Identification of Crash Contributing Factors in AV involved Crashes

MD TANVIR ASHRAF
ma00001@mix.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Part of the Civil Engineering Commons, and the Transportation Engineering Commons

Recommended Citation
ASHRAF, MD TANVIR, "Identification of Crash Contributing Factors in AV involved Crashes" (2021). Graduate Theses, Dissertations, and Problem Reports. 8271.
https://researchrepository.wvu.edu/etd/8271

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.
Identification of Crash Contributing Factors in AV involved Crashes

Md Tanvir Ashraf

Thesis submitted
to the Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University

in partial fulfillment of the requirements for the degree of
Master of Science
in
Civil Engineering

Kakan C. Dey, Ph.D., Chair
David R. Martinelli, Ph.D.
V. Dimitra Pyrialakou, Ph.D.

Department of Civil and Environmental Engineering
Morgantown, West Virginia
2021

Keywords: Decision tree, Association rule, AV, autonomous mode, non-autonomous mode.

Copyright 2021 Md Tanvir Ashraf
ABSTRACT

Identification of Crash Contributing Factors in AV involved Crashes

Md Tanvir Ashraf

Autonomous Vehicles (AVs) can dramatically reduce the number of traffic crashes and associated fatalities by eliminating the avoidable human-errors related crash contributing factors. Many companies have been conducting pilot tests on public roads in several states in the United States and other countries to accelerate AV mass deployment. AV pilot operations on California public roads saw 251 AV-involved crashes (as of February 2020). These AV-involved crashes provide a unique opportunity to investigate AV crash risks in a mixed traffic environment. This study collected the AV crash reports from the California Department of Motor Vehicles and applied the Decision Tree (DT), and Association Rule methods to extract the pre-crash rules of AV-involved crashes. Extracted rules revealed that the most frequent types of AV crashes were rear-end crashes and predominantly occurred at intersections when AVs were stopped and engaged in the autonomous mode. AV and Non-AV manufacturers, and transportation agencies can use the findings of this study to minimize AV-related crashes. AV companies could install a distinct signal/display to inform the operational mode of the AVs (i.e., autonomous or non-autonomous) to human drivers around them. Moreover, the Automatic Emergency Braking system in non-AVs could avoid a significant number of rear-end crashes as often rear-end crashes occurred due to the failure of following non-AVs timely slow down behind AVs. Transportation agencies can consider separating the AVs from the non-AVs by assigning “AV only lanes” to eliminate the excessive rear-end crashes due to the mistakes of human drivers in non-AVs at the intersections.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr. Kakan Dey, for his continuous and valuable guidance, inspiration, and support throughout my MS study period and thesis research work. Without his active support and motivation, it would not have been possible for me to complete this research work. Besides my advisor, I would also like to Dr. David Martinelli and Dr. Dimitra Pyrialakou, for serving on my thesis committee, and providing valuable feedback.

I am also thankful for the wonderful research environment and institutional supports provided by West Virginia University, and Dr. Dey’s research group where I had the chance to grow as a researcher.

Finally, I would like to dedicate my thesis to my parents whose unconditional love and support made me who I am today. Also, I would like to thank my wife Khosnur Jahan Aurin for her immense support and encouragement during the past two years which helped me to navigate the ups and downs of my graduate student life.
Table of Contents

ABSTRACT................................................................................................................................... ii
ACKNOWLEDGEMENTS ........................................................................................................ iii
CHAPTER 1: INTRODUCTION.......................................................................................... 1
CHAPTER 2: LITERATURE REVIEW........................................................................... 3
CHAPTER 3: METHODOLOGY...................................................................................... 5
   3.1 Decision Tree (DT) Model ...................................................................................... 5
   3.2 Association Rule Data Mining ............................................................................... 7
CHAPTER 4: DATA DESCRIPTION............................................................................ 10
   4.1 Study Area ............................................................................................................. 10
   4.2 CA DMV AV CRASH DATA ................................................................................ 11
CHAPTER 5: RESULT AND DISCUSSION................................................................. 16
   5.1 Results of the DT model ...................................................................................... 16
   5.2 Rules Generated from Association Rules ............................................................. 21
   5.3 Implications of the research findings ................................................................. 23
CHAPTER 6: CONCLUSION....................................................................................... 24
DECLARATION............................................................................................................. 26
REFERENCES.................................................................................................................... 26
List of Figures

Figure 1: Locations of AV crashes in California (2016-February 2020) .................................................... 10
Figure 2: AV crashes in CA, by year ........................................................................................................ 12
Figure 3: Company-wise Number of AV crashes ................................................................................. 13
Figure 4: AV miles per AV crash in different years for Waymo and GM Cruise LLC ....................... 13
Figure 5: AV operational mode at the time of crashes ........................................................................ 15
Figure 6: DT of AV involved crashes, crash types as response variable (TM- Turning Movement, PS-Proceeding Straight) ........................................................................................................ 19
List of Tables

Table 1: Summary Statistics of the AV crashes and Explanatory Variables........................................... 11

Table 2: Extracted rules from the DT model .......................................................................................... 20

Table 3: Rules Generated from Association Rule Method based on Crash Type ............................... 21
CHAPTER 1: INTRODUCTION

Traffic crash-related fatality has been one of the major causes of death in the U.S. and around the world. In 2018, thirty-seven thousand people died on the US roadways due to traffic crashes and caused an economic loss of $147 billion (Webb, 2018). Transportation agencies in the U.S. have developed “Vision Zero” initiatives to eliminate all traffic fatalities and severe injuries (Fagnant and Kockelman, 2015). Over 90 percent of all crashes happen due to driver errors, and over 40 percent of fatal crashes happen due to some combination of alcohol, distraction, drug involvement, and/or fatigue. As most crashes occur due to driver errors, emerging autonomous vehicles (AVs) have the potential to eliminate driver errors and associated traffic crashes and help agencies to achieve the Vision Zero goal. Besides, AV can reduce congestion and fuel consumption and can improve mobility performance by reducing vehicle ownership and providing mobility services to elderly and disabled people (Fagnant and Kockelman, 2015). Since September 2014, the California Department of Motor Vehicles (CA DMV) allowed permit-holding companies to test AV technologies on the public roadways. Human test/safety drivers in these AVs are responsible for disengaging from autonomous mode to manual mode when the autonomous mode faces any technical issues or approaches an unfamiliar driving environment. While simple/typical driving actions (e.g., maintaining position on a lane) are relatively easy for an AV to operate, designing an autonomous system that can perform safely in nearly every situation is challenging (Campbell et al., 2010). Navigating complex situations such as intersections where the number of potential conflict points is high poses a great challenge for the AVs. Testing companies operating in CA are required to submit a Traffic Collision Involving an Autonomous Vehicle Report (OL 316 Form) to CA DMV. From September 2014 to February 2020, 251 AV-related crashes occurred in California.

