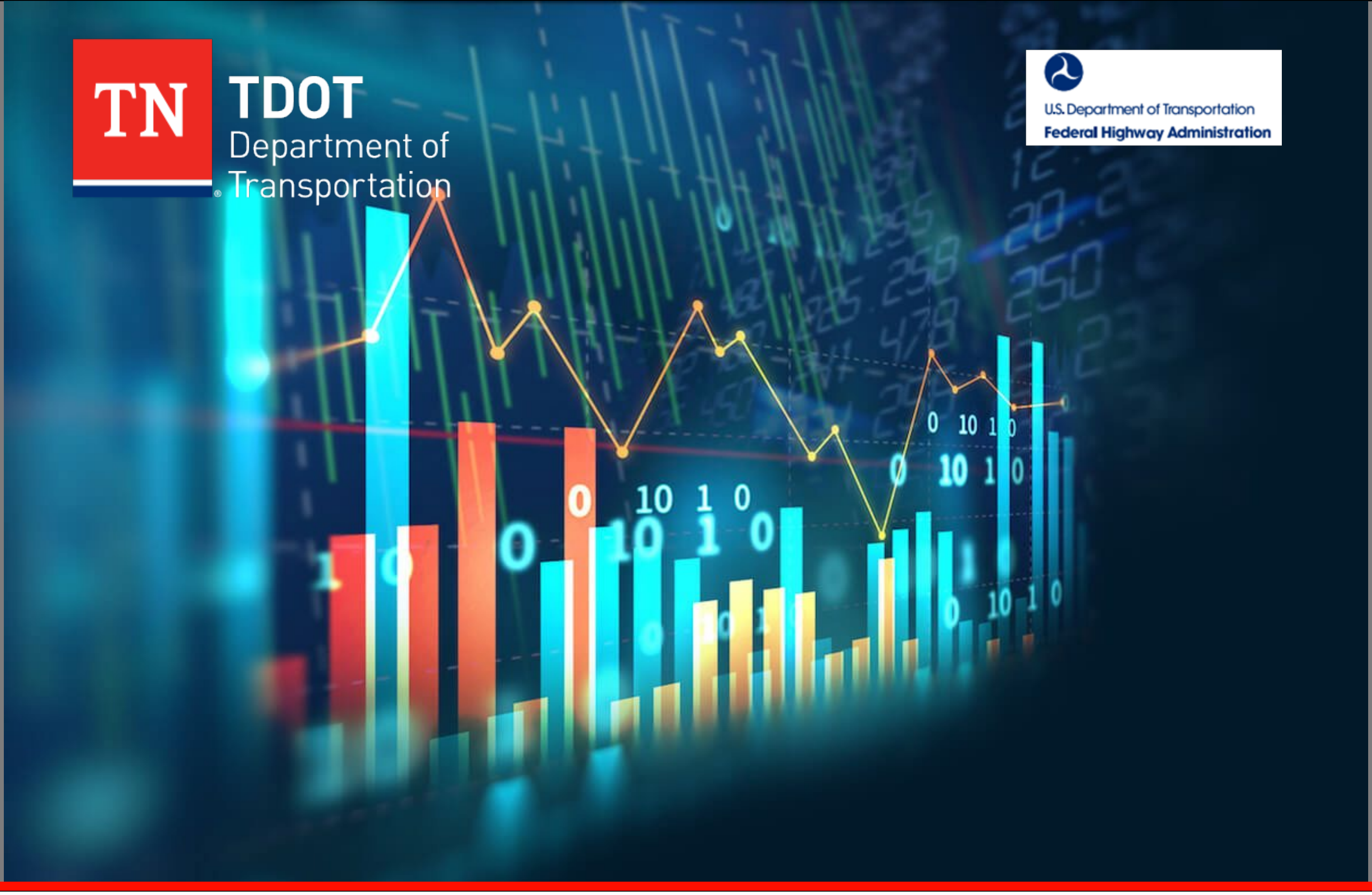


Connected and Automated Vehicles Investment and Smart Infrastructure in Tennessee

Part 4: Eco-system for Connected and Automated Vehicles: Investments in Data Collection, Analytics, and Simulations

Research Final Report from University of Tennessee | Asad Khattak, Subhadeep Chakraborty, Iman Mahdinia, Austin Harris, & Mina Sartipi | May 31, 2022

Sponsored by Tennessee Department of Transportation Long Range Planning Research Office & Federal Highway Administration

THE UNIVERSITY OF TENNESSEE KNOXVILLE

DISCLAIMER

This research was funded through the State Planning and Research (SPR) Program by the Tennessee Department of Transportation and the Federal Highway Administration under ***RES2019-07: Research on Connected and Automated Vehicles Investment and Smart Infrastructure in Tennessee.***

This document is disseminated under the sponsorship of the Tennessee Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Tennessee and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the views of the author(s) who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Tennessee Department of Transportation or the United States Department of Transportation.

Technical Report Documentation Page

<table>
<tr><td>1. Report No.
RES2019-07</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle

Eco-system for Connected and Automated Vehicles: Investments in data collection, analytics, and simulations</td><td colspan="2">5. Report Date
May 2022</td></tr>
<tr><td colspan="2">6. Performing Organization Code
R011313595</td></tr>
<tr><td colspan="2">7. Author(s)
Asad Khattak, Subhadeep Chakraborty, Iman Mahdinia, Austin Harris, & Mina Sartipi</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
University of Tennessee, Knoxville, Tennessee, 37996
University of Tennessee, Chattanooga, Tennessee, 37403</td><td colspan="2">10. Work Unit No. (TRAIS)</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
RES2019-07</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Tennessee Department of Transportation
505 Deaderick Street, Suite 900
Nashville, TN 37243</td><td colspan="2">13. Type of Report and Period Covered
Final Report.
May 2019-May 2022</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.
The project has developed a series of five (5) reports that support intelligent mobility strategies in Tennessee. This is report 4 of 5.</td></tr>
<tr><td colspan="4">16. Abstract
As Connected and Automated Vehicles (CAVs) diffuse through the transportation system, research, development, and deployment of such technologies are critical to unlocking their potential. It is critically important for any state Department of Transportation to lay the foundation for smart infrastructure. Such infrastructure can include supporting 1) roadside and onboard devices for connected vehicles and new sensors, 2) selection of context-relevant applications and user services, 3) installing roadside cameras and dynamic message signs, 4) deploying fiber for fast communication of data, and 5) installing traffic control device improvements. This report aims to provide the Tennessee Department of Transportation with information to support investment decisions in creating a complete ecosystem for connected and automated vehicles. Such a system will require investments in physical infrastructure and CAV data collection, analytics, modeling, and simulation. This report discusses the process of CAV data collection, which can be supplemented with smart infrastructure non-CAV data (e.g., cameras). Archiving, sharing, and visualization examples are discussed. Next, the data analytics and modeling for infrastructure technologies provide case studies to demonstrate the application of powerful statistical and Artificial Intelligence techniques. Finally, results from a diverse set of simulations are presented, e.g., SUMO and CARLA, which can be used for scenario analysis of mixed CAV and conventional transportation as well as analyzing Tennessee-specific edge cases. A discussion of recommendations related to CAV data collection, analysis, and simulation investments is provided.</td></tr>
<tr><td colspan="2">17. Key Words
Data Collection, Data Analysis, Simulation, Connected And Automated Vehicles, Smart Infrastructure</td><td colspan="2">18. Distribution Statement

No restriction. This document is available to the public from the sponsoring agency at the website http://www.tn.gov/.</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
62</td><td>22. Price</td></tr>
</table>

Acknowledgment

We are very grateful to the Tennessee Department of Transportation staff for their input and particularly to the following individuals for their support throughout the project: Mr. Lee Smith, Mr. Jon Storey, Mr. Brad Freeze, and Ms. Melanie Murphy.

