because RF has been shown to improve modeling performance relative to a single tree classifier (e.g., DT) and LR. RF enables multiple uncorrelated DTs to grow, thus creating a forest. RF uses a technique called feature bagging, where features are selected randomly for individual DTs, which is similar to bagging procedure. With feature bagging, the correlation between each DT is reduced but the overall accuracy of the model increases. RF performs better compared to LR and DT as it is more robust to noise, and able to capture the non-linear tendencies by putting all the weak learners in an ensemble that is used to make the prediction. It also avoids overfitting because those individual learners are weak, so it is not one massive model that could lead to overfitting the data (A.C. Muller et.al.,2017).

SVM is a supervised machine learning method that uses a decision boundary to divide the data, based on the features, into either a fatality or incapacitating outcome or a less severe injury outcome. It calculates a maximal margin hyperplane separating the two classes. The boundary decision can be a linear or non-linear function. When comparing the efficacy of RF and SVM performance, RF performance is sensitive to the number of features randomly selected at each node and the number of trees and its choice can lead to performance variation leading to opportunities for overfitting, unlike SVM parameters that have minor effects on error. A major advantage of SVM is its insensitivity toward unbalanced datasets and avoiding overfitting. We chose to use a linear decision function in SVM based on the observed gap between the testing and training Receiver Operating Characteristic (ROC) values for RF (A.C. Muller et.al.,2017)

In this study, we applied LR, followed by DT, RF, and SVM to observe the model prediction outcome. It is not necessary to use models that build on the previous ones; however, this was done to tune the classifier and improve model performance.

The dependent variable was defined as a numerical Boolean variable, with a value of 1 indicating a KII outcome and 0 otherwise (i.e., minor injury or no injury). Before conducting model estimation, data pre-processing was performed to remove records with missing data. This resulted in the selection of the following candidate crash factors (attributes) to be considered as independent variables in model estimation: location, functional class, number of lanes, speed limit, average annual daily traffic (AADT), impaired driver, weather, lighting, weekday, and time-of-day. Categorical values for location (roadway, intersection, bridge, ramp), functional class (urban, rural), impaired driver (yes, no), weather (clear, cloudy, rain, fog, snow, severe crosswind, sleet, hail), lighting (dark, dawn, daylight, dusk), weekday (yes, no) and time-of-day (early morning, morning, AM peak, afternoon, PM peak, late evening) were converted to numerical Boolean variables (0 or 1). AADT, speed limit, and number of lanes were scaled to help decrease the variation of magnitude for these features compared to the other features. Scaling reduces model performance fluctuations by interpreting the features on the same scale.

The data set was divided such that 80% of the records were used for training and the remaining 20% for testing. We attempted to balance the training data before model insertion. Note that as shown in Figure 4.6, the dependent variable is unevenly distributed in the training dataset, with 28% of 3,055 pedestrian crashes resulting in a KII (i.e., minority class), and 72% of 3,055 pedestrian crashes resulting in non-KII (i.e., majority class). There are several techniques to handle imbalanced datasets, but broadly speaking data can be balanced by decreasing the majority class sample size (under-sampling) or increasing the minority class sample size (oversampling). We will consider two widely used algorithms for under and oversampling i.e., Near-Miss and SMOTE. Additionally, we look at misclassification costs as a way to address imbalanced classification.

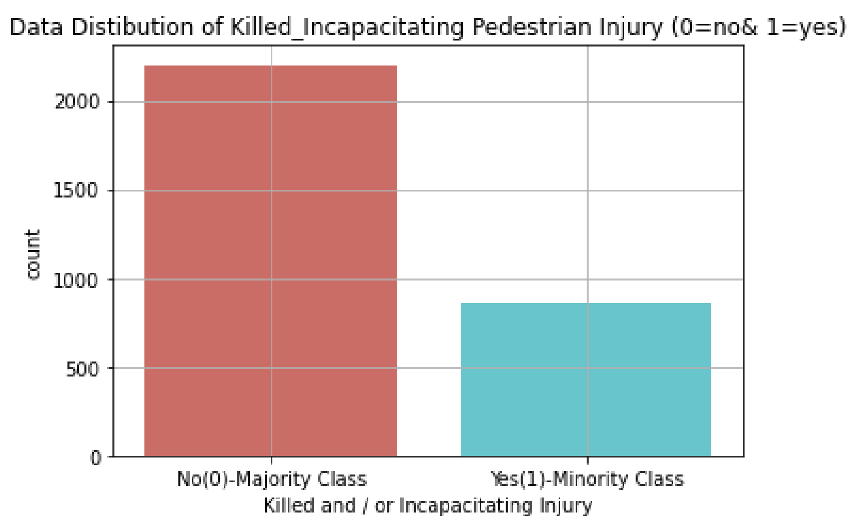

Figure 4.6: Unbalanced Data for Dependent Variable

Three sampling techniques were applied to training data as part of the modeling effort to gauge which method gave the best prediction capability. We used the NearMiss algorithm for under-sampling to prevent the problem of information loss in most traditional under-sampling techniques. Here, the majority class is reduced to the total number of the minority class. In a near-miss algorithm, distance between all the points in the majority and minority class is calculated; for all instances where this distance is the shortest this group of points in the majority class is selected for elimination. For each example in the minority class, a given number of the closest majority class is selected. This method guarantees that every minority point is surrounded by some majority samples. Synthetic Minority Oversampling Technique (SMOTE) was applied for oversampling, a technique where synthetic samples are generated for the minority class that helps to overcome the overfitting problem posed by traditional random oversampling techniques. By linear interpolation of the minority class, synthetically more training data is generated by randomly selecting one or more of the k-nearest neighbors for each minority class by calculating the Euclidean distance between a point and every other sample point in the minority class. Another balancing approach involves the use of cost-sensitive learning (CSL), whereby a larger weight is assigned to the minority class and a smaller weight is applied to the majority class. Since the data points for a killed and/or incapacitating injury are way smaller compared to injured and/or no injured cases and this chapter focuses on being able to detect killed and /or incapacitating injury, CSL can be used in especially such cases. In CSL, each class is given a misclassification cost when training a model, where the aim is to minimize the total misclassification cost. When the class weight is set according to the imbalanced ratio, it implies a modification in the loss function thus improving the training model by pushing the decision boundary that allows improvement in the minority class.

Even though RF and SVM is insensitive to many features, we applied the following feature selection methods to improve the classification performance of LR and DT so that we can examine and compare the models unbiasedly. As the dataset consists of both numerical and categorical inputs, three methods of feature selection were applied sequentially: 1) Correlation coefficient, 2) DT feature importance, and 3) Recursive Feature Elimination (RFE). Feature selection using correlation is a filter approach that is based on feature-to-feature correlation using the standard Pearson correlation coefficient value. The goal here is to find a subset of features that are highly correlated with one another and drop them as they may influence the performance of the model performance. A correlation coefficient threshold of ±0.7 was applied to eliminate highly correlated features (Hulse, J.V. et al., 2012). The second step of feature selection involved the use of DT after dropping the highly correlated features. By using all the features in a DT, we can quickly observe the portion of the features DT uses for the full classification. We drop the features which do not contribute to the classification. The final filter selection method used was RFE, which works by starting with all the features in the training dataset and subsequently removing the undesired features untill a subset of the desired features remains. The RFE starting point was the set of features filtered using DT in the training dataset. The RFE starting point was the set of features filtered using DT in the training dataset. The core of the model used here is DT, where features are ranked by importance, the least important features are discarded, and the model is refitted. This process is repeated until only the desired features remain by performing a cross-validation evaluation of the different number of features and selecting the number of features with the best mean score.

Finally, to understand and explain the output for a killed and/or incapacitating bicycle injury (class 1) for the selected model, Shapley additive explanations (SHAP) is used. SHAP helps interpret the predictions by measuring each feature's contribution (known as Shapley value) to the output (class 1). Shapley values are a concept adopted from the game theory field, whose objective is to measure each player's contribution to the game (Shapley, 1953). We used Kernel Shap to calculate the Shapley value as it can interpret any machine learning model regardless of its nature. Kernel Shap is based on weighted linear regression where the coefficients of the solution are the Shapley value (Lundberg et. al., 2017).

Some limitations were identified in the crash data. The data used for this study consists of only pedestrian impacts associated with motor vehicle crashes. All the data is recorded at the crash scene by law enforcement officers. Once this police report is filed, it is then entered into the data platform. Hence, data efficacy can suffer from human error when reported, collected, and processed at the various stages. Additionally, TDOT does not include near misses and unreported pedestrian incidents. The crash record also does not provide information on the cause of crash, which party involved in the crash was injured (although we assume that if any injuries are reported, it at a minimum involves a pedestrian), nor any details on circumstances prior to the crash.

4.5 Model Results

The overall model performance measure is the extent to which the model can accurately predict whether a pedestrian crash results in a fatality or incapacitating injury. Appendix II provides a list of relevant metrics and their corresponding definitions for each of indicators considered in evaluating the efficacy of model performance.

Table 4.4 summarizes the performance metrics for various models that were estimated using the features that emerged from the aforementioned elimination process: number of lanes, speed limit, AADT, weekday, location type (roadway) and lighting (dark).

As shown in Table 4.4, although oversampling applied to both LR and SVM performs well, SVM with weighted CSL performs slightly better, due to its nearly even but also high true negativity rate (TNR=0.68) and true positivity rate (TPR=0.66), and with lower Type I and Type II errors. Moreover, SVM with weighted CSL has among the highest values of G-mean (0.67), and weighted accuracy (0.67).