Ensuring the safety of AV technology development is a complex challenge, and the improvement of AV technology depends on addressing the wide array of risk factors such as the behavior of non-AV drivers, driving environments, and AV technology limitations. To minimize the crash risk of AVs, AV technology has to detect and predict unsafe movements of other vehicles and roadway users in complex driving environments. CA AV crash reports provide a unique opportunity to identify the challenges faced by AVs during the pilot operation on public roads and develop
potential solutions for AV technology improvement. This study compiled the CA AV crash reports to analyze the crash contributing factors that led to rear-end, sideswipe, broadside, or head-on types of AV-involved crashes. The results of this study developed new insights on the safety performance of AVs in real-world driving environments and potential measures that can be implemented by different AV stakeholders to reduce AV safety deficiencies.

The structure of the paper is as follows: the literature review section provides a comprehensive overview of the past studies focused on AV crashes, followed by the gap in the existing literature and how this study addresses the gap. The next four sections explain the AV-related crash data, the adopted research method for the crash rules extraction, the results of the rules extraction, and potential implications of the findings. The last section presents concluding remarks.
CHAPTER 2: LITERATURE REVIEW

Since the beginning of testing on public roadways, AV technology has made significant progress. An AV uses a combination of hardware and software technologies such as cameras, a Global Positioning System (GPS), a Radio Detection and Ranging (RADAR) sensor, Light Detection and Ranging (LiDAR) sensor, digital maps, video camera, and computing platform to sense the driving environment and determine AV movement actions. The number of companies conducting AV testing on public roads has been increasing, and many of the tests have been performed in a closed environment (i.e., test tracks/proving grounds) (Wang and Li, 2019). Many companies and researchers have been using driving simulators as a tool for AV technology development. Several studies investigated the driver’s behavior and performance in a highly automated driving system (Körber et al., 2016; De Winter et al., 2014; Merat et al., 2014). Complex driving environments could lead to a manual takeover of the AV by the safety driver, which is known as “Disengagements” of AV technology. In addition to the disengagement reports, CA DMV maintains a separate database for all AV-involved crash reports since 2014. AV crash reports document some important crash information such as manufacturer’s information, crash information, serious injuries to people, other associated factors such as weather, lighting, and pavement condition (Wang and Li, 2019). As of February 2020, the DMV has received 251 Autonomous Vehicle Collision Reports. Prediction and estimation of traffic crashes and crash contributing factors have been an important research topic for decades. The traditional approach for non-AV crash modeling is to develop statistical models to explore the underlying causes of the crashes (Fountas, Anastasopoulos and Abdel-Aty, 2018; Iranitalab and Khattak, 2017; Xie, Lord and Zhang, 2007). Besides these statistical modeling approaches, researchers also used different Machine Learning (ML) and Neural Network methods to analyze traffic crashes (Huang et al., 2018; Riviere et al., 2006; Abdelwahab and Abdel-Aty, 2002; Mussone, Ferrari and Oneta, 1999). Many studies applied non-parametric methods such as Decision Trees (DT) or Association Rule to identify the crash contributing factors. For example, Montella et al. (2012) used the Classification and Regression Tree (CART) model and Association rules to generate rules for powered two-wheeler crashes in Italy (Montella et al., 2012). De Oña et al. (2013) and Abellán et al. (2013) used tree-based methods to analyze the crash severity and crash types (De Oña et al., 2013; Abellán et al., 2013).
The availability of the AV crash database is very limited, and only CA DMV has released AV crash data to the public. Few researchers explored the mechanism of AV crashes using the CA AV crash data. Previous research on AV disengagements and crashes focused on the exploratory analysis of the crash data to find the relation between different contributing factors. Favaro et al. (2017) examined the factors of AV crashes by exploring the drivers’ engagement at pre-and post-crash scenarios, the relative speed of the involved vehicles, and the location of the crashes (Favaro et al., 2018). Boggs et al. (2020) conducted a statistical analysis of the CA AV rear-end types of crashes and injury crashes using text mining techniques and a hierarchical Bayesian heterogeneity-based approach (Boggs et al., 2020). The chance of rear-end crash was higher in situations when the AV is driven in an autonomous mode compared to the situations when the AV was driven in non-autonomous mode. Wang and Li explored the AV crash severity and crash types by applying the ordinal logistic and CART classification tree models. Several studies also explored the AV disengagement data (Wang and Li, 2019). Favarò et al. (2018) analyzed the disengagement triggers, trends, and contributory factors, and found that 1 in every 178 disengagements resulting in a crash (Favarò et al., 2018).

A key gap in the limited AV-related crash studies is the comprehensive analysis of the AV and Non-AV movements in the mixed traffic conditions that resulted in different types of AV-involved crashes. This study investigated the AV and Non-AV movements as well as other pre-crash factors (e.g. driving mode, location of the crash, etc.) in AV involved rear-end, sideswipe, broadside, and head-on crashes. Previous literature on non-AV crashes suggests that analyzing crash types have three primary benefits: 1) identification of locations deficiency to specific crash types which are not likely to be revealed by total crash model 2) implementation of the improved and efficient type of countermeasures as there are target crashes revealed through crash types analysis, 3) better development of crash predictive models as the association of individual crash types with various covariates such as roadway geometry, traffic volume, and environmental factors (Sarker et al., 2017; Mishra and Zhu, 2015; Geedipally, Patil and Lord, 2010; Ye et al, 2009; Kim, Washington and Oh, 2006). Thus, in this study, the Decision Tree (DT) and Association Rule method model were applied to generate and identify major combinations of crash contributing factors in AV-involved crashes based on model performance criteria.
CHAPTER 3: METHODOLOGY

Each AV-involved crash was the result of a combination of crash contributing factors, and each crash can be expressed as a logical combination of crash contributing factors. These conditional statements are called crash rules. In this study, rules were extracted by applying two different techniques—Decision Tree (DT) and Association Rule methods. Both methods are non-parametric and provide complementary insights on AV-involved crashes. The non-parametric method such as DT and association rules are better suited for this study compared to other parametric methods as the AV crash database has a relatively small sample size and the CA DMV AV crash database provided only a few factors associated with these crashes. The results from the DT and association rules provide intuitive and direct correlation factors that could lead to the appropriate mitigation measures to improve the safety performance of AV during the critical ongoing technology development phase. Thus, the rules generation by DT and association rules identified a set of important rules in the AV crash database, which can be used by AV safety analysts to improve the safety of the AV technology.