Several students have contributed to this report. They include Ms. Nastaran Moradloo, Mr. Mohammad SafariTaherkhani, Mr. Latif Patwary, Ms. Antora Haque, and Mr. Zachariah Nelson.

Executive Summary

Background

> *The goals of TDOT's smart infrastructure project are to provide:*
> - *A complete picture of relevant research, development, and deployment (RDD)*
> - *Discuss key research findings and investment opportunities*
> - *Provide recommendations for investments in intelligent mobility*

Research, development, and deployment of Intelligent mobility technologies to improve transportation system performance are essential. It is critically important for any state Department of Transportation (DOT) to lay the foundation for a smart technology infrastructure ecosystem. Such infrastructure includes supporting 1) roadside and onboard devices for connected vehicles and new sensors, 2) selection of context-relevant applications/user services, 3) installing roadside cameras and dynamic message signs, 4) deploying fiber for fast communication of data, and 5) installing traffic control device improvements. Establishing the appropriate cyber-physical ecosystem is critical, which also entails the collection, processing, management/storage, and harnessing of connected and automated vehicle (CAV) communications data. For the operation of connected vehicles, such data are continuously being transferred (streamed) between roadside units and onboard units. The research team has worked on supporting the future efforts of the Tennessee Department of Transportation (TDOT) in terms of readiness for data collection, data analysis, and the use of simulation for emerging CAV technologies. Focusing on investments in smart infrastructure and intelligent mobility, actions and activities needed for supporting the CAV data collection, data analysis, modeling, and simulation efforts are provided. These are meant to assist in deploying the entire cyber-physical ecosystem for CAV technologies and smart infrastructure.

Key Findings

Focusing on smart infrastructure, the findings of investments in a CAV ecosystem are summarized in three areas:

Collection of CAV data. The whole CAV system is based on the fast movement of data over wireless networks, and hence a critical component of operating CAV systems is data collection. Data transfer in real-time enables 1) the applications and user services that improve traffic operations, 2) archived data helps improve planning and related models for the future, and 3) assists with an independent evaluation of emerging technologies. CAV data refers to the continuous streaming of Basic Safety Messages (BSMs), Traveler Information Messages (TIMs), Signal Phase and Timing (SPaT) messages, and logs of alerts or warnings, most of which are transmitted over wireless networks. For example, if alerts or warnings are given, then event logs can be created from BSM, TIM, and SPaT messages in a vehicle before and after the alert or warning was issued to the driver. Such data can be stored on Aftermarket Safety Devices (ASD) at the time of collection and pushed Over-The-Air from the ASD to the roadside unit (RSU), from where it can be archived on a secure server. Notably, CAV data can be collected, archived, and harnessed in different ways. Details are provided about CAV and non-CAV data sources, data archival, processing, and sharing, with specific use case examples from Tennessee (MLK smart corridor and Shallowford road in Chattanooga) and around the country covering the implications for smart infrastructure technology deployments in the future.

Data analytics and modeling are needed to use the CAV data effectively. This can include visualizing the collected data to measure system performance in real-time and tactical/strategic planning. CAV data are increasingly being shared through dashboards, data hubs, and data lakes. The analytics include visualization of CAV data. Specifically, CAV user services such as red-light running alerts or curve-speed warnings use standardized BSMs, which are data packets related to a vehicle's position, heading, speed, acceleration, state of control, and predicted path. These data can be transmitted from one vehicle to another via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, collectively known as vehicle-to-everything (V2X) communications. In a real-life application, they are analyzed by the receiving onboard unit (OBU) to determine the presence of hazardous situations and alert the driver of the host vehicle accordingly. Storing and analyzing these messages can provide insights into whether the alerts were given appropriately and if they were effective in avoiding hazardous situations.

Similarly, TIM provides drivers with information about traffic incidents, major events, and even evacuations. These messages typically utilize V2I communications and are sent to vehicles by RSUs. Furthermore, SPaT messages contain data about the state of signal phases at an intersection and related information. SPaT messages are processed by vehicles to support driver/vehicle decision-making at an intersection, e.g., whether to stop or go at a signalized intersection. The point is that these data are analyzed to improve the transportation system's performance, e.g., in terms of safety and mobility, as well as these messages can be analyzed for their effectiveness and harnessed more generally to improve system performance. Modeling the data and applications of Artificial Intelligence have gained momentum in this realm.

- Case studies highlight the experience with V2I technologies in the Chattanooga MLK smart corridor, analysis of BSM and alert data from bus drivers with access to “Enhanced Pedestrian Collision Warning Systems,” analysis of data on cooperative merging systems at on-ramps, and application of Artificial Intelligence techniques for smart traffic signal control strategies at intersections. New performance measures based on BSM data for safety (e.g., driving volatility and time to collision), energy, and emissions have also emerged.
- Case studies also feature experiences with specific CAV applications such as adaptive cruise control that utilized V2V technologies.
- Case studies highlight how CAV data can be more generally harnessed for proactive planning without a specific CAV application or user service.

The application of a key set of tools for CAVs is simulations. Several simulation tools are available for envisioning CAV scenarios, sensitivity testing, and identification of edge cases. Simulations can range from 1) using tools such as Simulations of Urban MObility (SUMO) and Car Learning to Act (CARLA) for insights about CAV performance at the levels of transportation network or vehicle sensors (Light Detection and Ranging (LiDAR), radar, and cameras), 2) hardware-in-the-loop simulations, e.g., the Rototest driving simulator for a realistic representation of vehicle (drivetrain) components, 3) multi-user virtual reality simulators for understanding driver behavior at different levels of automation and connectivity, and 4) digital twins to represent a real-time digital counterpart of an operating transportation system. Simulations can provide a system or vehicle-level testing and analysis of vehicle sensors and components. The tools can be viewed as "virtual testbeds" for developing and testing emerging technologies. Moreover, the toolsets can be integrated (e.g., combining SUMO and CARLA). Generally, simulations are needed as part of the CAV ecosystem because they can envision future strategic planning scenarios, e.g., mixtures of conventional vehicles and CAVs, anticipate the operation of high-level automated vehicles' that are merging at on-ramps and intersections, as well as explore "edge-cases" where extreme situations can be anticipated and addressed proactively. Case studies of simulations are provided in this report, e.g., studies using SUMO to anticipate future safety and CARLA to identify edge cases, and the digital twin using a representation of the transportation system in Chattanooga, Tennessee. The highlighted work represents a collaboration between The University of Tennessee and Oak Ridge National Laboratory.