Table 4.4: Performance Metrics for Various Model Estimation Techniques

Performance Metrics	True Negative Rate	True Positive Rate	False Negative Rate	False Positive Rate	Geometric-Mean	Weighted Accuracy	Receiver Operating Characteristics-Train	Receiver Operating Characteristics-Test
LR-Unbalanced	0.12	0.94	0.056	0.88	0.34	0.53	0.57	0.53
LR-Under sample	0.61	0.7	0.3	0.39	0.65	0.655	0.63	0.65
LR-Over sample	0.7	0.65	0.35	0.3	0.67	0.675	0.68	0.66
LR-Weighted CSL	0.67	0.66	0.34	0.33	0.66	0.665	0.67	0.67
DT-Unbalanced	0.29	0.83	0.17	0.71	0.49	0.56	0.68	0.56
DT-Under sample	0.56	0.63	0.37	0.44	0.59	0.595	0.68	0.59
DT-Over sample	0.4	0.81	0.19	0.6	0.57	0.605	0.78	0.6
DT-Weighted CSL	0.53	0.7	0.3	0.47	0.61	0.615	0.72	0.62
RF-Unbalanced	0.2	0.9	0.095	0.79	0.43	0.555	0.67	0.56
RF-Under sample	0.59	0.59	0.41	0.41	0.59	0.59	0.73	0.59
RF-Over sample	0.49	0.77	0.23	0.51	0.61	0.63	0.79	0.63
RF–Weighted CSL	0.53	0.75	0.25	0.47	0.63	0.64	0.74	0.64
SVM-Unbalanced	0.088	0.96	0.04	0.91	0.29	0.524	0.55	0.52
SVM-Under sample	0.59	0.71	0.29	0.41	0.65	0.65	0.63	0.65
SVM-Oversample	0.7	0.65	0.35	0.3	0.67	0.675	0.67	0.68
SVM-Weighted CSL	0.68	0.66	0.34	0.32	0.67	0.67	0.67	0.67

The Receiver Operating Characteristic (ROC) curve value for SVM with weighted CSL is also high for the testing data, with no change from the training data. This curve plots two parameters: true positive rate (TPR) and false positive rate (FPR). This measure is derived from a curve plotted on a graph showing the performance of a classification model at different classification thresholds The ROC curve can help identify the threshold by balancing the TPR and FPR than manually checking which threshold works best. A cut-off point of 0.5 is taken for ROC, which means below this value, the model is unable to distinguish between class 1 and class 0. In RF, you obtain the probability of the prediction belonging to a class when you aggregate the indication functions from its decision trees. When you do the inference on the train and test dataset, you get a distribution, and the ROC curve represents the precision of the chosen point of the corresponding probability space. The ROC measures the area under the curve; when the ROC is closer to 1 but greater than 0.5, it indicates a strong model.

The importance of each feature in contributing to the selected model results (SVM using weighted CSL) appears in Figure 4.7. We use Shapley additive explanations (SHAP), which measures the contribution of a feature in model prediction. Here, it can be seen that each of the features contribute equally towards a pedestrian KII (Class 1) or otherwise (Class 0), a desirable condition.

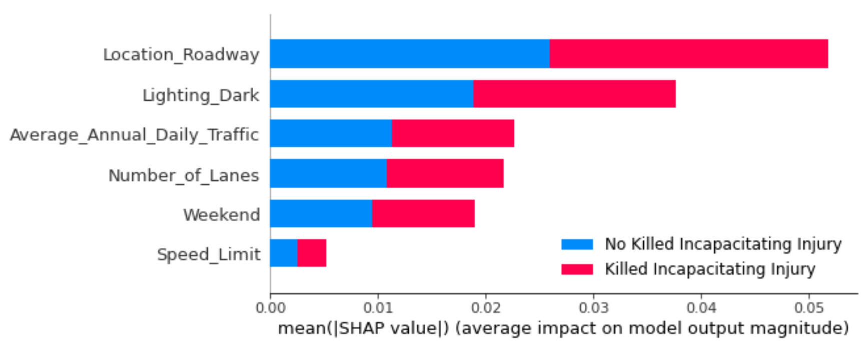

Figure 4.7: Summary Plot Displaying SHAP Values for Model Features

Figure 4.8 displays a bee swarm plot for the study data. This plot helps us understand how a variable may influence model prediction. In this plot, every record in the database is shown as a dot on each row. The color of the dot represents the value of that feature for the event, with red indicating a higher value and blue a lower value.

From this plot, we can observe that even though inadequate lighting conditions and roadway location type are the top two features affecting the model, they have a negative impact a KII outcome. This can be explained by the likelihood that more pedestrians walk during the daytime and cross at street intersections rather than along the roadway. Similarly, roads with a large number of lanes and pedestrians who travel during weekends experience fewer severe crash outcomes.

By contrast, increases in AADT and speed limit on roads positively contribute to a model outcome resulting in a KII crash outcome.

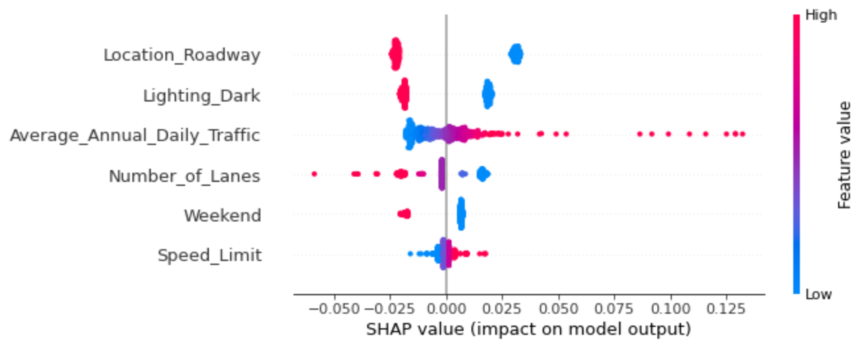

Figure 4.8: Summary Plot Combining Feature Importance with Feature Effect for Class 1 (Killed and/or Incapacitating Pedestrian Injury)

That AADT emerges as an explanatory factor for pedestrian injury severity can be expected as a larger number of motor vehicles are potentially interacting with pedestrians, providing greater opportunity for crash exposure (Obeng and Rokonuzzaman, 2013; Haleem et al. 2015).

The relationship between speed limit and pedestrian injury severity concurs with results supported by other researchers (Li et al., 2016; Haleem et al., 2015; Sasidharan et al., 2015; Gårder, 2004; Kong et al. 2010, Richards 2010, Rosén et. al. 2009 Tefft 2011). As a reference point, Table 4.5

summarizes the relationship between vehicle impact speed and pedestrian injury risk (Nilsson, 2004). The magnitude of this factor in our model suggests that a significant improvement in pedestrian safety would be to direct efforts at reducing pedestrian crashes on higher speed urban roads.

In reviewing these findings, it is important to acknowledge the potential differences between factor correlation and causation, particularly absent any information on traffic volume to normalize the results. Consequently, one must be careful in interpreting how to associate these results with potential risk mitigation considerations. It is entirely possible, for example, that pedestrian fatalities and incapacitating injuries are actually occurring more often in well-lit places, not because the individual safety risk is greater, but because the volume of pedestrians is so much higher that it confounds this relationship. This issue is addressed in Chapter 5.

Table 4.5: Pedestrian Injury as a Function of Vehicle Crash Speed (Nilsson, 2004)

Pedestrian Injury Scale	**Impact Speed**
10%	17 mph
25%	25 mph
50%	33 mph
75%	40 mph
90%	48 mph

Figure 4.9 helps identify groups of similar instances by using hierarchical agglomerative clustering to order the instances. Each position on the x-axis is an event in the database, where red plots

increase the model prediction and blue decreases it. A cluster is observed towards the right of the curve with high prediction of KII.

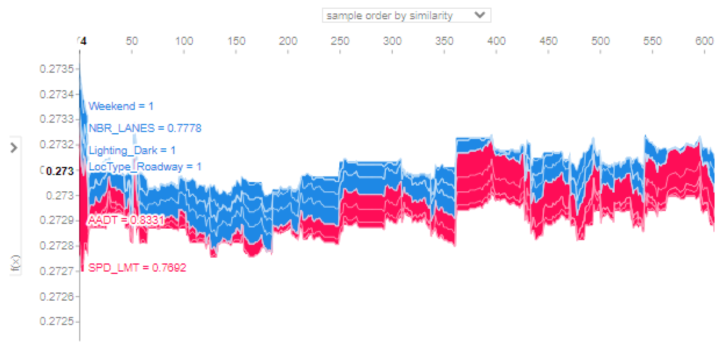

Figure 4.9: Clustering Based on Features for Class 1 (Killed and/or Incapacitating Pedestrian Injury)

4.6 Policy Implications

The feature importance associated with the selected model, as shown in Figure 4.8, provides insights into critical factors that most influence serious pedestrian safety outcomes and their relative contribution. The following discussion provides a general description of policies that may be cost-effective in reducing serious pedestrian safety risk based on the model results. However, the extent to which a particular strategy makes sense is dependent on the site-specific conditions that are present at the location of interest. For example, implementation of a pedestrian bridge may be possible in one location that would be physically infeasible at another site or the benefit-cost may not be sufficient to justify allocation of construction resources. As noted in Chapter 3, it is also possible that the unintended consequences of implementing a supposed safety improvement actually creates greater risk, due to pedestrian perception of enhanced safety when sufficient risk still remains.

AASHTO's Roadway Design Speed Classification identifies a low-speed facility when it is 45 mph or lower. Based on the model results regarding speed limits, we recommend reviewing all urban streets with speed limits at or above 30 mph to assess whether the limit can be lowered. When this is not considered feasible, speed-reducing mechanisms (speed bumps) are an alternative, mainly where frequent pedestrian activity requires that motorists slow down. An added advantage of lowering speed is an increase in a driver's cone of vision. Another strategy would be to remove any conflict points between vehicles and pedestrians by providing a wider shoulder width or separate pedestrian access routes such as a pedestrian bridge. This latter option would be more viable for streets with heavy traffic, multiple lanes, and 30mph or greater speeds.

Additional risk mitigation interventions include introducing measures such as a road diet (where the number and width of travel lanes is reduced on a roadway to cater to other uses like street parking or other modes of travel), pedestrian medians/barriers, or raised sidewalks. When combined with adequate signage, posts, and other demarcations, these intervention strategies could help alleviate at least some crashes and reduce the impact of others when they occur.

Although not explicitly addressed in this paper, conflict with bus stops at the intersections and pedestrian crossings should be reviewed towards implementing improved complete street designs. For lighting conditions, relatively simple risk mitigation strategies would include placement of street lighting (Abdur et al, 2021), especially where the power source is renewable energy such as solar, along popular pedestrian routes to improve visibility. An illumination study can be conducted for new and retrofit projects to suggest the street light locations, height, and type based on existing topography and physical obstructions. Active or passive flashing beacons at pedestrian crossings are another means of notifying motorists of potential safety considerations.

Jaywalking should be strongly discouraged, and the placement of additional pedestrian crossings in locations of heavy pedestrian use should be considered.

The multimodal project manual released in 2018 by the Tennessee Department of Transportation (TDOT), represents a proactive policy in this regard. It advocates for a risk-based approach with built-in flexibility regarding pedestrian facility design. Nashville and Memphis Vision Zero draft plans adopt a similar approach, where comprehensive efforts are made to eliminate pedestrian fatalities and severe injuries while increasing safe and equitable mobility for all.