In the DT method, each terminal node of the tree represents a rule, with all the preceding nodes that are connected to the terminal node acting as the crash contributing factors. Each terminal node of the tree gives the probability of each crash type. On the other hand, association rules are based on the relative frequency of the combination of factors that occurred in the crash database. The rules can have crash types such as rear-end, sideswipe, broadside, or head-on as consequent. The explanatory variables used for the rule extraction were driving mode of AVs, disengagement of the autonomous mode, location of the crash, and movement of AVs and Non-AVs at the moment of the crash.

3.1 Decision Tree (DT) Model

A Decision Tree (DT) is a non-parametric predictive model that does not require prior probabilistic knowledge on the study phenomena (de Oña et al., 2013). DT can be used for classification and regression modeling purposes. If the dependent variable in the DT is a class variable (e.g., crash type) and has a limited number of possible options, then it is called classification. DT represents a simplistic and graphically hierarchical structure and is used to extract rules and to identify the
effect of different explanatory variables (i.e., crash contributing factors). In the hierarchical structure of a DT, each node represents a feature, and each branch represents one of the conditions of that feature variable. The terminal node of the tree, known as the tree leaf, represents the expected value of the class variable. Each node is associated with the most informative variable that has not already been selected in the path from the root to that node.

Several algorithms can be used to build a DT. The splitting criteria of the tree formation are the main difference between these methods (de Oña et al., 2013). Among all algorithms, the Classification and Regression Tree (CART) model developed by Breiman et al. is the most widely used method in traffic crash data analysis (Breiman et al., 1984). In this study, the CART algorithm was used to build the DT in the JMP pro program. The CART model uses the Gini index as splitting criteria, which measures the degree of impurity at each node until their purity cannot be increased by splitting further. The Gini Index for a variable C (for example Operational Mode of the AVs: Autonomous or non-autonomous) can be defined as the following Equation 1:

$$gini(C) = 1 - \sum_j p^2(C) ; C = c_j$$

Where p is the probability of a variable C being classified to a particular class c_j at a node of the DT (e.g. for the variable “driving mode” p is the probability of being the driving mode autonomous or non-autonomous at that node). The split criteria based on the Gini Index can be defined as Equation 2:

$$GIx(C, X) = gini(C|X) - gini(X)$$

Where $gini(C|X) = \sum_t P(X_t) gini(C|X = x_t)$ and X is another known independent variable (for example, the location of the AV crashes). The best split minimizes the GIx (C, X) (de Oña et al., 2013). Similar to other classification methods, the performance of the CART model is measured by accuracy, which is defined as the percentage of cases correctly classified by the classifier.

A decision rule is a logical conditional expression that states the sequence of the events that lead to the occurrence of a particular class. It takes the form “IF A THEN B” where A is the antecedent, and B is the consequent. The rules generated from a DT depend on the direction of the variable at
the root node, where the conditional IF structure begins. To separate the rules that are significant, three parameters (i.e., support (S), Population (Po), and Probability (P)) were used in this study. The support is the percentage of the dataset where A and B occur, the population is the percentage of the dataset where A occurs, and Probability P is the percentage where the rule is accurate. The minimum value of these parameters depends on the nature of the dataset (balanced or unbalanced), rare events, and the size of the sample. For example, Montella et al. (2012) used a low value of probability (P) because their data was unbalanced, whereas De Oña et al. (2013) used a high value of P as their data was balanced (de Oña et al., 2013; Montella et al., 2012). As the data used in this study contains a limited number of crash events, and the data is unbalanced, a probability value of 40% as a minimum value was used to separate the significant rules.

3.2 Association Rule Data Mining

Association rule is one of the most popular non-parametric data mining techniques that can be used in traffic safety causality analysis. The main advantage of the association rule data mining method is that the dependent variable does not need to follow a specific functional distribution to generate the rules. Generated association rules can be used to develop countermeasures to break the association between the contributing factors to reduce the crash risk (Weng et al., 2016). The association rule discovery is usually done in two steps. In the first step, factors that are frequent (e.g., AV movements, locations) and have a support value greater than the threshold minimum value are selected. Factors are divided into two types- antecedent and consequent. From the list of contributing factors in a crash, one factor is set as consequent, and others are set as antecedents and are expressed as conditional forms called rules. Then, the confidence value of the rules (generated in the first step) is used to determine the significance of the rules. Next, one item is subtracted from the antecedent and added to the consequent to check the confidence value of the new rules. This process is iterated until the antecedent becomes empty. The second step of the association rule discovery method is more straightforward. In this step, rules generated in the first step are organized based on their lift values (AL-Zawaidah, Jbara and Marwan; 2011).