Key Recommendations

Associated with the selection of context-relevant connected vehicle user services is creating an effective ecosystem. A set of actions include the following:

- ***Invest in collecting CAV data.*** This entails developing a CAV data management system, given the large scale of such streaming data, and identifying the types of CAV data that can support core TDOT functions, including operations, maintenance, planning, and the required workforce for data collection and management. Data collection also comes with investments in cybersecurity, given the potential for adversarial attacks on the large-scale streaming data generated by CAVs. Notably, cybersecurity is a national challenge, and, in this regard, TDOT can follow the guidelines provided by the National Highway Traffic Safety Administration (NHTSA). Some of the best practices in cybersecurity in the automotive industry are gathered and discussed in the NHTSA cybersecurity best practices report [1]. TDOT should consider developing CAV data sharing procedures within the agency and a sharing policy with external partners, including other agencies, industry, research institutions, and the general public. Such policies can enhance traffic operations and freight supply chains and support smart city initiatives.
- ***Invest in CAV data analytics and modeling.*** Procedures are needed that fully utilize data from CAVs and other sources to successfully operate CAV user services and understand/improve transportation system performance. Data analytics, modeling, and artificial intelligence techniques are critical in designing highly efficient, safe, and sustainable transportation systems and providing smart mobility services to passengers and freight customers. TDOT should consider creating CAV data dashboards to monitor the performance of the transportation system and the deployed CAV technologies. Specifically, to manage data, TDOT can create and maintain a CAV data dashboard through centralized servers. Such dashboards can provide information that helps oversee operations and inventory and assists stakeholders in tracking resources and activities across the State. TDOT can emulate the connected data platform (CDP), similar to the Georgia DOT use case, to begin integrating diverse data sources. Specifically, CDP can overlay road inventory, WAZE data, CAV device information, highway patrol data, traffic, and crash data in a user-friendly interface.

- ***Innovative uses of CAV data.*** TDOT can use new data sources related to CAVs to support planning activities and assess modeling tools and the methodology they are applied to reflect future uncertainty about CAV adoption. This includes developing transportation models based on CAV data and other data (e.g., crowdsourcing) to accurately estimate and predict transportation system performance and develop proactive and multimodal transportation management plans. Another use of data is providing short-term traffic performance predictions and locating hazardous sites. The data can further be harnessed to improve traffic signal performance by incorporating new performance measures such as driving volatility of the CAV trajectories and using CAV data in high-uncertainty situations such as incidents and special events for lane recommendations and determining dynamic speed limits. TDOT's partner agencies can also use the data, such as Fire and Emergency Medical Services. Further, CAV data can fill data gaps for various functions provided by TDOT, e.g., by maintenance or environmental divisions. All the potential uses will require analysis of the CAV and related data, with some requiring research.
- ***Invest in simulations to create virtual testbeds and digital twins to enhance transportation system performance.*** More investments in "virtual testbeds" through simulation methodologies such as digital twins and the use of software SUMO and CARLA simulations can be valuable for CAV data integration and processing, anticipating future scenarios, doing sensitivity analysis, as well as identifying Tennessee-specific "edge" (fringe) cases. Additionally, simulations can evaluate operational and planning strategies across large-scale networks. Notably, TDOT can further leverage modeling and simulation capabilities available in Tennessee through the universities and Oak Ridge National Laboratory. This can involve leveraging high-performance computing, data science, and advanced sensors and communications protocols to develop, test and deploy emerging technologies and algorithms for vehicle-to-everything communications (including, of course, the infrastructure and the grid) that enable applications for smart routing, smooth and safe traffic flow, and higher operational efficiency of the network. TDOT investments in applied research should be considered, e.g., using big data and machine learning to improve traffic signals' delay and safety performance in Tennessee or harnessing basic safety message data from CAV.
- ***Future research on data collection, processing, analysis, and dissemination.*** In terms of future CAV research, it is crucial to invest in:
    - Developing sophisticated visualizations of CAV data. Specifically, TDOT can invest in creating a data visualization platform that will process real-time data and show different performance metrics. For instance, the visualization may include throughput, arrivals on green, progression ratio, and travel time index on signalized arterials.
    - Using modeling, artificial intelligence, and simulation capabilities based on data generated by CAVs and smart infrastructure enablers to enhance the diffusion of higher automation levels.
    - Accurately estimate and predict transportation system performance and develop proactive and multimodal transportation management systems.

Table of Contents

DISCLAIMER i

Technical Report Documentation Page ii

Acknowledgment iii

Executive Summary iv

List of Tables x

List of Figures xi

Glossary of Key Terms and Acronyms xiii

Chapter 1 Introduction 1

1.1 Background 1

1.2 Purpose of the Report 2

1.3 Organization of the Report 2

Chapter 2 Connected and Automated Vehicle Data Collection (Task 4) 3

2.1 Introduction 3

2.2 Data Sources 3

2.2.1 Connected Vehicles Data 3

2.2.2 Non-Connected Vehicle Data 5

2.3 Data archival, processing, and sharing 5

2.4 Data collection issues and examples from Chattanooga, Tennessee 6

2.5 Data Visualization and Dashboard 8

2.5.1 Data Visualization in Chattanooga, Tennessee 8

2.5.2 THEA CV Pilot Dashboard 12

2.6 Synthesizing findings for Tennessee: Lesson Learned 14

Chapter 3 Data Analytics for Infrastructure Technologies (Task 5) 16

3.1 Introduction 16

3.2 Study Design and Modeling Methods 17

3.3 Vehicle-to-Infrastructure Applications 18

Study 1: MLK Smart Corridor 18

Study 2: Synthesis of V2I applications in THEA Pilot Project 20

Study 3: Enhanced Pedestrian Collision Warning System (EPCW) for bus drivers 24

Study 4: Cooperative Merging at Ramps 27

3.4 Vehicle-to-Vehicle Applications 28

Study 1: Safety-critical Applications: Safe Pass Advisory 28

Study 2: Platooning-Cooperative Adaptive Cruise Control 29
Study 3: Pedestrian Crash Prevention Systems 31
3.5 Harnessing CAV and non-CAV Data 32
Study 1: Driving volatility helps identify hazardous intersections 32
Study 2: Easing in Automated Vehicles-Experimentation in mixed traffic 34
Study 3: Predicting future crashes more accurately with CAV Data 35
Study 4: Identifying hazards through Automated Vehicle disengagements 36
3.6 Smart infrastructure: Traffic Signal Control Strategies Using AI 38
3.7 Synthesizing findings for Tennessee: Lesson Learned 41
Chapter 4 Simulations for Connected and Automated Vehicle Technology (Task 6) 42
4.1 Introduction 42
4.2 SUMO simulation-How future CAV scenarios are represented 43
4.3 Case Studies applying simulation 44
Study 1: The future of mixed traffic at intersections-SUMO simulations 44
Study 2: Identifying edge cases-CARLA Simulations 46
Study 3: Cooperative on-ramp merging simulations of high-level automated vehicles 49
4.4 Digital twin-application in Chattanooga 50
4.5 Research on techniques in CAV simulations and AI applications 54
4.6 Synthesizing findings for Tennessee: Lesson Learned 59
Chapter 5 Conclusions and Recommendations 60
References 64

List of Tables

TABLE 1: LIST OF DEPLOYED CAV TECHNOLOGIES FOR TDOT CONSIDERATION 18
TABLE 2: SUMMARY OF THE KITTI AND VKITTI2 DATASETS USED 55