4.7 Conclusion

Pedestrian safety has been a much-discussed topic, particularly of late, as interest in walking as a sustainable transportation alternative continues to gain popularity. Consequently, policy analysts and planners have been grappling with cost-effective methods to reduce pedestrian crashes, particularly those with serious outcomes. We believe the results of this study have shed additional light on the subject, in particular: 1) providing policy makers with evidence-based recommendations to address pedestrian safety , 2) demonstrating the use of support vector machine and select sampling techniques as having the potential to provide greater accuracy in predicting the likelihood of serious pedestrian crash outcomes, and 3) utilizing the feature weighting of the predictive model to prioritize the types of risk mitigation strategies that offer the greatest safety impact. Ultimately, we hope this study has broadened collective knowledge and awareness of risk-informed decision-making that will lead to lives saved and greater use of walking as a safe and therefore more viable transportation alternative.

Chapter 5

Severe Crash Rates Involving Pedestrians and Bicyclists in Urban Areas

5.1 Introduction

Pedestrians along with bicyclists are the lifeblood of many urban areas (AASHTO, 2018). Walking and biking encourage the use of non-motorized modes for short trips, especially in urban areas, and are essential sustainable urban design concepts. This is in addition to promoting the many health benefits from such an active lifestyle.

Unfortunately, pedestrians and bicyclists are among the most vulnerable road users, especially when a motor vehicle is involved. Safety interventions are needed on such heavily utilized urban corridors by understanding the interactions and factors resulting in a serious crash outcome, namely the pedestrian or bicyclist being killed or suffering an incapacitating injury (KII) outcome. While the frequency of where these outcomes occur is important, it may mask the individual risk of a pedestrian or bicyclist. This requires accurate pedestrian and bicyclist traffic counts, which is also often referred to as exposure (USDOT, 2010).

This chapter discusses a study performed to identify pedestrian and bicyclist crash safety, both in the aggregate user population and from an individual risk perspective, on specific roadway segments in the urban area of Nashville, TN, where significant pedestrian and bicyclist severe crash outcomes have been reported (Dash et al., 2022, 2022).

5.2 Literature Review

Various methods have been explored to calculate the exposure of pedestrians and bicyclists. Population measures have been proposed to estimate motor vehicle and pedestrian/bicyclist exposure to risk. It assumes that crashes between motor vehicles or between pedestrians/bicyclists and motor vehicles are more likely to occur when there are more residents, drivers, motor vehicles, pedestrians, bicyclists, or bicycles in a given area. Over the past several years, NHTSA has reported the number of motor vehicle fatalities and fatality rates based on three population types in the United States : motor vehicle crashes per 100,000 residents, per 100,000 licensed drivers, and per 100,000 registered motor vehicles (Marshall et al., 2011, DiMaggio et al., 2016). However, such population-based methods have limited use when examining pedestrian/bicyclist crashes since these methods do not consider the opportunity of exposure to motor vehicles, especially at a specific type of location (e.g., roadway), nor is it sensitive to the amount of time or distance that a pedestrian or bicyclist is exposed to motor vehicle traffic. Additionally, population metrics do not account for external changes in behavior patterns, such as changes in walking or bicycling behavior for health or environmental reasons with a constant population of residents, bicyclists, and/or bicycles. Such a metric also runs counter to the notion of safety in numbers which has been hypothesized in the literature (Elvik, 2009) that the denser the population of pedestrians or bicyclists, the lower the probability of a crash.

The amount of time that a pedestrian or bicyclist engages in certain activities may be taken as a measure of exposure. This measure has the advantage of capturing time differences between pedestrians who walk more and those who walk less. However, the measure is not sensitive to where people walk. As a result, it includes time walking on various unspecified locations such as sidewalks, trails, and other facilities not shared with motor vehicles. The time spent walking in these facilities represents an overestimate of exposure because the likelihood of a crash between a pedestrian and a motor vehicle is extremely small at these locations. If the measure had specified time spent walking in locations where pedestrians and motor vehicles share the same facility, the time metric would have represented a variant of the original metric, the difference being time walking in the facility would have replaced distance walking in the facility. For a constant walking speed, the distinction between time walked and distance walked is minimal; therefore, a distance metric is preferred such as vehicle miles traveled (VMT) for motor vehicles. However, the distance metric uses total distance traveled rather than the total distance traveled on the facility shared with motor vehicles. Thus, exposure is overestimated, and risk underestimated.

One measure of pedestrian exposure that has been investigated in the past is the number of pedestrians observed on the roadway. Pedestrian safety analyses based on only the number of pedestrians observed in the roadway is better than population density as the latter could distort risk values. This helped observe and reflect only changes in walking behaviors Some authors included the number of motor vehicles and the distance factor to the pedestrian volume as an exposure metric which could address driving behaviors to some extent. The drawback of volume metric manual counts was that it was conducted for short durations, at only a few specific locations, and these were extrapolated using an adjustment factor for a longer duration and to other similar locations. Ferenchak et al. (2020) uses age-specific metrics based on fatality data from FARS and exposure data from national surveys such as US Census, the National Household Travel Survey (NHTS), American Community Survey (ACS), and National Sporting Goods Association (NSGA). The socio-economic demographic population used as a measure of exposure (Ferenchak et al, 2019 & 2021; Rebentisch et al., 2019) also suffers from the same drawbacks as population metrics.

Studies conducted for the metrics mentioned above mostly used social surveys and lacked the accuracy of observational counts of pedestrian and bicyclist activity. They are also prone to error as they are based on a person's memory or behavior rather than an empirical observation.

Some studies address perceived risk as an exposure metric to identify high-risk urban infrastructures (Stülpnagel et al., 2022, Ryerson et al., 2021, Bigazzi et al., 2022) while others have conducted studies in a stimulator or laboratory to address driver inattention at intersections for vehicle-pedestrian/bicyclist crashes using eye-tracking technology (Kaya et al, 2021).

Procedures and guidebooks exist on how to perform pedestrian and bicyclist volume counts. These include FHWA Exploring Pedestrian Counting Procedures, FHWA Bicycle-Pedestrian Count: Technology Pilot Report, and NCHRP Report 797 - Guidebook on Pedestrian and Bicycle Volume Data Collection. However, performing such counts for non-motorized modes have faced challenges that include organizational interest, availability of technological tools, and funding. Unfortunately, this lack of volume data has been a barrier to planning a more effective roadway facility for all users, especially when considering locations of high pedestrian and cyclist interactions with motor vehicles.

Most non-motorized counts nowadays are still performed manually and are also difficult to scale. Currently, the best practice is to establish permanent counts at strategic locations, which can help conceptualize and adjust short-term manual counts. Some state transportation agencies, including Oregon, Washington, Vermont, Colorado, and North Carolina, have established permanent manual counts at strategic locations as part of a statewide data collection program. As Washington aims to double the number of cyclists and pedestrians by 2027, these measures bring important attention to the need to improve the safety of vulnerable road users (WSDOT, 2008).

Fortunately, technological data collection advancements are helping to create large volumes of data compared to traditional counting methods while requiring less effort and fewer resources, especially for non-motorized travel. These techniques use either a passive or an active method. A passive method requires little or no input and effort from the traveler. These include leveraging global positioning system (GPS) enabled devices or location-based services (LBS). INRIX and StreetLight are two web-based service providers that use passive methods. Active data collection requires user participation and input, such as using fitness apps or involvement in bicycle-share programs. Strava is a fitness activity tracking map that uses an active data collection method (CDOT, 2018).

StreetLight is unique as it combines mobile data from GPS and LBS data types, geospatial information, and textual datasets (census data) where socioeconomic factors can be considered. It uses a methodology that takes anonymized location records from smart phones and navigation devices in connected cars and trucks, supplemented by other sources (e.g., parcel and digital road network data). This information is utilized in an algorithm that transforms the data into contextualized travel patterns. The final step involves normalizing and aggregating the results to deliver insights into how vehicles, including bicycles and pedestrians, move by road and Census block. A version of StreetLight can directly analyze and visualize pedestrian and bicycle activity metrics with an online platform (StreetLight Data, 2021).

5.3 Approach

Pedestrian and bicyclist crash counts and their respective features contributing to a KII outcome in Nashville, TN, have been analyzed and presented in previous work published by the authors (Dash et al., 2022 and 2022). While crash frequency is important, we also need to know the risk to an individual pedestrian or bicyclist. Hence, facility usage information for pedestrians and bicyclists is used as an exposure metric. KII percentage and rates are used toward improving safety of non-motorists on roadways. Here both percentages and rates are given equal weights simply because we want to identify roadways with high KII cases and where the exposure of crashes is high. In the absence of permanent and short-term activity counts, the StreetLight online platform was used to determine volume counts for pedestrians and bicyclists at a facility level. These volume counts were available from 2018 to the present, as were the respective crash counts. Roadways with a KII percentage (i.e., ratio of total KII crashes to total crashes from 2018 to 2020) of five were the minimum requirement for inclusion in the pedestrian and bicycle assessment.

It has been suggested in some literature such as Ryerson et al. (2021) that counting killed or serious injuries (KSI) and even adding exposure or near misses, which is an objective safety metric, is insufficient; rather it can be improved by addressing the perceived safety context for a pedestrian or bicyclist (i.e., by assessing the cognitive workload) also termed a proactive safety metric. However, the major limitation of this study is that the data is collected for a 0.3-mile urban mixed-use arterial road with 43 cyclists wearing eye-tracking glasses for an unspecified duration or time. It does not address the fact that safety perception varies from person-to-person due to reasons such as individual experience, risk tolerance level, frequency of use, and gender. The existing crash data from TDOT also does not include any pedestrian or bicyclist exposure metrics at a facility level, nor does it include information on near misses, information on the non-motorists, factors prior to the crash, their reactions, or factors that they think would have prevented a non-injury scenario. In the absence of permanent counters, it is reasonable to use KII percentage and KII rate where the pedestrian and bicycle activity, especially at the facility level, can be obtained from Streetlight for the high crash count locations in Nashville, TN.

The data validation process for pedestrian and bicyclist volumes for the selected roadways of Nashville, TN obtained via Streetlight using manual counts or permanent counts was not possible. However, pedestrian and bicyclist metrics have been validated against permanent counters in San Francisco, Philadelphia, and Ottawa and a correlation of 0.7 to 0.9 have been reported between the daily trip counts from counters and trip counts from Streetlight (Streetlight, 2020). Another limitation is that the number of crashes recorded even over three years may result in too small a sample size for pedestrians and bicyclists in some locations.