Several algorithms can be used for developing association rules. Apriori and Eclat are the two most commonly used algorithms by researchers over the years. The Apriori algorithm was first introduced by Agarwal et al. (1993). In this study, the Apriori algorithm is used for mining the
association rules from the CA DMV AV crash database using the program R. The working principle of this algorithm is discussed below:

Let, \( I = \{i_1, i_2, i_2, \ldots, i_n\} \) be a set of \( n \) attributes called items such as driving mode, lighting condition, vehicle movement type. and \( D = \{c_1, c_2, c_2, \ldots, c_m\} \) contains \( m \) numbers of AV crash causality data. Each crash event in \( D \) occurred due to some unique combination of attributes from \( I \) and has a unique id (i.e., crash ID). So, each causality in \( D \) contains a subset of the items in \( I \). A rule is defined as an implication of the form \( X \Rightarrow Y \) where \( X, Y \subseteq I \), and \( X \cap Y = \emptyset \). The part of an association rule (\( X \)) is antecedent(s), while the second part (\( Y \)) is consequent. In this study, crash types (i.e., rear-end crash, sideswipe crash, broadside, and head-on) were considered as consequent, and different crash contributing factors such as the location of the crash, driving mode were considered as antecedents. To separate significant rules from the whole set of rules generated by the Apriori algorithm, various measures of interest and significance can be used. The best-known measures are minimum thresholds on support and confidence (Hornik, Grün and Hahsler, 2005). The support of a rule is the percentage of the entire dataset covered by the rule and expressed as **Equation 3:**

\[
\text{Support} \ (X \Rightarrow Y) = \frac{\#(X \cap Y)}{N} \tag{3}
\]

Where \( N \) is the total number of AV crash events. The confidence of a rule is the conditional probability of occurrence of the consequent, given that the antecedent(s) have occurred and is expressed as **Equation 4,**

\[
\text{Confidence} = \frac{\text{Support} \ (X \Rightarrow Y)}{\text{Support} \ (X)} \tag{4}
\]

In addition, the lift value was used to measure the degree of interdependence among antecedents and consequents of the rule. The lift measures the ratio of confidence of a rule and the expected confidence that the consequent occurs depending on the occurrence of the antecedents, expressed as **Equation 5,**

\[
\text{Lift} = \frac{\text{Support} \ (X \Rightarrow Y)}{\text{Support} \ (X) \times \text{Support} \ (Y)} \tag{5}
\]
The lift value shows the frequency of the occurrence of antecedent and consequent (Weng et al., 2016). Lift value of a rule less than 1 indicates negative interdependence between the antecedent and the consequent, and greater than 1 indicates that the antecedent and the consequents are positively interrelated. If the lift value is equal to one, then the antecedent and consequents are independent (Montella et al., 2011). A large confidence and a lift value greater than one are desirable for the rules to have a high level of support. Moreover, lift value is more important for determining the strength of an association rule than support and confidence.
CHAPTER 4: DATA DESCRIPTION

4.1 Study Area

Although several states in the US have allowed testing AVs on public roadways, only California (CA) Department of Motor Vehicle (DMV) made the AV crash reports publicly available. Most of the AV technology companies in CA have been operating AVs in San Francisco and Santa Clara areas. The AV operators are required to fill form OL 316 for each AV-involved crash in CA. The AV crash locations in CA are shown in Figure 1.

Figure 1: Locations of AV crashes in California (2016-February 2020)
4.2 CA DMV AV CRASH DATA

One hundred and ninety-eight AV-involved crashes that occurred between January 2016 to February 2020 were used in this study. A crash database was developed for this study using the crash reports. A summary statistic of different characteristics of the AV crashes is provided in Table 1.

Table 1: Summary Statistics of the AV crashes and Explanatory Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Category</th>
<th>No.</th>
<th>Percentage (%)</th>
<th>Variables</th>
<th>Category</th>
<th>No.</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td></td>
<td></td>
<td></td>
<td>Lighting Condition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alameda</td>
<td></td>
<td>2</td>
<td>1.0</td>
<td>Dark-Street Light</td>
<td>30</td>
<td>15.2</td>
<td></td>
</tr>
<tr>
<td>Contra Costa</td>
<td></td>
<td>1</td>
<td>0.5</td>
<td>Daylight</td>
<td>121</td>
<td>61.1</td>
<td></td>
</tr>
<tr>
<td>San Francisco</td>
<td></td>
<td>124</td>
<td>62.6</td>
<td>Dusk/Dawn</td>
<td>5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Santa Clara</td>
<td></td>
<td>71</td>
<td>35.9</td>
<td>N/A</td>
<td>42</td>
<td>21.2</td>
<td></td>
</tr>
<tr>
<td>Crash Types</td>
<td></td>
<td></td>
<td></td>
<td>Crash Severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad Side</td>
<td></td>
<td>15</td>
<td>7.6</td>
<td>None</td>
<td>19</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td>Head-On</td>
<td></td>
<td>17</td>
<td>8.6</td>
<td>Moderate</td>
<td>21</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>Side Swipe</td>
<td></td>
<td>40</td>
<td>20.20</td>
<td>Minor</td>
<td>135</td>
<td>68.2</td>
<td></td>
</tr>
<tr>
<td>Rear End</td>
<td></td>
<td>122</td>
<td>61.6</td>
<td>Major</td>
<td>1</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td>4</td>
<td>2</td>
<td>AV Movement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stopped</td>
<td>78</td>
<td>39.4</td>
<td></td>
</tr>
<tr>
<td>Operational Mode</td>
<td></td>
<td></td>
<td></td>
<td>Slowing/Stopping</td>
<td>22</td>
<td>11.1</td>
<td></td>
</tr>
<tr>
<td>Non-Autonomous</td>
<td></td>
<td>55</td>
<td>28</td>
<td>Proceeding Straight</td>
<td>75</td>
<td>37.9</td>
<td></td>
</tr>
<tr>
<td>Autonomous</td>
<td></td>
<td>114</td>
<td>57</td>
<td>Turning Movement</td>
<td>17</td>
<td>8.6</td>
<td></td>
</tr>
<tr>
<td>Disengaged</td>
<td></td>
<td>29</td>
<td>15</td>
<td>Non-AV Movement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stopped</td>
<td>6</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Location of the Crash</td>
<td></td>
<td></td>
<td></td>
<td>Slowing/Stopping</td>
<td>14</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>Ramp</td>
<td></td>
<td>5</td>
<td>2.5</td>
<td>Proceeding Straight</td>
<td>128</td>
<td>64.6</td>
<td></td>
</tr>
<tr>
<td>Intersection</td>
<td></td>
<td>126</td>
<td>63.6</td>
<td>Turning Movement</td>
<td>31</td>
<td>15.7</td>
<td></td>
</tr>
<tr>
<td>In Roadway</td>
<td></td>
<td>61</td>
<td>30.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parking Lot</td>
<td></td>
<td>6</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.1 AV Crash by Year

The CA DMV allowed AV manufacturers to test their AV technologies on the public roadways by establishing an Autonomous Vehicle Tester Program in 2014. Permit holders are required to comply with California DMV regulations in order to participate in the program, as stated in
California Vehicle Code Section 38750. The two primary requirements of the program are - manufacturers must report all AV crashes through the OL 316 form and must fully retain all the details. From 2016 to February 2020, a total number of 198 AV crashes occurred in CA. The year-wise distribution of the AV crashes is shown in Figure 2.