List of Figures

Figure 2.1 Data sets available at ITS Data Hub (https://www.its.dot.gov/data/). 6
Figure 2.2 Region level highway traffic visualization 9
Figure 2.3 [Left] Radar detection sensor and [Right] a GridSmart camera sensor 9
Figure 2.4 Incident visualization (a) Heat map and point data, (b) Sunburst plot 10
Figure 2.5 Energy cost over a day in 5-minute increments 10
Figure 2.6 Daily traffic intensity at the I-75S junction with Shallowford road (a) whole week, (b) Tuesday to Thursday 11
Figure 2.7 Chord diagram to represent directional traffic flow at intersections 12
Figure 2.8 THEA CV Pilot Dashboard Components 13
Figure 2.9 Measurement Dashboard Page 13
Figure 2.10 Performance Dashboard Page 14
Figure 2.11 Warning Event Profile Option 14
Figure 3.1 MLK Smart Corridor Data Dashboard and Visualization 20
Figure 3.2 End of Ramp Deceleration Warning application 21
Figure 3.3 Pedestrian Collision Warning application 22
Figure 3.4 Transit Signal Priority overview 23
Figure 3.5 Transit Signal Priority functional flows 23
Figure 3.6 Traffic Progression Physical architecture 24
Figure 3.7 Onboard Subsystem (top) and Infrastructure Subsystem (bottom) (28) 26
Figure 3.8 Definition and visualization of the cooperative merging scenario 28
Figure 3.9 Definition and visualization of the platooning scenario 30
Figure 3.10 Pedestrian Crash Prevention System Evaluation Framework 32
Figure 3.11 Create Map from BSM Data (left), Data Preparation Framework (right) 33
Figure 3.12 Framework of the Pedestrian Crash Prevention Evaluation Study 34
Figure 3.13 Methodological Approach of the Study 36
Figure 3.14 Categorization of Disengagements Causes 37
Figure 3.15 Case Study of a grid traffic network containing 20 traffic signal controllers in the downtown district of Springfield, Illinois 40
Figure 3.16 Time versus Emissions in the Studied Scenarios 41
Figure 4.1 Intersection of Washtenaw Ave. and Huron Parkway in Ann Arbor, MI. The study area is shown by yellow boundaries 45
Figure 4.2 Number of conflicts observed at simulated intersection and confidence intervals in the study area, in 10% increments of AV market penetration (Left). Speed volatility at the simulated intersection for ACC and CACC vehicles (Right). 46
Figure 4.3 Framework to Incorporate AV Data into Crash Investigation Analysis 48
Figure 4.4 The proposed Data Pipeline 48
Figure 4.5 Visualization of CARLA sensors in CARLA Simulation #1 49
Figure 4.6 Visualization of CARLA Simulation #2 49
Figure 4.7 GridSmart camera field-of-view 51
Figure 4.8 A schematic of queue estimation using probabilistic modeling 51
Figure 4.9 A relevant section of the Shallowford road corridor divided into observation zones . 52
Figure 4.10 SUMO representation of the corresponding Shallowford section with all signal phases and actuated traffic control .... **Error! Bookmark not defined.**

Figure 4.11 Ground-truth queue length and estimated queue lengths show a high level of accuracy in queue prediction using the probabilistic estimation method 54
Figure 4.12 An example of the YOLOv5 architecture applied to an image from the KITTI dataset 55
Figure 4.13 The testing results of two Yolov5 object detectors 56
Figure 4.14 A LiDAR point cloud taken from a Velodyne LiDAR is shown in blue, containing 360 degrees of information around the vehicle. A Pseudo-LiDAR point cloud is shown in green, a point cloud where every pixel from a depth estimation is converted to 3D space. 57
Figure 4.15 cGAN based Pseudo-LIDAR 58
Figure 4-16 Image-to-image translation methods are shown using the CycleGAN framework ...59

Glossary of Key Terms and Acronyms

ACC - Adaptive Cruise Control
API - Application Programming Interfaces
ASD - Aftermarket Safety Devices
BSM – Basic Safety Message
CACC - Cooperative Adaptive Cruise Control
CARLA - Car Learning to Act
CAV – Connected and Automated Vehicle
C-V2X - Cellular Vehicle-to-Everything
DSRC - Dedicated Short-Range Communication
EPB - Electric Power Board
EPCW - Enhanced Pedestrian Collision Warning
FHWA - Federal Highway Administration
I-SIG - Intelligent Traffic Signal System
OBU – Onboard Unit
PCW - Pedestrian Collision Warning
PSM - Pedestrian Safety Message
RSU – Roadside Unit
SDR - Software Defined Radio
SPaT - Signal Phasing and Timing
SRM - Signal Request Message
SSM - Safety Surrogate Measure
SUMO - Simulation of Urban MObility
TIM - Traveler Information Message
TSP - Transit Signal Priority
UTC - University of Tennessee at Chattanooga
US DOT - U.S. Department of Transportation
V2I - Vehicle-to-Infrastructure
V2V – Vehicle-to-Vehicle
V2X – Vehicle-to-Everything
WAVE - Wireless Access in Vehicular Environments
WSPM - WAVE Short Messaging Protocol
WWE- Wrong-Way Entry

Chapter 1 Introduction

1.1 Background

The purpose of this chapter is to describe eco-system for connected and automated vehicles.

The cyber-physical CAV ecosystem must be able to harness the data.The eco-system can be enhanced by developing simulations.

To create ecosystems for connected and automated vehicles (CAVs), state Departments of Transportation (DOTs) are investing in smart infrastructure and associated data collection. Additionally, they are investing in data analysis, application of Artificial Intelligence techniques (for smart control strategies), and the development of simulations, e.g., digital twins that can monitor system performance and accelerate CAV development and deployment. To collect and store connected vehicle data, large servers are devoted to archiving CAV and related traffic/inventory data. Integrating data from different sources can be challenging but very useful in analyzing and measuring the performance of CAV technologies and determining their appropriate uses and value. Given that CAV data comes as continuous streams, this microscopic level CAV data (e.g., in the form of basic safety messages or warnings and alerts given to drivers) provides opportunities to operate the system more efficiently and explore relationships that were previously too difficult to infer.

The cyber-physical CAV ecosystem must be able to harness the data. Analysis of CAV infrastructure technologies includes the smooth operation of specific vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) applications and services, e.g., adaptive cruise control (ACC), curve speed warning, red-light running warnings, and assessments of their impacts on safety and mobility. CAV data can also be harnessed to improve the transportation system's performance, e.g., by identifying hazardous locations where high levels of speed volatility are observed. These sites can be potentially unsafe where improvements can be targeted. Integrated CAV databases can enable high-quality assessments of how the impacts of CAV technologies could vary relative to, say, the intersection geometry and local demographics, such as local population, urban or suburban settings, density, and the mix of uses of surrounding intersections. The integration of inventory and user data in a rich Geographic Information Systems (GIS) environment can help identify high-risk situations for drivers and high-risk intersections. It will also provide opportunities to conduct specialized analyses such as corridor analysis and exploit new modeling techniques, including machine learning, cluster analysis, and time-series analysis. The idea is to turn CAV data into CAV information.

The eco-system can be enhanced by developing simulations, e.g., combining data analysis with scenario analysis of mixed high-level CAV and conventional transportation. To explain further, a simulation supporting the safety impact of mixed traffic can be insightful in terms of strategic planning. The simulation can anticipate future safety impacts of mixed traffic. A simulation framework that accounts for travelers who are not driving CAVs and those with partially automated CAVs that relieve drivers of car-following and steering tasks is provided. Different market penetration scenarios are simulated using Simulation of Urban MObility (SUMO) software for demonstration. Car Learning to Act (CARLA) software can be used to explore "edge-cases" unique to Tennessee, where extreme situations for CAVs can be anticipated and addressed proactively.