5.4 Analysis

A total of thirty-three road segments for pedestrians and thirty-one road segments for bicyclists met the criteria mentioned above. For pedestrians and bicyclists, respectively, each roadway's KII percentage was normalized (on a scale from 0 to 1) according to the location with the highest KII percent value. These values serve as a proxy measure for severe crash frequency.

Pedestrian and bicyclist volumes for each segment were obtained from Streetlight Data and averaged for the same three-year period. Using this information, a KII rate was calculated as a ratio of total KII crashes to the corresponding volumes for pedestrians and bicyclists, respectively, on each segment. Using a similar approach to the KII percentage ranking, KII rates were normalized to values between 0 and 1. Figures 5.1 and 5.2 show the KII percentage (total KII/total crashes) and KII rate (total KII/average volume) for pedestrians and bicyclists, respectively, by individual road segment.

Identification of high-risk locations for pedestrians and bicyclists, respectively, was obtained by adding the rank order values for KII percentage and KII rate (see Table 5.1 for pedestrians and Table 5.2 for bicyclists). This, in effect, represents an equal weighting of high-frequency locations where KII events have occurred and where the rate at which an individual is at risk for a KII crash. Note that Murfreesboro Pike, Shelby Avenue, Commerce Street, West Trinity Lane, South Old Hickory Boulevard, Nolensville Pike, East Old Hickory Boulevard, Antioch Pike, and Dickerson Pike each appear in both pedestrian and bicycle top ten streets that should be prioritized for pedestrian and bicyclist safety enhancements. This suggests that designing safety enhancements at these locations could simultaneously benefit pedestrians and bicyclists.

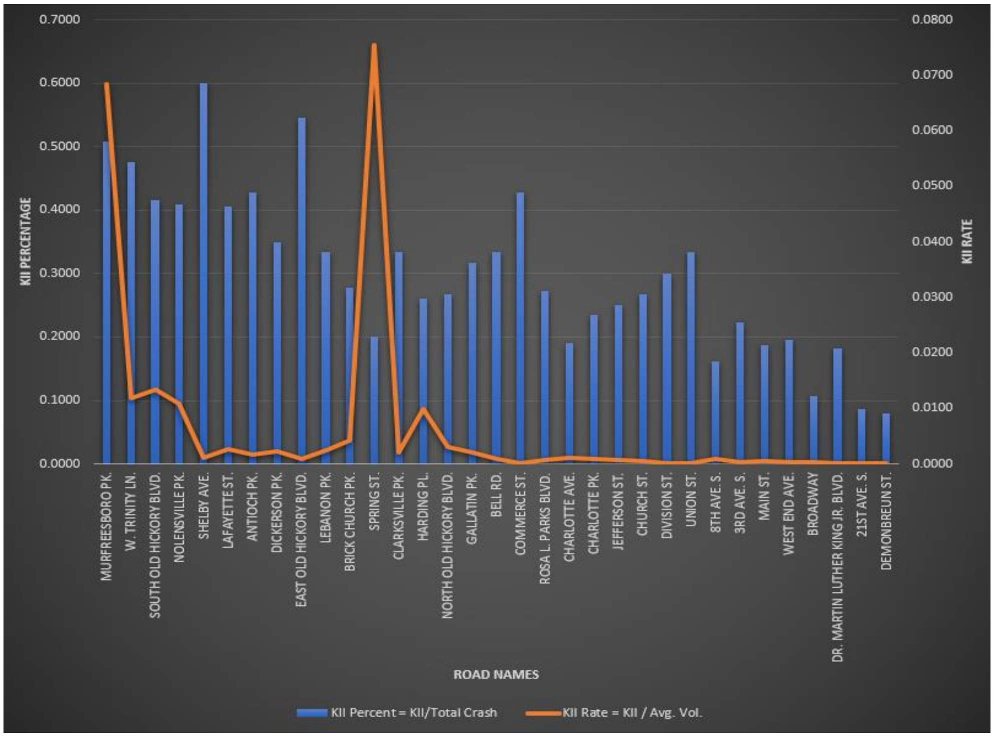

Figure 5.1: Pedestrian KII Percentage and Rate - Davidson County (2018-2020)

Table 5.1: Rank Order of Top Ten KII Pedestrian Crashes on Roadways

Segment No.	Road Name	KII Percent Rank	KII Rate Rank	Final Rank
1	MURFREESBORO RD.	0.846	0.906	1.753
2	SPRING ST.	0.333	1	1.333
3	SHELBY AVE.	1	0.015	1.015
4	W. TRINITY LN.	0.793	0.156	0.950
5	EAST OLD HICKORY BLVD.	0.909	0.012	0.921
6	SOUTH OLD HICKORY BLVD.	0.694	0.176	0.871
7	NOLENSVILLE PK.	0.681	0.143	0.825
8	ANTIOCH PK.	0.714	0.023	0.737
9	COMMERCE ST.	0.714	0.0006	0.714
10	LAFAYETTE ST.	0.677	0.034	0.711

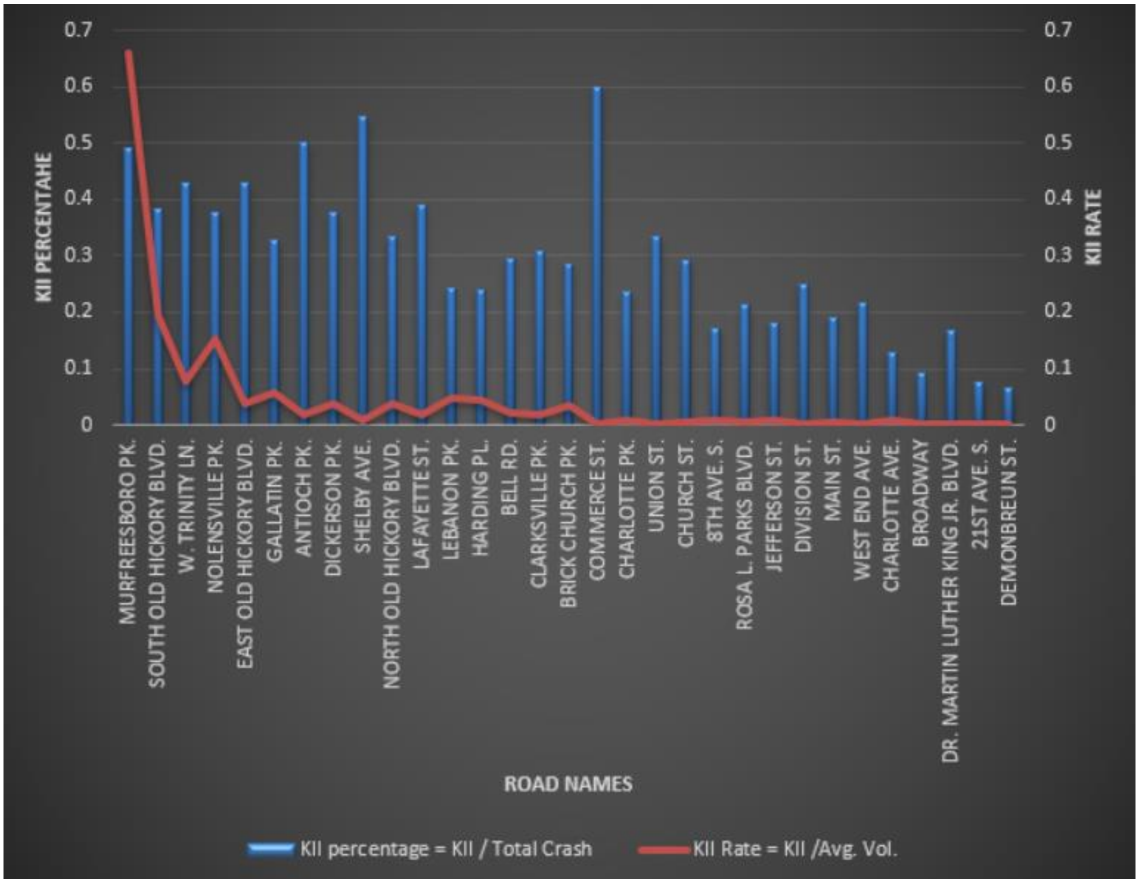

Figure 5.2: Bicyclist KII Percentage and Rate - Davidson County (2018-2020)

Table 5.2: Rank order of Top Ten KII Bicyclist Crashes on Roadways

Segment No.	Road Names	KII Percent Rank	KII Rate Rank	Final Rank
1	MURFREESBORO PK.	0.820	1	1.820
2	COMMERCE ST.	1.000	0.001	1.001
3	SOUTH OLD HICKORY BLVD.	0.641	0.294	0.935
4	SHELBY AVE.	0.909	0.012	0.921
5	ANTIOCH PK.	0.833	0.028	0.861
6	NOLENSVILLE PK.	0.628	0.229	0.857
7	W. TRINITY LN.	0.714	0.114	0.829
8	EAST OLD HICKORY BLVD.	0.714	0.059	0.773
9	DICKERSON PK.	0.625	0.058	0.683
10	LAFAYETTE ST.	0.648	0.028	0.676

5.5 Comparative Assessment

The Nashville-Vision Zero Plan draft includes a systemic safety analysis that assesses crashes and roadway characteristics for pedestrians and bicyclists. While the findings are in alignment with the results of our previous studies on the factors contributing to pedestrian and bicyclist crash severity in urban areas (Dash et al., 2022 & 2022), there are differences in the Plan's list of locations with high pedestrian and bicyclist injuries, respectively. This is due to the approach and limitations identified in the Plan. In particular: 1) data consisted of motor vehicle, motorcycle, pedestrian, and cyclist crashes (minor, serious, and fatal injuries), 2) freeways and ramps were not considered, and 3) weights were assigned based on injury severity (minor, serious, and fatal injuries), vulnerable users (pedestrians and bicyclists), and equity (vulnerable areas).

Moreover, the final designation of high injury roadways was derived by multiplying weights for each crash on that road segment (i.e., severity index * vulnerable user index * equity index). This method identifies a high pedestrian injury network, bicyclist high injury network, and dangerous intersections for pedestrians and cyclists, as shown in Table 5.3. It is unclear if all of the identified roadways in the high injury network are rank-ordered. The Plan uses Random Forest regression models with eight features for pedestrian predictive analysis. The three important variables identified are AADT, the number of lanes, and the fraction of commercial land use within 500 feet. Due to the small sample size for bicyclist crashes, no predictive analysis was conducted; hence it would be difficult to draw conclusions. AADT for local roadways was assumed to be a constant of 2,400 vehicles per day.