**Figure 2: AV crashes in CA by year**

### 4.2.2 AV Crash by Company

As of January 2020, 64 companies have valid permits to test autonomous vehicles on CA public roadways with a safety driver, and one company has the permit for safety driver-less testing (CA DMV, 2020). Fourteen companies have reported 198 crashes from 2016 to February 2020. Among the companies that reported AV crashes, GM Cruise, LLC has the highest number of crashes, which is about 52% of total AV crashes (Figure 3). Google/Waymo started testing AVs in 2015, reported 60 AV crashes from 2016 to February 2020, which is almost 31% of total AV crashes. Another company, named Zoox Inc. had 17 AV crashes during this period. Other companies such as Lyft, Toyota, Nissan, Apple, Aurora had relatively fewer AV crashes during this period.
4.2.3 Autonomous miles driven and AV Crashes

Figure 4 shows the autonomous mile driven per AV crash by Waymo LLC and GM Cruise LLC. After 2016, the autonomous miles per AV crash was increasing for Waymo LLC. However, in 2019 the autonomous miles per AV crash decreased for this company. On the other hand for GM Cruise LLC, the autonomous miles driven per AV crash showed a gradually increasing trend which indicates a gradual increase in their AV technology safety.
4.2.4 Crash Types and AV Movement
Four main types of AV-related crashes (rear-end, sideswipe, broadside, and head-on) were considered in this study. Non-AVs hitting the rear-side of the AVs were defined as rear-end crashes. For this study, all the rear-end crashes happened when the AV was in front of the non-AVs. The summary statistics in Table 1 show that about sixty-two percent of the AV crashes were rear-end crashes. Most of these rear-end crashes happened when the AVs were stopped or slowing down to a full stop at intersections. The second common AV crash type was the sideswipe crash. Sideswipe crashes were defined as the crashes where the non-AVs impacted the side of the AVs and vice versa. Forty percent of the sideswipe crashes happened when the AV and non-AVs were moving in the same direction (Table 1). Broadside crashes were defined as crashes when the side of the AVs was impacted by the non-AVs at an angle. In this study, all the broadside crashes happened when non-AVs hit the AVs. Fifteen crashes in the crash database were broadside crashes, which was only 7.6% of the total number of AV-involved crashes (Table 1). Broadside crashes occurred when AVs were traveling straight at an intersection and hit by a non-AV making a left-turn or right-turn. Lastly, the head-on was the type of crashes when the front of AVs was impacted by the non-AVs, or the AVs struck other vehicles or objects. In the AV crash database, only 8.6% of the total crashes were head-on crashes (Table 1). The AV movement data in Table 1 shows that the most frequent AV movements during the crashes were “AV Proceeding Straight” and “AV Stopped”. On the contrary, the predominant non-AV movements were “Proceeding Straight” (64.6%) and Turning Movements (15.7%), as in most of the cases, AVs were hit by the moving non-AVs (Table 1).

4.2.5 Crash Severity
CA AV crash reports classify crash severity into four types (Major, Minor, Moderate, and None) based on the types of damage sustained by AVs. In most frequent rear-end crashes, the relative speed of the AV and non-AV were very low (i.e., less than 10 mph). As a result, most of the rear-end crashes caused minor damage to the AVs, except for one rear-end crash with major damage to the AV. Twenty-one crashes out of 198 AV crashes experienced moderate damage to the AVs, which was only 10.6% of total crashes (Table 1).
### 4.2.6 AV Driving Mode

AV manufacturers are required to report the driving mode (i.e., autonomous or non-autonomous mode) for each AV crash event. Crash reports showed that 57% of the crashes occurred when the AV was operating in autonomous mode. In some instances, AVs were disengaged to non-autonomous mode right before the crashes. These disengagements generally happened when the autonomous technology failed or when the safety driver identified an issue and took over the AV operation. The AV operation mode distribution found in the AV crashes is shown in Figure 5. Fifteen percent of the AV crashes occurred right after the autonomous mode was disengaged, and 28% of the crashes occurred when the AVs were driving in non-autonomous/conventional mode (Figure 5).

![Figure 5: AV operational mode at the time of crashes](image)
CHAPTER 5: RESULT AND DISCUSSION

5.1 Results of the DT model

The DT was developed using the CART model, where 70% of the crash events were used for training the model, and 30% of the data was used to validate the model. The accuracy of the CART model using the training dataset and validation dataset was about 69% and 65%, respectively. The observed model accuracy was similar to the previous studies that focused on crash severity modeling and identified the associated crash contributing factors, where we focused on crash type modeling in this study. De Oña et al. (2013) used the DT method to extract rules from the police accident reports, where the accuracy of the CART model was 55.87% (De Oña et al., 2013). Abdel Wahab and Abdel-Aty (2001), De Oña et al. (2011) also reported similar accuracy on their crash severity modeling studies using data mining and Bayesian networks, respectively. Figure 6 shows the developed DT for AV-involved crashes compiled for this study.

“Non-AV Movement” was the root node that generated the DT, which was then divided into two branch nodes (node 1 and node 2), as shown in Figure 6. Each node was then further divided into branch nodes based on the AV movement, driving modes, and location of the crashes. There were total of 26 nodes in the DT developed using the CART model. The terminal node represents the probability of each crash type as a combination of factors that appear on the previous branch nodes leading to the terminal node. The decision rules can be expressed as a conditional statement of the nodes in different levels in the DT. A total of 14 rules (in terms of crash probability) were generated from the DT, and the top 10 significant rules are listed in Table 2. Each rule estimated the crash risk of different crash types for a combination of crash contributing factors.