1.2 Purpose of the Report

This report aims to summarize the findings and research on CAV data collection, analysis, and the application of simulations. The report provides details of CAV data collection, visualization, analytics, and modeling and simulation through use cases that demonstrate the application of these emerging technologies. The report also summarizes the lessons learned and recommendations for investments in smart infrastructure and deployments in Tennessee.

1.3 Organization of the Report

The report is organized into the following sections:

Chapter 2 – Connected and Automated Vehicle Data Collection. This chapter discusses the processes that can be used to collect connected vehicle (CV) data. Supplementing this, non-CV data and archiving and sharing approaches are also discussed. Ideas on presenting the CV performance data online are also provided.

Chapter 3 – Data analytics for infrastructure Technologies. This chapter describes the powerful statistical and Artificial Intelligence methods that are being developed and applied to improve system performance by applying emerging technologies.

Chapter 4 – Simulation System Impact Evaluation Results. This chapter presents the results of simulations, especially the tools (e.g., SUMO and CARLA) that can be used for scenario analysis of mixed CAV and conventional transportation as well as analyzing Tennessee-specific edge cases.

Chapter 5 – Conclusions and recommendations. The findings are summarized along with the contributions of the reported work. A discussion of recommendations and lessons learned is provided.

Chapter 2 Connected and Automated Vehicle Data Collection (Task 4)

2.1 Introduction

Investment in CAV data collection on emerging technologies are critical for TDOT

Investments in CAV data collection on emerging technologies are critical for TDOT to create the ecosystem needed for smart infrastructure. The Tennessee-specific data collection related to connected vehicle technologies has occurred mainly in Chattanooga in the MLK smart corridor and Shallowford Road corridor. For the TDOT-sponsored I-24 Smart Corridor project, CAV data collection is planned along with identifying performance measures and relevant data to assess the impacts of intelligent mobility strategies being implemented in the corridor. Additionally, substantial CAV deployment activity is anticipated across the state, including Oak Ridge, Knoxville, Hendersonville, Memphis, Mount Juliet, Farragut, Lenoir City, Nashville, Goodlettsville, and LaVergne. The research team has collected and analyzed data on new and emerging technologies (e.g., Basic Safety Message data, alert and warning data) from other key sources, including transportation technology pilot tests around the country. These data provide national trends and specific examples of emerging data and insights into unmet data needs for the future.

The reason for investing in CAV data collection is for TDOT to take full advantage of the mobility and safety benefits, which can be maximized through the communication of CAV data. The connected and automated vehicle data can come from three major sources: (1) CV testbed generated data, 2) non-CAV data, (can include externally generated data, e.g., traffic data, and safety data, Bluetooth, weather data, and emissions data), and 3) behavioral surveys of testbed users. CV data in this report comes from CV devices transmitting and receiving information. Such data are typically stored on dedicated servers (samples of relevant data are available from US DOT Intelligent Transportation Systems (ITS) Data Hub and the Secure Data Commons). Notably, the collected data can have Personally Identifiable Information (PII), which must be removed before sharing the data publicly. This chapter discusses the data collection and documents the procedures for data processing, archival, and sharing. A database server typically allows archiving CAV data. This is often integrated with data from different (non-CAV) sources, which helps analyze and measure the performance of CAV technologies and determine their appropriate uses and value. Using the data, researchers will be able to check for relationships that were previously too difficult to infer.

2.2 Data Sources

2.2.1 Connected Vehicles Data

The CV-generated datasets are based on vehicle-to-vehicle communications between onboard units (OBUs) and vehicle-to-infrastructure enabled through OBU/roadside unit (RSU) interactions. Vehicles moving through the transportation system and having OBUs can generate data in the form of Basic Safety Messages (BSMs), which can be collected by RSUs and transferred to a server. The roadside units also broadcast relevant information to OBUs, e.g., traffic signal timing from a controller to equipped vehicles in the vicinity of the intersection.

RSU Data

Roadside units can transmit and collect the following data:

- BSM data from equipped vehicles, usually up to 10 Hz (or ten times a second). BSMs can also be collected by a vehicle operating in the range of a roadside unit.
- Signal Phase and Timing Message (SPaT) data from RSUs typically at 10 Hz
- Map Data Message (MAP) from RSUs, typically at 1 Hz
- Traveler Information Message (TIM) from RSUs, typically at 1 Hz
- Traffic Signal Request Message (SRM) transmitted by OBUs within range of an RSU (about 1,000 ft).
- Traffic Signal Status Message (SSM) broadcast by RSUs for conveying back to OBUs the status of its SRM.

Additionally, depending on the system deployed, the Multimodal Intelligent Traffic Signal System can take advantage of CV data for various vehicle types (including transit, emergency, freight) and pedestrians. Such information is often transferred through JavaScript Object Notation. Whether Pedestrian Crossing (PED-X) application is used, the information can include Pedestrian Safety Message (PSM) that triggers a collision alert. This report provides more detailed information about the use of BSM data in the context of passenger vehicles and public transit (enhanced pedestrian collision warning system).

OBU Data

The equipped vehicles can record received and transmitted data from interaction with proximate vehicles and roadside units. OBU data can include:

- Warnings issued to the driver of a vehicle.
- Internal system monitoring events (e.g., data from SD cards, security audits).
- Messages sent or received using WAVE Short Messaging Protocol (WSMP); WAVE is Wireless Access in Vehicular Environments, which is the mode of operation for 802.11 devices.

Note that driver warning event records (e.g., for transit vehicle drivers) are created whenever the application (a pedestrian) triggers a warning. The OBU creates a unique warning ID. Each record represents the warning and the time and location it was triggered. The record also contains the host vehicle's BSM at a given time. Warnings that result from receiving a PSM can be captured from the vulnerable road user triggering the pedestrian collision warning.

To reiterate, the whole connected and automated vehicle system is based on the fast movement of data over wireless networks. Hence a critical component of operating CAV systems is data collection. Data transfer in real-time enables 1) the applications and user services that improve traffic operations, 2) archived data helps improve planning for the future, and 3) assists with an independent evaluation of emerging technologies. CAV data refers to the continuous streaming of BSM, TIM, and SSPaT messages, and logs of alerts or warnings, most of which are transmitted over wireless networks. For example, if alerts or warnings are given, then event logs can be created from BSM, TIM, and SPaT messages in a vehicle before and after the alert or warning was issued to the driver. Such data can be stored on Aftermarket Safety Devices (ASD) at the time of collection and pushed Over-The-Air from the ASD to the RSU, from where it can be archived on a secure server. Notably, CAV data can be collected, archived, and harnessed differently. This report provides details about CAV and non-CAV data sources, data archival, processing, and sharing, with specific case study examples from Tennessee (MLK smart corridor and Shallowford Road in Chattanooga) and around the country, covering the implications for smart infrastructure technology deployments in the future.

2.2.2 Non-Connected Vehicle Data

Data can be collected from other sources for integration with CAV data. Some of these data and associated technologies are a part of smart infrastructure. The data can be harnessed to enhance the performance of the transportation system. The following datasets are often collected in CAV testbeds:

- Crash and traffic incident data.
- Road inventory data including segment and intersection, e.g., traffic signals data.
- Traffic flow data, e.g., from loop detectors and probes, including vehicle counts
- Weather event data.
- Crowdsourced traffic data such as WAZE and INRIX data.
- Field camera data (images and videos), e.g., fisheye Gridsmart cameras.