Table 5.3: Top Pedestrian and Bicyclist Injury Roadways and Intersections in Nashville (Nashville Vision Zero Draft)

High Pedestrian Injury Network	High Bicyclist Injury Network	Dangerous Intersections-Pedestrians	Dangerous Intersections-Bicyclist
West Trinity Lane	Church Street	Lafayette St. & Charles E. Davis Blvd.	Division St. & 12th Ave. S.
Murfreesboro Pike	28th Ave North	Gallatin Pike S. & Neely's Bend Rd	Division St. & 12th Ave. S.
Lafayette Street	Charlotte Ave	Gallatin Pike S. & Berkley Drive	Gallatin Pike S & Emmit Ave
Nolensville Pike	West Trinity Lane	Dr. MLK Blvd & Rep John Lewis Way	Highland Ave & 25th Ave S.
Gallatin Pike	Gallatin Pike	Gallatin Pike S. & Madison St.	E. Thompson Lane & Old Glenrose Ave
Dickerson Pike	East Trinity Lane	Nolensville Pike & Welshwood Dr.	
Harding Place	Lafayette Street	Murfreesboro Pike & Millwood Dr.	
Rosa L Parks Blvd	25th Ave South		
Main Street	8th Ave South		
Old Hickory Blvd	Murfreesboro Pike		

5.6 Results

In our previous work examining explanatory factors associated with severe crash outcomes for bicyclists and pedestrians in urban areas (Dash et al., 2022 & 2022), when comparing opportunities for introducing pedestrian and bicycle policy improvements, the same six features (number of lanes, speed limit, AADT, weekday, roadway location type, and dark lighting) emerged as critical factors, albeit with different explanatory effects. This suggests that there may be opportunities to mitigate fatal and severe pedestrian and bicyclist injuries simultaneously by improving specific roadway design and operational elements at high-risk locations.

Shelby Avenue is a case in point, which ranked fourth highest in KII pedestrian and bicyclist crashes. We observe a total of 10 and 11 pedestrian and bicycle crashes, respectively, and each has six KII crashes. By examining the 12 KII crash locations on Shelby Avenue (Figures 5.3 & 5.4), one can assess the opportunity to improve speed management (reduction to 20 mph) as a viable strategy for relieving pedestrian and bicycle crashes. We also observed that all pedestrian crashes occurred within one-half mile of a transit stop. A case could be made for transit stops to be located along the roadway length rather than at intersections. To address inadequate lighting conditions, enhanced crosswalk markings, flashing lights, median refuge islands, and crosswalk signals (e.g., High-Intensity Activated CrossWalk beacon - HAWK) warrant consideration. Improving roadway design (with signal timing for non-motorized modes) at intersections, and construction of a barrier between the bicycle lane and motor vehicle travel lanes can reduce bicycle crashes. Closer examination of vehicular right and left turns at intersections is also worthy of consideration. Additionally, education can help modulate safe travel behaviors for all modes (e.g., motorist, cyclist, pedestrians, transit).

Figure 5.3: Pedestrian (Yellow Mark-up) and Bicycle (Blue Mark-up) KII on S. 6th St. & Shelby Ave

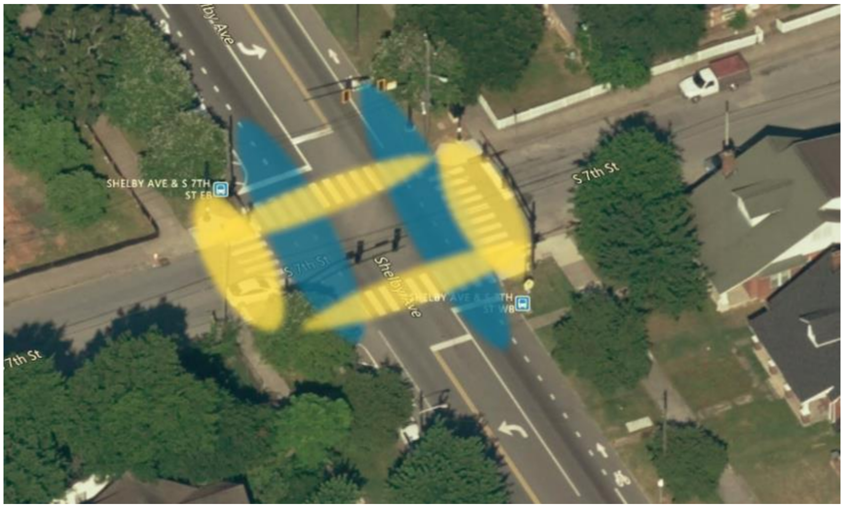

Figure 5.4: Pedestrian (Yellow Mark-up) and Bicycle (Blue Mark-up) KII on S. 7th St. & Shelby Ave

In 2010, the United States Department of Transportation adopted a policy (based on the Code of Federal Regulations (CFR) Title 23 – Highway, Title 49 – Transportation, and Title 42 – Public Health and Welfare) that supports the development of fully integrated activity transportation systems. This policy includes provisions to incorporate and improve conditions for safe and convenient walking and bicycling facilities in transportation projects. It also includes a need to modify minimum design standards and requirements that have proven inadequate to create safe and accessible pedestrian and bicycle facilities. Guidance exists in the FHWA Lighting Handbook, Urban Street Design Guide, Complete Street Guide, Proven Safety Countermeasure (PSC) Tools, and TDOT's Roadway Design Guidelines, describing engineering design solutions involving lighting provisions, pedestrian and bicycle facility design, and traffic calming measures.

Ultimately, the goal is equity for all road users which can be achieved by prioritizing pedestrian and bicycle safety relative to motor vehicle utilization. This will help balance safety on the streets for all users, especially in urban areas. Where implemented, further research will be needed to gauge the efficacy of these recommendations.

Walking and biking are essential transportation alternatives for short distances, which help close the transport loop for an individual. Hence, establishing safe and connected pedestrian and bicycle facilities is a necessity. By doing so, we can create greater access to various locations, particularly for underserved communities.

5.7 Conclusions

Pedestrian and bicyclist safety has been a much-discussed topic, particularly of late, as interest in walking and bicycling as sustainable transportation alternatives continue to gain popularity. Policy analysts and planners have been grappling with cost-effective methods to reduce pedestrian and bicycle crashes, particularly those with serious outcomes.

The results of this study can help shed additional light on the subject, in particular demonstrating how to leverage existing tools to obtain pedestrian and bicycle volumes and severe crashes. By comparing the pedestrian and bicycle KII to the volume for a specific roadway, one can prioritize the streets targeted for risk mitigation. Additionally, we can utilize the predictive model's feature weighting to prioritize the types of risk mitigation strategies that offer the most significant benefit. Ultimately, we hope this study has broadened collective knowledge and awareness of risk-informed decision-making, leading to saved lives and greater use of walking and bicycling as safe and therefore more viable transportation alternatives.

Chapter 6

Conclusion

By executing a data-driven approach, we can better understand transportation risk and safety issues, which can help inform and shape policy, especially for future transportation investment. Advancements in information technology and data analytics are transforming our ability to make risk-informed decisions. through the exploration of new opportunities for evidence-based quantitative feedback.

In Chapter 2, a conceptual design was presented for deploying smart detection systems and their communication technology for use by rail hazmat shippers. It was observed that since implementing SHRIS for its high-hazard rail shipments, Olin Corporation had experienced a dramatic reduction in the risk of transporting hazardous materials as well as realizing cost savings. This proven and affordable system offers an opportunity for hazmat rail shippers of all sizes to leverage this capability and not just a select few within the industry. Such widespread adoption benefits not only each shipper but the industry as a whole.

In Chapters 3 and 4, the study here aims at evaluating the degree to which three main classification algorithms are impacted by class imbalance, with the goal of identifying the algorithms that perform best and worst on imbalanced data and other forms of balanced data. In particular, this study assesses the relative impact of class imbalance and three class balancing techniques on three/four different classification algorithms as they are applied to data sets with varying levels of class-balanced data. The results from this study demonstrate that certain classification algorithms, such as random forest and support vector machine, perform very well in the presence of cost-sensitive learning while other algorithms, perform poorly. Improved modeling techniques were introduced for determining critical factors influencing bicyclist and pedestrian severe accident outcomes on an imbalanced dataset. This provides an opportunity to make more cost-effective bicycle and pedestrian safety policy decisions based on improved explanatory models. The critical factors influencing pedestrian and bicyclist severe crash outcomes alike were inadequate lighting conditions, crashes on roadways, speed limits, average annual daily traffic, number of lanes, and weekends. Based on this finding, specific suggested policy changes were presented for implementation consideration.

Chapter 5 demonstrated that the ability to evaluate the efficacy of the potential risk mitigation strategies presented in Chapters 3 & 4 requires an ability to consider not just the frequency of pedestrian and bicyclist crash outcomes at specific locations, but also the demand for services in these locales. In the absence of this exposure information, the rate at which an individual pedestrian or bicyclist may suffer severe harm remains unknown. Available methods and technologies for measuring pedestrian and bicyclist severe crash exposure were utilized, illustrated in a case study using Nashville, Tennessee to identify the most critical urban area locations using both severe crash frequencies and rates. These techniques are scalable and transferable to other locations, both temporally and spatially. Of particular note is the similarity among the top ten road segments for pedestrian and bicycle risk, respectively, suggesting that not only should they be prioritized for pedestrian and bicyclist safety enhancements, but there may be cost-effective solutions that provide mutual benefits.

The research performed in this dissertation can be expected to contribute to transportation safety methods and practices. However, there remain several avenues for future research, in particular the following:

1. Integration of innovative information technologies and data analytics to enhance the safety and security of rail transport of hazardous materials, with the potential transferability to other cargo types and freight modes.

Both the barge and truck modes can benefit from the implementation of a SHRIS-type system, and the vast majority of highly valued features can either be adopted directly using technologies embedded in the SHRIS system or can be implemented by making modifications to SHRIS system elements. There are no dramatic differences in the development effort involving the implementation of such a system for either mode, but rather certain feature elements add greater value for one mode and vice-versa. While the development of a SHRIS-type system for hazmat movements by either the truck or barge mode is likely to offer safety and security benefits, a greater need and opportunity appears to exist for developing and deploying such a system in the barge domain. This rationale rests with the communication technology gap found in the maritime industry created by the dependence on a paper-based system kept on board each vessel. Further, the significantly larger cargo volumes per barge shipment create the potential for a more consequential impact in the event of a material release.