Rule 1, 2, 3, 4, 7, and 8 (six rules out of top 10 rules) in Table 2 identified the higher probability of rear-end crashes. Four out of these six rules (Rule 1, 3, 4, and 7) had intersections as the location of crashes which means that rear-end AV crashes were more likely to happen at intersections. According to the pre-crash factors of DT Rule 1, when an AV stopped at an intersection and non-AV was slowing or proceeding straight behind AVs, and there was 80% chance of rear-end crash for non-autonomous mode of driving. DT Rule 2 indicates that the probability of rear-end crash was also high (51%) when the AV was proceeding straight on a roadway in a non-autonomous driving mode and non-AVs were slowing down or proceeding straight behind AVs. Pre-crash
factors of the DT rules 3 and 4 show that the rear-end AV crashes also occurred when the AVs were traveling at the ramps. DT rule 7 had the highest probability (87%) of rear-end crash when the AV was stopped in autonomous mode and the non-AV was making left or right turn movement at an intersection. Most of the rear-end crashes occurred due to the unsafe movement of human-driven non-AVs as they failed to stop before hitting the back of the AVs.

The second most common types of AV-related crashes were the sideswipe crashes. Sideswipe crashes were more likely to happen when both the AVs and non-AVs were proceeding forward or making turning movements in roadways or at intersections. DT rules 1 and 2 shows that when the location of the crash was on a roadway (i.e., not at an intersection) and the AV was proceeding forward (rule 2) instead of stopped at an intersection (DT rule 1) the chance of sideswipe crashes increased from 9% (DT rule 1) to 38% (DT rule 2). Chances of sideswipe crashes were 46% and 42% according to DT rules 9 and 10, respectively where crashes happened when the AVs and non-AVs were making turning movements at the intersections. Sideswipe crashes occurred during AV’s non-autonomous and autonomous driving mode. Rule 2 showed that sideswipe crashes occurred when both AVs and non-AVs were proceeding straight in a roadway. Most of the sideswipe crashes happened due to the fault of non-AVs. At an intersection, sideswipe crashes occurred when both AV and non-AV were making turning movements while the non-AV crossed its lane and contacted the AV. In other cases where sideswipe crashes occurred in a roadway, the non-AV sideswiped the AV while passing and merged back in the AV’s lane too closely.

On the other hand, DT rule 5 indicates 86% chance of a head-on crash when the AV was proceeding straight, and the non-AV was stopped or making left or right turning movements. According to the rule, head-on crashes were more likely to happen at locations such as middle or rightmost lanes, right-turning lanes rather than intersections. Rule 10 also showed that the chance of head-on crash was 42% when the location of the crash was an intersection, AV was in non-autonomous mode, and non-AV was performing turning movement or stopped. Analysis of the movements of AVs and non-AVs from these two rules (rules 5 and 10) showed that head-on crashes occurred mainly because of the non-AV. According to these rules, the non-AV traveled in the wrong direction in a one-way road after making a left/right turn or aggressively swerved lanes and stopped in front of an AV. Few of the head-on crashes also occurred because of the AVs while it was running in the non-autonomous mode (rule 10). Finally, rule 6 (in Table 2) describes the
combination of factors that created a scenario where the chance of broadside crash was 51.8%, and the chance of rear-end crash was 43.7%. This rule showed that the predominant movement of AV and non-AV at the moment of crashes was “proceeding straight,” and the location of the crash was “intersection.” Further investigating the crash reports, it was found that almost all broadside crashes occurred because of the non-AVs, where the non-AV either disobeyed the STOP sign at the intersection or violated the red light and proceeded through the intersection.

Overall, the rules indicated that AVs performed well in preventing sideswipe, broadside, and head-on crashes, leaving the rear-end crashes where AV in front of the non-AVs was the most critical crash scenario to be addressed. AVs are designed to follow the traffic rules and maintain posted speed limits, where human drivers in non-AVs choose a safe and comfortable speed often higher than the posted speed limit. Thus, the speed differential between AVs and non-AVs could lead to rear-end crashes. Intersections in a mixed traffic condition (i.e., a mix of AVs and non-AVs) presented a challenging environment for the AVs. Overcautious (e.g., longer startup delay by AVs at intersection stop bar) during the turning or forward movements at the intersections lead the non-AV drivers (with shorter startup delay) hit the AV and caused rear-end crash or passed the AVs and caused a sideswipe crashes.
Figure 6: DT of AV involved crashes, crash types as response variable (TM- Turning Movement, PS-Proceeding Straight)
Table 2: Extracted rules from the DT model

<table>
<thead>
<tr>
<th>Rule No.</th>
<th>Leaf Label</th>
<th>Crash Type Probability</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Non-AV Movement (Non-AV PS, Non-AV Slowing) &amp; Location (In Roadway) &amp;</td>
<td>Broadside</td>
<td>0.009</td>
<td>0.156</td>
<td>0.802</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>Driving Mode (Non-Autonomous) &amp; AV Movement (AV Stopped)</td>
<td>Head-on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Non-AV Movement (Non-AV PS, Non-AV Slowing) &amp; Location (In Roadway) &amp;</td>
<td>Rear-end</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Driving Mode (Non-Autonomous) &amp; AV Movement (AV PS)</td>
<td>Side Swipe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Non-AV Movement (Non-AV PS, Non-AV Slowing) &amp; Location (Ramp, Intersection) &amp;</td>
<td>Broadside</td>
<td>0.004</td>
<td>0.004</td>
<td>0.626</td>
<td>0.366</td>
</tr>
<tr>
<td></td>
<td>Driving Mode (Autonomous) &amp; AV Movement (AV PS)</td>
<td>Head-on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Non-AV Movement (Non-AV PS, Non-AV Slowing) &amp; Location (Ramp, Intersection) &amp;</td>
<td>Rear-end</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Driving Mode (Non-Autonomous) &amp; Disengagement (No) &amp; AV Movement (AV Slowing, AV Stopped)</td>
<td>Side Swipe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Non-AV Movement (Non-AV TM, Non-AV Stopped) &amp; Disengagement (No) &amp;</td>
<td>Broadside</td>
<td>0.013</td>
<td>0.864</td>
<td>0.090</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>Location (In Roadway) &amp; AV Movement (AV PS)</td>
<td>Head-on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Non-AV Movement (Non-AV PS, Non-AV Slowing) &amp; Location (Ramp, Intersection) &amp;</td>
<td>Rear-end</td>
<td>0.518</td>
<td>0.011</td>
<td>0.437</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>Driving Mode (Non-Autonomous) &amp; Disengagement (No) &amp; AV Movement (AV TM, PS)</td>
<td>Side Swipe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Non-AV Movement (Non-AV TM, Non-AV Stopped) &amp; Disengagement (No) &amp;</td>
<td>Broadside</td>
<td>0.007</td>
<td>0.103</td>
<td>0.870</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>Location (Intersection) &amp; Driving Mode (Autonomous) &amp; AV Movement (AV PS, AV Stopped)</td>
<td>Head-on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Non-AV Movement (Non-AV TM, Non-AV Stopped) &amp; Disengagement (No) &amp;</td>
<td>Rear-end</td>
<td>0.154</td>
<td>0.312</td>
<td>0.506</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>Location (In Roadway) &amp; AV Movement (AV TM, AV Stopped, AV Slowing)</td>
<td>Side Swipe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Non-AV Movement (Non-AV TM, Non-AV Stopped) &amp; Disengagement (No) &amp;</td>
<td>Broadside</td>
<td>0.153</td>
<td>0.019</td>
<td>0.367</td>
<td>0.460</td>
</tr>
<tr>
<td></td>
<td>Location (Intersection) &amp; Driving Mode (Autonomous) &amp; AV Movement (AV TM)</td>
<td>Head-on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Non-AV Movement (Non-AV TM, Non-AV Stopped) &amp; Disengagement (No) &amp;</td>
<td>Rear-end</td>
<td>0.007</td>
<td>0.414</td>
<td>0.156</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td>Location (Intersection) &amp; Driving Mode (Non-Autonomous)</td>
<td>Side Swipe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: DT= Decision Tree; AV= Autonomous Vehicle; PS= Proceeding Straight; TM= Turning Movements.
5.2 Rules Generated from Association Rules