The raw data typically requires a dictionary that can assist with analyzing the data.

2.3 Data archival, processing, and sharing

A server is typically used to receive and archive all CV and non-CV data. Given the large size of CV data, they are typically stored as highly compressed files. CV data will contain BSMs, TIMs, MAPs, SPaT messages, SSMs, and OBU data logs, with PII removed. Information Technology specialists set the data governance procedures and manage access to the generated data.

Visualization can process the data and obtain insights into the transportation system's performance and CAV technologies, as demonstrated in this report at the Shallowford corridor in Chattanooga, Tennessee. Also, the status of the CAV fleet (OBU functioning), RSUs, BSMs, and warnings received by drivers can be observed, often in near-real-time, and stored for future analysis.

Sharing of the CAV data is done via portals. Publicly available archived CAV data is illustrative of what is collected in CV Pilots and shared openly. The US DOT is disseminating the archived data. The data are categorized into CV messages, trajectories, and connected equipment, as shown in Figure 2.1. For further analysis, these data are available for download via the ITS Data Hub at https://www.its.dot.gov/data/.

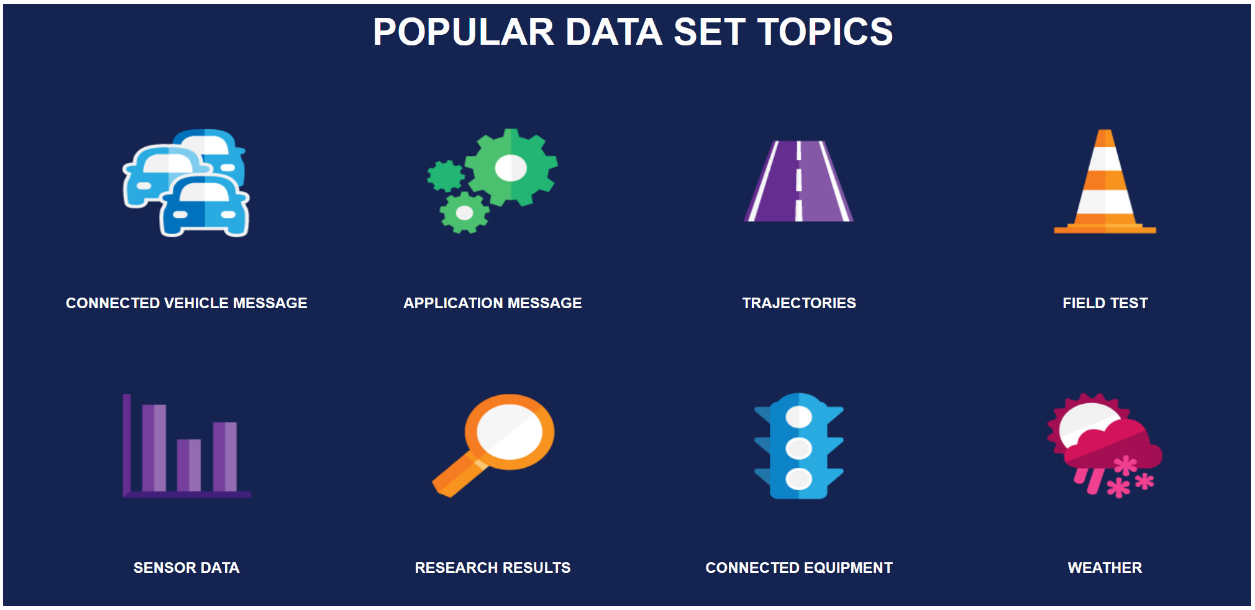


Figure 2.1 Data sets available at ITS Data Hub (https://www.its.dot.gov/data/).

2.4 Data collection issues and examples from Chattanooga, Tennessee

To remain competitive in the rapidly changing landscape of ITS, it is critical for cities and states to make investments in the right technology at the right time. However, predicting the future trajectory of an ITS application based on the current, remarkably fluid state-of-the-art is a challenge. For example, even just five years ago, Dedicated Short Range Communication (DSRC) was the technology of choice for V2V and V2I communication. However, it has been transitioning to the 5G Cellular Vehicle-to-Everything (C-V2X) cellular communication framework more recently.

Given the emergence of Artificial Intelligence (AI) applications in traffic control, the discussion will focus on the specific problem of investing in AI-based technology for improving traffic efficiency and safety at intersections and along state and local highways by collecting and analyzing traffic data such as vehicle volume, vehicle class, speed, turn direction, lane occupancy, queue lengths, and delay. There are many potential solutions offered in literature to this classic problem, ranging from drone-based surveillance methods to crowd-sourced data scraping techniques, as well as more deployment-ready, vision and AI-based commercial turnkey setups.

These techniques offer exciting new possibilities for collecting and analyzing traffic data. However, these new techniques need to be carefully evaluated based on a few constraints and requirements before adoption. Any ITS technology, to be a serious contender for adoption into a traffic operation workflow, should be

- Near-deployment ready, yet future proof and potentially expandable;
- Reliable with quantifiably high accuracy, yet not reliant on an expert operator;
- Scalable in terms of computation, communication bandwidth requirements, and cost.

The combined sensing and analysis package offered by Gridsmart and Miovision comes close to fulfilling these requirements with turnkey solutions that require very little operator knowledge.

These advantages need to be weighed against the significant initial investment cost, potentially high maintenance cost, and lack of flexibility in deployment and operation.

No matter which technology is used, an essential aspect for intelligent traffic systems to operate more effectively is to ensure that such systems receive relevant information. Regardless of the complexity of machine learning (ML) or statistical methods to govern behavior, any optimization system will not perform well if the information supplied is delayed or does not adequately reflect the environment. To avoid the issue of erroneous information being supplied to a system, which is sometimes referenced as "Garbage in, garbage out," is to provide sensors at locations that may allow for traffic technologies to perform at more optimal conditions.

While it can be assumed that CAV systems can relay their relative position to any intersection with the V2I communication, it should be assumed that in a realistic environment, not every vehicle along a roadway will have V2I communication. A mixed flow of CAV vehicles and conventional vehicles may hinder the performance of intelligent traffic system controllers, which can hinder the benefits of CAV platoons in urban environments and V2I that benefits vehicles.

Use of CV data

Some of the early testings of CAV technologies and associated CAV data collection in Tennessee occurred in Knoxville and Johnson City. This was done by installing DSRC units in Knoxville and Johnson City, connecting them with traffic signal controllers, and exchanging BSMs between roadside and onboard units. This was done in collaboration and with the help of city traffic engineers. The study explored an advisory system that provided drivers with traffic signal information and speed recommendations to reach the destination intersection and analyzed the computational challenges associated with such a system. The system used real-time SPaT information and MAP data retrieval for advisory calculation and DSRC for V2I communication. The advisory application helped drivers make informed decisions following audio and visual recommendations presented in a less distractive way. This testing was similar to the ISIG application, which is discussed more fully in the context of the Tampa Hillsborough Expressway Authority (THEA) CV Pilot project.

Use of Non-CV data

An important aspect of intelligent transportation systems is to ensure that such systems receive relevant and real-time information. Regardless of the detail or complexity of a machine learning or other statistical system to govern behavior, any system will not perform well if the information supplied is delayed or does not adequately reflect the environment. To avoid the issue of delayed information, sensors can be located more optimally.