2. Implementation of improved modeling techniques to identify evidence-based risk mitigating strategies and policies to enhance safety. Leverage existing methods and technologies to measure activity level on a transport segment or facility, which can help prioritize cost-effective safety enhancements at specific locations. This approach is transferable to other road users to achieve an optimal safety for all modes.

An analysis is as good as the available data. It has been noted that although pedestrian and bicyclist severe crash outcomes typically involve motor vehicle crashes, current databases do not include near misses and are plagued by unreported incidents, as well as lacking cause and circumstances prior to the crash. Including such information and standardizing the reporting system is a desirable future task.

Much of this may be accomplished through the advent of more sophisticated technology for studying pedestrian and bicycle safety. This includes putting eye-tracking cameras on pedestrians, cyclists and motorists to track what they are looking at on a frequent time scale; installing accelerometers on pedestrians, cyclists and motorists to track when they turn their heads to look to their side or behind; and instrumenting streets with cameras to observe interactions among motorized vehicles, cyclists and pedestrians, particularly at intersections.

References

2019 Pedestrians Traffic Safety Fact Sheet, (2021) DOT HS# 813079

2019 Traffic Safety Facts A Compilation of Motor Vehicle Crash Data (Annual FARS/GES/ CRSS Report), (2021) DOT HS# 813141

Abdel-Aty, M., and J. Keller. Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections. In Accident Analysis & Prevention, Vol. 37 No. 3, (2005) pp. 417-425

Abdur, R., Aya, K., Teppei, K., Hisashi,K., A mechanism to enhance bicycle conspicuity and visibility and increase detection distances: New insights into bicycle safety, IATSS Research, Volume 45, Issue 2, (2021) pp. 241-250

Al-Ghamdi, A. S. Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity. In Accident Analysis & Prevention, Vol. 34 No. 6, (2002) pp. 729-741

Allen-Munley, C., Daniel, J. and Dhar, S. Logistic Model for Rating Urban Bicycle Route Safety, Transportation Research Record 1878, TRB, National Research Council, Washington DC (2004), pp. 107-115

Almasi, S.A., Behnood, H.R., Arvin, R. Pedestrian crash exposure analysis using alternative geographically weighted regression models. Journal of Advanced Transportation, (2021), 10.1155/2021/6667688

American Association of State Highway and Transportation Officials (AASHTO). Green Book: Policy on Geometric Design of Highways and Streets. Edition 7. (2018)

American Association of State Highway and Transportation Officials (AASHTO). Guidelines for Geometric Design of Low-Road Volume Roads. Edition 2. (2019)

American Association of State Highway and Transportation Officials (AASHTO). Roadway Lighting Design Guide. Edition 7. (2018)

American Association of State Highway and Transportation Officials (AASHTO). Guide for Planning, Designing, and Operating Pedestrian Facilities. Edition 2. (2021)

American Association of State Highway and Transportation Officials (AASHTO). Guide for the Development of Bicycle Facilities. Edition 4. (2012)

Apley, Daniel W., and Jingyu Zhu. Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82.4 (2020): 1059-1086

Asgarzadeh, M., Fischer, D., Verma, S.K., Courtney, T.K., Christiani, D.C. The Impact of Weather, Road Surface, Time-of-Day, and Light Conditions on Severity of Bicycle-Motor Vehicle Crash Injuries. Am. J. Ind. Med., 61 (7) (2018), pp. 556-565

Asgarzadeh, M., Verma, S., Mekary, R.A., Courtney, T.K., Christiani, D.C. The Role of Intersection and Street Design on Severity of Bicycle-Motor Vehicle Crashes. Injury Prevention, (2017) Volume 23, page 179-185

Association of American Railroads, (2015) General Code of Operating Rules.

Association of American Railroads, (2020) High Tech Advances Improve Freight Railroad Safety & Efficiency, https://www.aar.org/wp-content/uploads/2020/11/AAR-High-Tech-Advancements-Fact-Sheet.pdf.

Behnood, A., Mannering, F.L. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. Analytic Methods in Crash Research, 12 (2016), pp. 1-17

Belin, Matts-Ake, Per Tillgren, and Evert Vedung. Vision Zero - A Road Safety Policy Innovation. International Journal of Injury Control and Safety Promotion 19, no. 2 (2012): 171-179.

Bentley, Rebecca, Damien Jolley, and Anne Marie Kavanagh. Local Environments as Determinants of Walking in Melbourne, Australia. Social Science & Medicine, (2010) 1806-1815.

Berdica, K. An Introduction to Road Vulnerability: What Has Been Done, Is Done and Should be Done. In Transport Policy, Vol. 9 No. 2, (2002) pp. 117-127.

Bigazzi, A., Ausri, F., Peddie, L., Fitch, D., Puterman, E. “Physiological markers of traffic-related stress during active travel,” Transportation Research Part F: Traffic Psychology and Behaviour, vol. 84, (2022), pp. 223–238

Buehler R, Pucher J. Cycling to work in 90 large American cities: New evidence on the role of bike paths and lanes. Transportation. (2011) 39(2):409-432

Chaudhari, A., Gore, N., Arkatkar, S., Joshi, G., Pulugurtha, S. Exploring pedestrian surrogate safety measures by road geometry at midblock crosswalks: a perspective under mixed traffic conditions IATSS Res., 45 (1) (2021), pp. 87-101, 10.1016/j.iatssr.2020.06.001

Chimba, D., Emaasit, D., Kutela, B. Likelihood Parameterization of Bicycle Crash Injury Severities. Journal of Transportation Technologies, (2012) 2, 213-219

Colorado Department of Transportation (CDOT). Strava Metro Data Analysis Summary (2018)

Dash, I., Abkowitz, M., Philip, C. Exposure Study for Pedestrians and Bicyclists in Urban Areas. Under review Journal of Safety Sciences (2022)

Dash, I., Abkowitz, M., Philip, C. Factors Impacting Bike Crash Severity in Urban Areas. in press. Journal of Safety Research (2022)

Dash, I., Abkowitz, M., Philip, C. Understanding Pedestrian Crash Severity in Urban Areas. Under review Journal of Safety Research (2022)

De Geus, B., Vandenbulcke, G., Panis, L.I., Thomas, I., Degraeuwe, B., Cumps, E., Meeusen, R. A Prospective Cohort Study on Minor Accidents Involving Commuter Cyclists in Belgium. In Accident Analysis & Prevention, Vol. 45, (2012) pp. 683-693.

DiMaggio, C., Frangos, S., Li, G.National safe routes to school program and risk of school-age pedestrian and bicyclist injury. Ann. Epidemiol., 26 (6) (2016), pp. 412-417

Dixon, K., C. Monsere, F. Xie, and K. Gladhill. Calibrating the Future Highway Safety Manual Predictive Methods for Oregon State Highways Publication No. OTREC-RR-12-02 (2012)

Dixon, K., R. Avelar, L. Brown, M. Mecham, and I. Van Schalkwyk. Quantifying Safety Performance of Driveways on State Highways Publication FHWA-OR-RD-13-02, FHWA, U.S Department of Transportation, (2012)

Dolatsara, H.A. Development of Safety Performance Functions for Non-Motorized Traffic Safety. (2014)

Dumbaugh E, Li, W. Designing for the safety of pedestrians, cyclists, and motorists in urban environments. Journal of the American Planning Association. (2010) 77(1):69-88

Eluru, N., C.R. Bhat, and D.A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. In Accident Analysis & Prevention, Vol. 40 No. 3, (2008) pp. 1033-1054.

Elvik R., Hoye A., Vaa T and Sorensen M, The Handbook of Road Safety Measures, 2nd ed. Oslo, Norway, (2009)

Environmental Protection Agency, sourced 2020, ALOHA Software, epa.gov/cameo/aloha-software.

Federal Highway Administration (FHWA) Bicycle-Pedestrian Count: Technology Pilot Report (2018)

Federal Highway Administration (FHWA) Exploring Pedestrian Counting Procedures (2017)

Federal Highway Administration (FHWA) Lighting Handbook (2012)

Federal Highway Administration (FHWA) Moving to a Complete Streets Design Model: A Report to Congress on Opportunities and Challenges (2022)

Federal Highway Administration (FHWA). Making Our Roads Safer - One Countermeasure at a Time (2021)

Federal Highway Administration (FHWA). Separated Bike Lane Planning and Design Guide (2015)

Federal Highway Administration. December 2016 Volume Traffic Trends. Washington D.C.: U.S. Department of Transportation (2016)

Federal Railroad Administration, Hazardous Materials Transportation, sourced 2020, railroads.dot.gov/program-areas/hazmat-transportation/hazardous-materials-transportation

Ferenchak, N. N., Marshall, W. E. "Bicycling facility inequalities and the causality dilemma with socioeconomic/sociodemographic change," Transportation Research Part D: Transport and Environment, vol. 97, (2021), p. 102920

Ferenchak, N. N., Marshall, W. E. "Equity Analysis of Proactively- vs. Reactively-Identified Traffic Safety Issues," Transportation Research Record, vol. 2673, no. 7, (2019), pp. 596–606

Ferenchak, N.N. and Marshall, W.E. "Is bicycling getting safer? Bicycle fatality rates (1985–2017) using four exposure metrics," Transportation Research Interdisciplinary Perspectives, Vol. 8, (2020) p. 100219

N. N. Ferenchak and Marshall, W.E., "The Relative (In)Effectiveness of Bicycle Sharrows on Ridership and Safety Outcomes," presented at the Transportation Research Board 95th Annual Meeting Transportation Research Board, 2016. https://trid.trb.org/view/1393928.

Ferenchak, N.N. and Marshall, W.E., "Advancing healthy cities through safer cycling: An examination of shared lane markings," International Journal of Transportation Science and Technology, vol. 8, no. 2, pp. 136–145, Jun. 2019, doi: 10.1016/j.ijtst.2018.12.003.

Ferster, Colin Jay, Trisalyn Nelson, Meghan Winters, and Karen Laberee. Geographic Age and Gender Representation in Volunteered Cycling Safety Data: A Case Study of BikeMaps.org. Applied Geography, (2017): 144-150.

Fields PhD, Billy, Jr. Bike Easy Audit Tool (BAT): A Bicycle Encouragement Evaluation. The Pedestrian Bicycle Resource Initiative at the University of New Orleans Transportation Institute, New Orleans: UNO Transportation Institute, (2013) 9.