For the association rules generation, six crash attributes (locations of the AV crash, mode of driving, disengagement, AV movement during the crash, non-AV movement during the crash, and crash types) were used. As rear-end and sideswipe were the most frequent types of AV crash, and the other two types of crashes (i.e., broadside, head-on) were rare in the database, a low value of support and confidence was used for the generation of the rules (Besharati and Tavakoli Kashani, 2018). In this study, by trial and error, the minimum support value of 0.03 and confidence value of 0.4 were selected, and the total number of rules generated after the initial training of the algorithm was 6,486. The lift value of all generated rules was greater than one, which means that the antecedent and consequents were positively interdependent in all rules. Next, the rules with crash types as the antecedent were separated to find the contributing factors behind the different types of crashes. The extracted rules from the association rule method using crash types as an antecedent are presented in Table 3.

Table 3: Rules Generated from Association Rule Method based on Crash Type

<table>
<thead>
<tr>
<th>Rule No.</th>
<th>Antecedent</th>
<th>Consequent</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Autonomous Mode, Intersection, AV Stopped, Non-AV Slowing</td>
<td>Rear-end</td>
<td>0.050</td>
<td>1.000</td>
<td>1.631</td>
</tr>
<tr>
<td>2</td>
<td>Autonomous Mode, Intersection, AV Slowing, Non-AV Proceeding Straight</td>
<td>Rear-end</td>
<td>0.030</td>
<td>1.000</td>
<td>1.630</td>
</tr>
<tr>
<td>3</td>
<td>Autonomous Mode, Intersection, AV Stopped, Non-AV Turning Movement</td>
<td>Rear-end</td>
<td>0.035</td>
<td>1.000</td>
<td>1.630</td>
</tr>
<tr>
<td>4</td>
<td>Autonomous Mode, Intersection, AV Stopped, Non-AV Turning Movement, Not Disengaged</td>
<td>Rear-end</td>
<td>0.050</td>
<td>1.000</td>
<td>1.631</td>
</tr>
<tr>
<td>5</td>
<td>Autonomous Mode, In Roadway, AV Proceeding Straight</td>
<td>Sideswipe</td>
<td>0.030</td>
<td>0.462</td>
<td>2.296</td>
</tr>
<tr>
<td>6</td>
<td>Non-Autonomous Mode, Intersection, AV Turning Movement</td>
<td>Sideswipe</td>
<td>0.035</td>
<td>0.438</td>
<td>2.177</td>
</tr>
<tr>
<td>7</td>
<td>Autonomous Mode, AV Proceeding Straight, Non-AV Proceeding Straight, Not Disengaged</td>
<td>Sideswipe</td>
<td>0.050</td>
<td>0.435</td>
<td>2.163</td>
</tr>
<tr>
<td>8</td>
<td>Non-Autonomous Mode, Intersection, AV Proceeding Straight, Non-AV Turning Movement, Non-AV Slowing</td>
<td>Broadside</td>
<td>0.030</td>
<td>1.000</td>
<td>13.266</td>
</tr>
<tr>
<td>9</td>
<td>AV Proceeding Straight, Non-Autonomous Mode, In Roadway, Disengaged, Non-AV Slowing</td>
<td>Head-on</td>
<td>0.040</td>
<td>0.571</td>
<td>5.415</td>
</tr>
<tr>
<td>10</td>
<td>AV Proceeding Straight, Non-Autonomous Mode, In Roadway, Disengaged, Non-AV Stopped</td>
<td>Head-on</td>
<td>0.035</td>
<td>0.538</td>
<td>5.100</td>
</tr>
</tbody>
</table>

Note: AV=Autonomous Vehicle

Most of the rules generated by the apriori algorithm supported the rules from the DT model, and also provided additional insights from the AV crashes. The top four rules shown in Table 3,
representing a combination of crash contributing factors related to the rear-end crash, were similar to the findings of the DT model. These four rules related to rear-end crashes identified autonomous driving mode as one of the crash contributing factors, which means that AV in autonomous mode had the rear-end crash risk higher than the non-autonomous driving mode. This finding indicates that fully engaged AVs were at high risk of rear-end crashes in mixed traffic conditions, where AV-related rear-end crashes happened because of the following non-AV driver’s inability to understand the movement of leader AV (e.g., whether AV stopped or slowing down).