Various devices have been developed to collect and record vehicle data, such as velocities or simple vehicle counts. A relevant data collection device is the application of cameras stationed at or near a traffic intersection. Traffic cameras benefit from using a single device to monitor multiple lanes and collect an extensive array of information through methods such as image processing. One example of existing technology is the Gridsmart fish-eye cameras, which could count the number of vehicles that have passed through an intersection, as well as approximate the speed and length of each vehicle that passed through. Data from the cameras are used to provide detailed information about the movement of vehicles and the associated traffic queues. Real-time data from the cameras and other sensors enables establishing a more intelligent and responsive machine learning strategy for improving signal timing. More details are explained in the following chapters [2].

2.5 Data Visualization and Dashboard

2.5.1 Data Visualization in Chattanooga, Tennessee

While it can be assumed that CAV systems can relay their relative position to any intersection with the vehicle to infrastructure (V2I) communication, it should be assumed that in a realistic environment, not every vehicle along a roadway will have V2I communication. A mixed flow of CAV vehicles and conventional vehicles may hinder the performance of intelligent traffic system controllers, which can hinder the benefits of CAV platoons in urban environments and V2I that benefits vehicles. In MLK Smart Corridor in Chattanooga, TN, various smart devices have been developed to collect and record vehicle data such as velocities or simple vehicle counts, given that no CAV is presented. A relevant data collection device is the application of cameras stationed at or near a traffic intersection. Camera systems such as GridSmart provide a wealth of real-time data that can be visualized and analyzed for gaining insights into traffic operation at the regional, corridor, and intersection levels. However, cameras can experience a handful of technical limitations as well, one of them due to adverse weather conditions. Heavy rain, snow, or fog, can obscure the camera's field of view or lead to lower efficiency in the computer vision's performance. Traffic cameras benefit from using a single device to monitor multiple lanes and collect an extensive array of information through image processing. One example of existing technology that was evaluated by the research team is the Gridsmart fish-eye cameras, which can count the number of vehicles that have passed through an intersection and approximate each vehicle's speed and length.

Region-level visualization

A camera-network-based monitoring system allows ingestion of real-time data that can be used to visualize incidents and identify network choking points and sources of congestion. With historical data, these real-time data can be used to predict short-term traffic patterns, which can inform routing, ramp-metering, and signal control adjustments. Figure **2.2** shows a tool for region-level visualization, where highway traffic is monitored by real-time data from 71 GridSmart cameras, providing time stamps, speed, length, approach direction, and turn direction for each vehicle detected. This data can be integrated from more than 200 radar detectors which provide lane-level information.

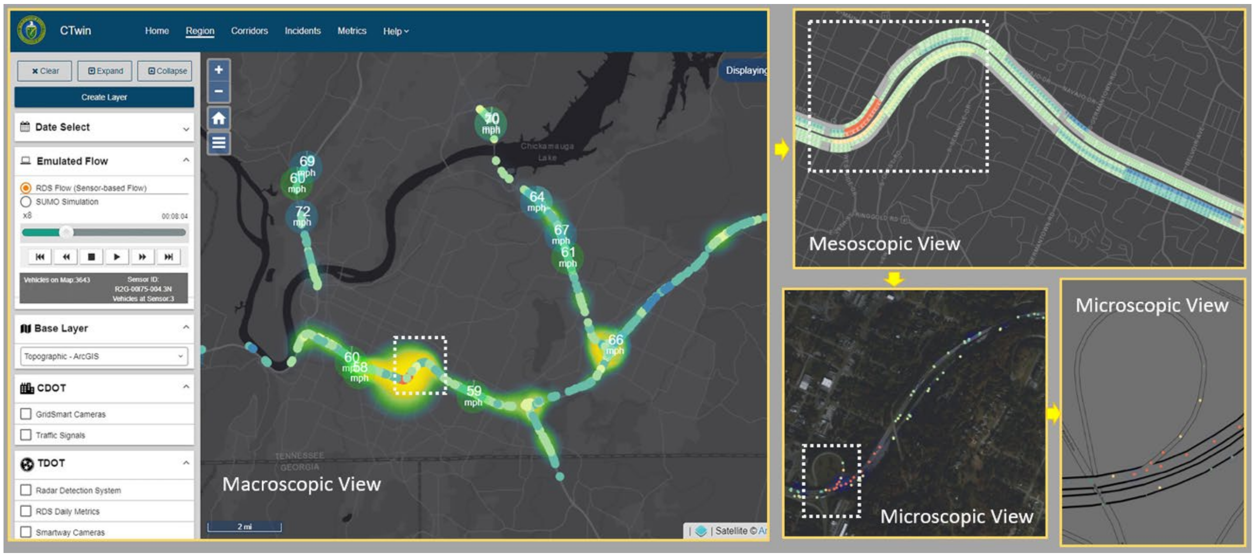
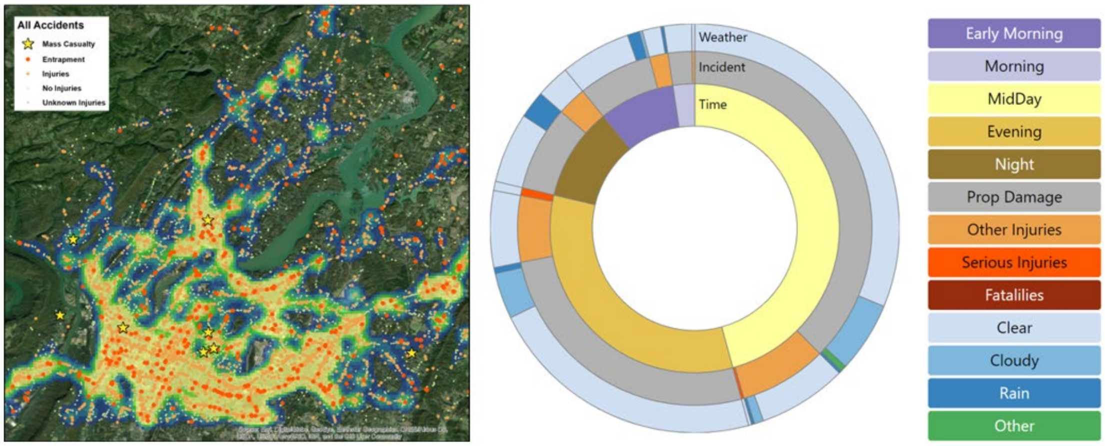


Figure 2.2 Region level highway traffic visualization

Figure 2.3 shows examples of radar and cameras deployed on the road. Occurrences of incidents can be analyzed temporally and spatially by synthesizing the region-level data. Figure 2.4 shows a heat-map and point data of all accidents over the monitored region, classified by varying degrees of severity (casualty to no injuries). This kind of visualization provides a snapshot of the frequencies. Thus the "hot spots" are related to the probability of an accident. Further analysis can correlate this spatial variance with localized driving metrics such as volatilities, leading to remedial measures such as re-examining road geometries and speed limits. Further causal analysis can be accomplished by studying a sunburst plot visualization of traffic incident data, as shown in Figure 2.4. This data can provide information such as what percentage of incidents occurred during mid-day versus late evening or how much weather played a role in the frequency of incidents.