Fridman, L., Ling, R., Rothman, L. et al. Effect of reducing the posted speed limit to 30 km per hour on pedestrian motor vehicle collisions in Toronto, Canada - a quasi-experimental, pre-post study. BMC Public Health 20, 56 (2020).

Furth, P.G., M.C. Mekuria, and H. Nixon. Network Connectivity for Low-Stress Bicycling. In Transportation Research Record: Journal of the Transportation Research Board, No. 2587, TRB, National Research Council, Washington D.C., (2016) pp. 41-49.

Haleem, K., and M. Abdel-Aty. Examining Traffic Crash Injury Severity at Unsignalized Intersections. In Journal of Safety Research, Vol. 41 No. 4, (2010) pp. 347-357.

Haleem, K., Alluri, P., Gan, A. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. Crash Analysis & Prevention, 81 (2015), pp. 14-23

Hamidun, R., Roslan, A., Sarani, R. Exploring factors for pedestrian fatalities at junctions in Malaysia. Pertanika Journal of Social Sciences and Humanities, 25 (4) (2017) pp. 1833-1840

Helak, K., Jehle, D., McNabb, D., Battisti, A., Sanford, S., & Lark, M. C. (2017) Factors influencing injury severity of bicyclists involved in crashes with motor vehicles: Bike lanes, alcohol, lighting, speed, and helmet use. Southern medical journal, 110(7), 441-444.

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

Huang, H., Abdel-Aty,M., Darwiche,A. County-level crash risk analysis in Florida: Bayesian spatial modeling. Transp. Res. Rec.: J. Transp. Res. Board, 2148 (2010), pp. 27-37

Hulse, J.V., Khoshgoftaar, T.M., Napolitano, A., Wald, R. Threshold-based feature selection techniques for high-dimensional bioinformatics data. Network Modeling Analysis in Health Informatics and Bioinformatics. 47 – 61 (2012)

Institute of Transportation Engineers (ITE). Designing Walkable Urban Thoroughfares: A Context Sensitive Approach (2010)

Isalsson-Hellman, Irene, Josefin Toreki. The effect of speed limit reductions in urban areas on cyclist injuries in collision with cars. Traffic Safety Prevention, Volume 20, (2019) - Issue sup3: 2018 International Cycling Safety Conference (ICSC). Pages 39-44

ITF. Road Safety Annual Report 2017. Paris: OECD Publishing, 2017, 581. Jacobsen, Peter L., and Harry Rutter. Cycling Safety. In City Cycling, by Ralph Buehler John Pucher, 141-156. Cambridge, Massachusetts: MIT Press (2012)

Jurewicz, C., Sobhani, A., Woolley, J., Dutschke, J., Corben, B. Exploration of vehicle impact speed – injury severity relationships for application in safer road design. Transportation Research Procedia 14 (2016), pp. 4247 – 4256

Kaya, N., Girgis, J., Hansma, B., Donmez, B. "Hey, watch where you're going! An on-road study of driver scanning failures towards pedestrians and cyclists," Accident Analysis & Prevention, vol. 162, (2021) p. 106380

Kaplan, S. and Prato, C.G. A spatial analysis of land use and network effects on frequency and severity of cyclist – motorist crashes in Copenhagen region. In Traffic Injury Prevention, Vol. 16, (2015), pp. 724-731

Khattak, A., Tung, L.-W. Severity of pedestrian crashes at highway-rail grade crossings. Transportation Research Board 92nd Annual Meeting, Washington DC, (2015)

Kim, J.K., S. Kim, G.F. Ulfarsson, and L.A. Porrello. Bicyclist Injury Severities in Bicycle– Motor Vehicle Accidents. In Accident Analysis & Prevention, Vol. 39 No. 2, (2007), pp. 238-251.

Klop, J.R., Khattak, A.J., Factors Influencing Bicycle Crash Severity on Two-Lane, Undivided Roadways in North Carolina. Transportation Research Record (1999) 1674(1):78-85.

Kong C., Yang J. Logistic regression analysis of pedestrian casualty risk in passenger vehicle crashes in China, Accident Analysis & Prevention Vol. 42 (2010) 987–993.

Lee, C., and Abdel-Aty, M. Comprehensive Analysis of Vehicle–Pedestrian Crashes at Intersections in Florida, Accident Analysis & Prevention, Vol. 37 No. 4 (2005) pp. 775-786.

Lee, J., Abdel-Aty, M., Jiang, X. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. In Accident Analysis & Prevention, Vol. 78 (2015), pp. 146-154

Li, D., Ranjitkar, P., Zhao, Y., Yi, H., Rashidi, S. Analyzing pedestrian crash injury severity under different weather conditions. Traffic Injury Prevention, 18 (4) (2017), pp. 427 430,

Li, Y., Song, L., Fan, Wei (David). Day-of-the-week variations and temporal instability of factors influencing pedestrian injury severity in pedestrian-vehicle crashes: A random parameters logit approach with heterogeneity in means and variances, Analytic Methods in Crash Research, Volume 29 (2021) 100152, ISSN 2213-6657.

Loftis, D. (2012) Wayside and On-Board Sensing. Available at: http://railcar-tech.org/wp-content/uploads/2018/10/2012-RSI_2012_Olin.pdf

Loftis, D. (2016) Coupling Speed Regulation, Coupling Speed - Damage Study. Available at: http://railcar-tech.org/wp-content/uploads/2018/11/Olin-Coupling-Speed-Study-RSI-Don-Loftis.pdf

Loftis, D. (2017) Managing Transportation Risks – Olin Corporation. Available at: https://www.mwrailshippers.com/wp-content/uploads/sites/2/2018/01/11-Managing-Transportation-Risks-MARS2018.pdf

Lu, J. Development of Safety Performance Functions for Safety Analyst Applications in Florida (2013)

Lundberg, S.M., Lee, S.I. A Unified Approach to Interpreting Model Predictions. 31st Conference on Neural Information Processing System (MIPS 2017)

Marshall, W.E., Garrick, N.W. Evidence on why bike-friendly cities are safer for all road users. Environmental Practice, 13 (1) (2011), pp. 16-27

Mokhtarimousavi, S. A time of day analysis of pedestrian-involved crashes in California: Investigation of injury severity, a logistic regression and machine learning approach using HSIS data Institute of Transportation Engineers. ITE Journal, 89 (10) (2019), pp. 25-33

Mokhtarimousavi, S., Anderson, J.C., Azizinamini, A., Hadi, M. Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and Artificial Neural Networks International Journal of Transportation Science and Technology (2020), pp. 1-16

Morgan, R., et al. (2006) System to Detect Truck Hunting on Freight Railroads. Available at: http://onlinepubs.trb.org/onlinepubs/archive/studies/idea/finalreports/safety/S-06_Final_Report.pdf

Mujalli, R.O., Garach, L., López, G., Al-Rousan, T. Evaluation of injury severity for pedestrian-vehicle crashes in Jordan using extracted rules Journal of Transportation Engineering Part A: Systems, 145 (7) (2019), pp. 1-13

Müller, A.C., Guido, S. Introduction to Machine Learning with Python: A Guide for Data Scientist, (2017) Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472

Müller, A.C., Guido, S. Introduction to Machine Learning with Python: A Guide for Data Scientist (2017)

National Association of City Transportation Officials (NACTO). Urban Street Design Guide (2013)

National Association of City Transportation Officials (NACTO). Urban Bikeway Design Guide (2011)

National Cooperative Highway Research Program (NCHRP). NCHRP Synthesis 436: Local Policies and Practices that Support Safe Pedestrian Environments: A Synthesis of Highway Practice. Transportation Research Board (2012)

National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation 2019 Traffic Safety Facts: A Compilation of Motor Vehicle Crash Data (Annual FARS/GES/CRSS Report)., Washington, D.C.: USDOT (2021)

National Institutes of Health, sourced 2020, Wireless Information System for Emergency Responders, https://wiser.nlm.nih.gov/.

National Transportation Safety Board (NTSB) (2019) Bicyclist Safety on US Roadways: Crash Risks and Countermeasures. Safety Research Report NTSB/SS-19/01. Washington, DC

National Cooperative Highway Research Program (NCHRP) Report 797 - Guidebook on Pedestrian and Bicycle Volume Data Collection. Transportation Research Board (2016)

Ngai, A. (2009) Wild Impact Load Detector. The Institution of Railway Signal Engineers (Hong Kong Section). Available at: http://irse.org.hk/eNewsletter/issue05/Technical-Articles/WILD/WILD.htm

National Highway Traffic Safety Administration (NHTSA) (2018) Traffic safety facts: Pedestrians. National Highway Traffic Safety Administration, Washington (DC). Report No.: DOT HS 812 493.

National Highway Traffic Safety Administration (NHTSA) (2021) Countermeasures that work: Pedestrians. National Highway Traffic Safety Administration, Washington (DC). Report No.: DOT HS 813 187.

National Highway Traffic Safety Administration (NHTSA) (2022) Motor Vehicle Traffic Fatalities and Fatality Rate by Sub-Categories in 2021. National Highway Traffic Safety Administration, Washington (DC). Report No.: DOT HS 813 298.

Nilsson, G. (2004) Traffic safety dimensions and the power model to describe the effect of speed on safety. Bulletin 221. Lund Institute of Technology, Lund, Sweden.

Noh, Y., Kim, M., Yoon, Y. Elderly pedestrian safety in a rapidly aging society—Commonality and diversity between the younger-old and older-old. Traffic Injury Prevention, 19 (8) (2018), pp. 874-879

Noland, R., and M. Quddus, M. Analysis of Pedestrian and Bicycle Casualties with Regional Panel Data. In Transportation Research Record: Journal of the Transportation Research Board, No. 1897, TRB, National Research Council, Washington D.C. (2004) pp. 28-33.

Nordback, K., W.E. Marshall, & B.N. Janson. Bicyclist Safety Performance Functions for a U.S. City. In Accident Analysis & Prevention, Vol. 65 (2014) pp. 114-122.

Obeng, K., Rokonuzzaman, S. Pedestrian injury severity in automobile crashes. Open Journal of Safety Science and Technology, 3 (2013), pp. 9-17

Oh, J., J. Jun, E. Kim, and M. Kim. Assessing Critical Factors Associated with Bicycle Collisions at Urban Signalized Intersections. Transportation Research Board, TRB, National Research Council, Washington D.C. (2008) pp. 1-17.