Association rules 5, 6, and 7 (Table 3) were related to AV sideswipe crashes. These three rules show similar findings revealed by the DT rules 2, 3, 9, and 10 (Table 2). However, the DT rules did not capture the influence of AV’s operating mode on head-on crashes. Association rules 9 and 10 showed that AV’s non-autonomous driving mode was associated with head-on crashes, and the AV was disengaged to non-autonomous mode right before the crash occurrence. This phenomenon could be attributed to potential issues related to AV technologies failure to take appropriate driving actions, and AV safety drivers recognized the risk and attempted to take over vehicle control. The DT model explained that the broadside crashes happened when the AVs were making turning movements, and the non-AVs were clearing the intersection. According to association rule 8 (Table 3), broadside crashes also occurred when the non-AVs were making turning movements, and the AVs were clearing the intersections in non-autonomous driving mode. High lift values in a rule indicate strong positive dependence between the antecedents. As shown in Table 3, association rule 8 has a high lift value of 13.27, indicating a strong positive association between the antecedent events. This high lift value means that the likelihood of a broadside crash was high when these antecedent events (i.e., AV proceeding straight, AV driving in non-autonomous mode, non-AV turning movement) were present at an intersection-related crash. Again, this scenario captured by rule 8 was not explained in the rules generated from the DT method. Overall, results from DT method and Association rules were complementary to each other as both the methods revealed new information that could be used to develop countermeasures to reduce AV-involved crashes in the future.
5.3 Implications of the research findings

The findings of this study have several practical implications and can be used to develop collaborative AV technology development efforts among multiple stakeholders such as AV developers, traffic safety officials, and law enforcement agencies. In the future, the roads will be shared by different levels of autonomous vehicle—from level 0 with no automation to level 5 with full automation. This mixed traffic poses challenges for both non-AVs and AVs. AV-related crashes mainly occurred when drivers of non-AVs failed to execute an appropriate driving task. The AV crash database showed that there were about 122 rear-end crashes occurred in CA between January 2016 to February 2020, which is about 62% of total AV-involved crashes. About 63% of the 122 AV-involved rear-end crashes happened when the AVs were stopped or slowing down to a full stop at intersections. This statistic indicates that the AV-involved rear-end crashes occurred mostly due to the following non-AVs failure to react appropriately to avoid rear-end crashes with leading AV. As the AV technology is designed to obey traffic laws, the possible reason for the high number of rear-end crashes is that the non-AVs were unable to recognize the driving pattern of AVs. Non-AV manufacturers could develop countermeasures to prevent these rear-end crashes. To avoid this type of crash, AV technology developers can modify the AV’s behavior to operate more like human drivers to minimize the consequence of non-AV’s mistakes. In addition to obeying the traffic rules, AVs have to execute driving strategies that account for uncertainty about the movement of non-AVs. AVs can also make non-AVs aware of them by using distinct active signals about the status of autonomous driving mode (i.e., active or inactive). Also, AVs can provide warnings (via signals or signs) to non-AV drivers when they drive too close to the AVs. On the other hand, the non-AVs can install the Automatic Emergency Braking (AEB) system to avoid hitting the rear-ends of the AVs. Recently, the majority of the automakers planned to equip all new passenger cars and all new trucks with a low-speed AEB system by September 1, 2022, and September 1, 2025, respectively (Boggs et al., 2020). As most AV-related crashes mainly occurred at or near the intersections, the transportation agencies can modify the intersections to accommodate the safe movement of AVs (e.g., separate lanes for AVs depending on availability of right-of-way).
CHAPTER 6: CONCLUSION

The future of significant improvement in traffic safety largely depends on the successful deployment of AV technologies. Although AVs have recorded a significant amount of vehicle miles in testing, AV technology is still maturing, and various companies have been fine-tuning their AV technology through extensive research programs, including testing on public roads. The AV crash reports provided insightful information on the unique nature of AV-related crashes. In this research, Association Rule and DT methods were used to identify major patterns in AV crash contributing factors.

The most significant rules identified the challenges of AVs sharing the roads with non-AVs. Almost 62% of the total crash were rear-end crashes, and predominantly occurred at intersections. Results of the generated rules suggested that the rear-end crash was more prominent when the AV was engaged in autonomous mode. Most of the rear-end crashes occurred when the AV was stopped at the intersection, and the non-AV was proceeding straight or slowing down behind AV at intersections. For other crash types, both autonomous mode and non-autonomous modes of AVs were frequent. Sideswipe crashes occurred both at intersections and on a roadway when the non-AV attempted to pass the AVs or making turning movements. Further analysis of the AV’s and non-AV’s movement at the moment of crashes showed that all types of AV involved crashes occurred mainly because of non-AVs. Rear-end crashes occurred because of the non-AV’s failure to react appropriately to avoid rear-end crashes with leading stopped AV at the intersection. On the other hand, broadside, sideswipe, and head-on crashes occurred because of the non-AV’s unsafe movements, red-light running, or disobeying the STOP sign at the intersection.

The findings indicate that rear-end crashes can be avoided by taking few simple precautionary measures such as separating the AVs from non-AVs using AV-only lanes, installing low-speed AEBs in non-AVs, or using signals/signs in the AVs letting non-AVs know about the AV driving modes. Currently, CA DMV collects limited information about AV-involved crashes. Although CA DMV modified the crash report (OL 316 form) over the years, crash severity levels of AV-involved crashes were not collected using the KABCO (K-Fatal, A-Serious Injury, B-Moderate Injury, C-Minor Injury, and O-No Injury) scale. A more comprehensive data collection of AV involved crashes, including crash injury severity, driver’s cognitive state, roadway traffic and environmental conditions, degree of engagement of the autonomous system, and vehicle
acceleration and deceleration rates, will enable a robust understanding of the factors influencing AV crashes, which can accelerate the development of effective safety remedies. The availability of a larger AV crash database with crash severity data scale in the future can be used to develop crash frequency-based and crash severity models as an extension to this research. Results from the crash severity and crash frequency models can be used to improve roadway design and as well as vehicle design to increase AV-related traffic safety. Furthermore, AV technology is evolving rapidly with more testing on public roads, and the engagement of human drivers in non-AVs with AVs is expected to change over time. Because of the continuous improvement of AV technology and more widespread testing in public roads in the near future, a more comprehensive AV crash database could provide new perspectives about the AV crash mechanism that is partially or completely different from what was discovered using the AV crash database in this study.
DECLARATION

This thesis work is based on the following journal article published in Transportation Research Record.

REFERENCES