Overview of Motor vehicle Crashes in 2019. National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation, Washington, D.C.: USDOT (2020)

Pai, C.W. Overtaking, Rear-End, and Door Crashes Involving Bicycles: An Empirical Investigation. In Accident Analysis & Prevention, Vol. 43 No. 3 (2011) pp. 1228-1235.

Per E Gårder. The impact of speed and other variables on pedestrian safety in Maine. Crash Analysis & Prevention. Volume 36, Issue 4, (2004) Pages 533-542, ISSN 0001-4575.

Petritsch, T., B. Landis, H. Huang, and S. Challa(2006). Sidepath Safety Model: Bicycle Sidepath Design Factors Affecting Crash Rates. In Transportation Research Record: Journal of the Transportation Research Board, No. 1982, TRB, Washington D.C., pp. 194-201.

Pucher J, Buehler R, Bassett DR, Dannenberg AL. Walking and cycling to health: A comparative analysis of city, state, and international data. American Journal of Public Health (2010) 100 (10) : 1986–1992

Quddus, M. Effects of geodemographic profiles of drivers on their injury severity from traffic crashes using multilevel mixed-effects ordered logit model. Transportation Research Record: Journal of the Transportation Research Board, 2514 (2015), pp. 149-157

Quick Facts 2019. National Highway Traffic Safety Administration (NHTSA), U.S. Government Department of Transportation, D.C. (2020)

Rahman, Md Sharikur, Applying Machine Learning Techniques to Analyze the Pedestrian and Bicycle Crashes at the Macroscopic Level (2018)

Railinc, (2019) Equipment Health Management System (EHMS) User Guide. Available at: https://www.railinc.com/rportal/documents/18/260627/EHMS_UG.pdf

Rebentisch, H., Wasfi, R., Piatkowski, D. P., Manaugh, K. "Safe Streets for All? Analyzing Infrastructural Response to Pedestrian and Cyclist Crashes in New York City, 2009–2018," Transportation Research Record, vol. 2673, no. 2, (2019), pp. 672–685

Reynolds, C. C., Harris, M. A., Teschke, K., Cripton, P. A., & Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature. Environmental Health 8:47

Richards DC Relationship between speed and risk of fatal injury: pedestrians and car occupants, Road Safety Web Publication No. 16, Transport Research Laboratory, Department for Transport, London, (2010)

Rosén E., Sander U. Pedestrian fatality risk as a function of car impact speed, In Accident Analysis & Prevention, Vol. 41 (2009) 536–542.

Samerei, S.A., Aghabayk K., Shiwakoti, N., et al. Modelling bus-pedestrian crash severity in the state of Victoria, Australia. International Journal of Injury Control and Safety Promotion, 28 (2) (2021), pp. 233-242

Sasidharan, L., Wu, K.-F., Menendez, M. Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. Crash Analysis and Prevention, 85 (2015), pp. 219-228

Schepers, J.P., P.A. Kroeze, W. Sweers, and J.C. Wüst. Road Factors and Bicycle–Motor Vehicle Crashes at Unsignalized Priority Intersections. In Accident Analysis & Prevention, Vol. 43 No. 3 (2011) pp. 853-861.

Schepers, P., and B. den Brinker. What Do Cyclists Need to See to Avoid Single-Bicycle Crashes? In Ergonomics, Vol. 54 No. 4 (2011) pp. 315-327.

Schepers, P., M. Hagenzieker, R. Methorst, B. Van Wee, and F. Wegman. A Conceptual Framework for Road Safety and Mobility Applied to Cycling Safety. In Accident Analysis & Prevention, Vol. 62 (2014) pp. 331-340.

Schneider, R.J., Stefanich, J. Application of the Location–Movement Classification Method for Pedestrian and Pedestrian Crash Typing. Transportation Research Record: Journal of the Transportation Research Board, No. 2601 (2016) p. pp 72–83.

Shapley, L. S. "A Value for n-Person Games" In Contributions to the Theory of Games (AM-28), Volume II edited by Harold William Kuhn and Albert William Tucker, 307-318. Princeton: Princeton University Press, 2016 (1953).

Shinkle D, Teigen A. Encouraging Bicycling and Walking: The State Legislative Role. National Conference of State Legislatures (2008)

Shubo Wu, Quan Yuan, Zhongwei Yan, and Qing Xu. Analyzing Accident Injury Severity via an Extreme Gradient Boosting (XGBoost) Model. Journal of Advanced Transportation, Volume (2021) Article ID 3771640, 11 pages

Siddiqui, C., Abdel-Aty, M., Choi, K. Macroscopic spatial analysis of pedestrian and bicycle crashes. In Accident Analysis & Prevention, Vol.45 (2012) pp. 382-391.

Smart Growth America. Dangerous by Design 2020. Smart Growth America, 2020

Stewart, M.F., et al. (2019). An Implementation Guide for Wayside Detector Systems

Strauss, J., Miranda-Moreno, L.F., Morency, P. Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. In Accident Analysis & Prevention, Vol. 83 (2015) pp. 132-142

StreetLight Data. Practitioner’s Guide to Solving Transportation Safety, Safety Data Handbook (2021)

StreetLight Data. Bike and Pedestrian Metrics Methodology, Data Sources, and Validation (2020)

Stülpnagel, R., Petinaud, C., Lißner, S. "Crash risk and subjective risk perception during urban cycling: Accounting for cycling volume," Accident Analysis & Prevention, Vol. 164, (2022), p. 106470

Sze, N., Wong, S. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes Crash Anal. Prev., 39 (2007), pp. 1267-1278

Tefft, B.C. (2011). Impact Speed and a Pedestrian's Risk of Severe Injury or Death (Technical Report). Washington, DC: AAA Foundation for Traffic Safety.

Tennessee Department of Transportation (TDOT). Roadway Design Guidelines (2022)

Theofilatos, A., Efthymiou, D. Investigation of pedestrians' crash patterns in greater Athens area. Procedia-Social and Behavioral Sciences, 48 (2012), pp. 1897-1906

Traffic Safety Facts: Bicyclists and other Cyclists 2019 Data. National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation, Washington, D.C.: USDOT (2020)

Turner, S., Wood, G., Hughes, T., Singh, R. Safety performance functions for bicycle crashes in New Zealand and Australia Transportation Research Record: Journal of the Transportation Research Board No, 2236 (2011), pp. 66-73

United States Department of Transportation (USDOT). Policy Statement on Bicycle and Pedestrian Accommodation Regulations and Recommendations (2010)

United States Department of Transportation (USDOT). Federal Railroad Administration. Office of Research, Development, and Technology. Available at: https://railroads.dot.gov/elibrary/implementation-guide-wayside-detector-systems
United States Department of Transportation, Bureau of Transportation Statistics, National Transportation Statistics (2021).

Vandenbulcke, G., Thomas, I. and Panis, L.I. Predicting Cycling Accident Risk in Brussels: A Spatial Case–Control Approach. In Accident Analysis & Prevention, Vol. 62 (2014) pp. 341-357.

Verzosa, N., Miles, R. Severity of road crashes involving pedestrians in Metro Manila, Philippines. Crash Analysis & Prevention, 94 (2016), pp. 216-226

Vision Zero. Mayor Cooper Commits to Vision Zero to Eliminate Traffic Fatalities and Severe Injuries in Nashville, Mayor's Office, Media Release, January 18, 2020

Von Stülpnagel, R., C. Petinaud, and S. Lißner, "Crash risk and subjective risk perception during urban cycling: Accounting for cycling volume," Accident Analysis & Prevention, vol. 164, p. 106470, Jan. 2022, doi: 10.1016/j.aap.2021.106470.

Wang, Y. and Nihan, N.L. Estimating the Risk of Collisions between Bicycles and Motor Vehicles at Signalized Intersections. In Accident Analysis & Prevention, Vol. 36 No. 3 (2004) pp. 313-321.

Wang, C., Lu, L., Lu, J. Statistical Analysis of Bicyclists' Injury Severity at Unsignalized Intersections, Traffic Injury Prevention, Vol. 16(5) (2015) pp. 507-512.

Wanvik, P.O. Effects of road lighting: an analysis based on Dutch crash statistics 1987–2006. Crash Analysis & Prevention, 41 (1) (2009), pp. 123-128

Washington State Department of Transportation (WSDOT). Washington State Bicycle Facilities and Pedestrian Walkways Plan (2008)

Wegman, F.C.M., Aarts ,L.T., Bax,C. Advancing Sustainable Safety: National Road Safety Outlook for the Netherlands for 2005-2020. Leidschendam: SWOV Institute for Road Safety Research (2006)

Welle, B., Sharpin, A.B., Adriazola-Steil, C., Bhatt, A., Alveano, S., Obelheiro, M., Imamoglu, C.T., Job, S., Shotten, M., Bose, D. Sustainable and Safe: A Vision and Guidance for Zero Road Deaths. Ross Center, World Resources Institute, Washington D.C.: World Resources Institute (2018) 70

Welle, B., Liu, Q., Li, W., Steil, C.A., King, R., Sarmiento, C., Obelheiro, M. Cities Safer by Design: Guidance and Examples to Promote Traffic Safety through Urban and Street Design. Work, WRI Ross Center for Sustainable Cities, World Resources Institute, Washington D.C.: World Resources Institute (2015)

Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R. An Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications for Land Use and Transportation Planning. In Accident Analysis & Prevention, Vol. 41 No. 1 (2009) pp. 137-145.

Xin, C., Guo, R., Wang, Z., Lu, Q., Lin, P. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: A random parameters generalized ordered probability model with heterogeneity in means and variances. Analytic Methods in Crash Research, 16 (2017), pp. 117-132

Yan,X., Ma,M., Huang,H., Abdel-Aty, M., Wu,C. Motor Vehicle-Bicycle Crashes in Beijing: Irregular Maneuvers, Crash Patters, and Injury Severity. In Accident Analysis & Prevention, Vol. 43 No. 5 (2011) pp. 1751-1758.

Yu, C.-Y., Zhu, X. Planning for safe schools: impacts of school siting and surrounding environments on traffic safety. Journal of Planning Education and Research (2015), pp. 1-11

Zafri, N., Prithul, A.A., Baral, I., Rahman, M. Exploring the factors influencing pedestrian-vehicle crash severity in Dhaka, Bangladesh. International Journal of Injury Control and Safety Promotion, 27 (3) (2020), pp. 300-307

Zangenehpour, S., Strauss, J., Miranda-Moreno, L.F., Saunier, N. Are signalized intersections with cycle tracks safer? A case-control study based on automated surrogate safety analysis using video data. Accident Analysis & Prevention. Vol 86 (2016) pp. 161-172