Figure 2.3 [Left] Radar detection sensor and [Right] a GridSmart camera sensor



(a) (b)

Figure 2.4 Incident visualization (a) Heat map and point data, (b) Sunburst plot

Corridor-level visualization

At the corridor level, traffic flow metrics can be converted into estimates of fuel usage along target corridors, providing further insight into the energy cost of specific routing strategies and incidents. Combined with simulation, this setup provides a powerful tool for assessing the efficacy of policies and strategies that control traffic flow now and projected into the future. Figure 2.5 provides an example of energy cost calculated along the Shallowford Road corridor in Chattanooga on a specific day. The two peaks correspond to 8 am and 5 pm traffic which hints at considerable energy savings by improving the traffic flow efficiencies through adaptive signal phase and timing (SPaT) control.

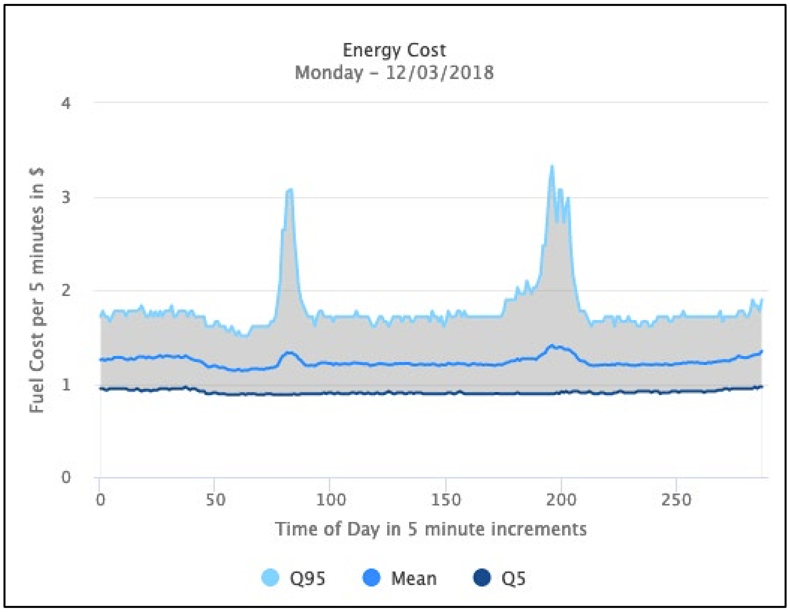
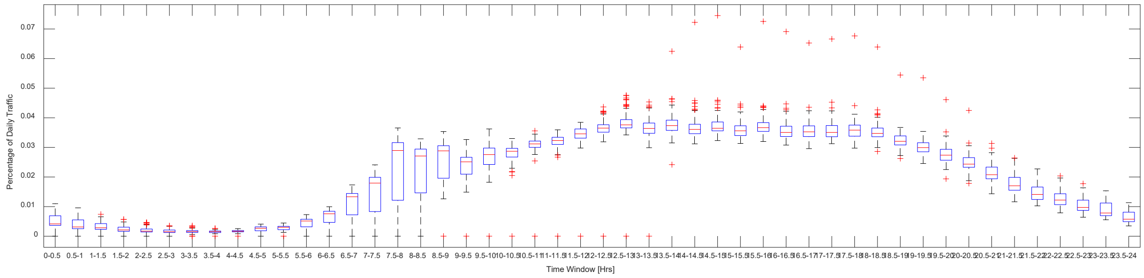
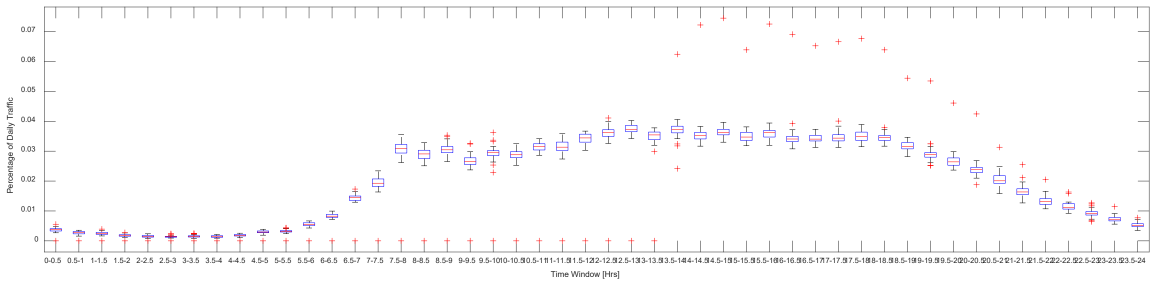


Figure 2.5 Energy cost over a day in 5-minute increments

Interesting trends can be observed by combining data from multiple sources and modalities. An example is shown in Figure **2.6**. The percentage of daily traffic is plotted for each 30-minute interval for the I-75S junction with Shallowford road. Figure **2.6** a and b are identical, except Figure **2.6**-b only plots data from Tuesday to Thursday, reducing the large variability seen in the full week's data (Figure **2.6**-a). This provides valuable information regarding the daily variation in traffic flow through the same region on weekdays versus weekends.



(a)



(b)

Figure 2.6 Daily traffic intensity at the I-75S junction with Shallowford road (a) whole week, (b) Tuesday to Thursday

Intersection level visualization

Intersection level data are crucial for developing historical trends and parameters, such as the split ratio between the volume of through traffic and turning vehicles. Notably, these data are also the main ingredient in developing future adaptive algorithms optimizing traffic flow. Raw data from the intersections provide information about each vehicle's time-stamp of detection, movement, and length as it enters the monitored zone. This cryptic data needs to be analyzed and visualized to produce easily accessible information in the form of chord diagrams. In these visualizations, the four cardinal directions are represented as arcs of a circle, and traffic flow between each pair of cardinal directions is represented by a band spanning between the pair of corresponding arcs. For example, in Figure **2.7**. the wide brown band informs us about the relatively large volume of traffic moving in the west to east direction, compared to the small volume of north-to-south traffic.

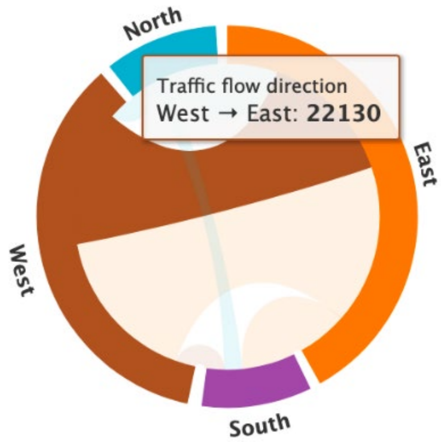


Figure 2.7 Chord diagram to represent directional traffic flow at intersections

2.5.2 THEA CV Pilot Dashboard

A dashboard is a visual display of information on an organization's operations that help managers, analysts, and other employees align their decisions with the organization's goals. Usually, dashboards are accessible to multiple decision-makers with different needs. The Performance Measurement and Evaluation Dashboard (PMED) was initially developed to meet US DOT's performance reporting needs for several measures. The components of the dashboard and their integration are shown in Figure 2.8. The dashboard is designed to completely support any form of device, including smartphones, tablets, personal computers, and so on. The data are provided to the dashboard by a secure backend service, making the design secure and reliable. The performance dashboard page (Figure 2.10) provides a snapshot of continued monitoring of the overall THEA CV Pilot progress using a near-real-time dynamic data feed from CV infrastructure toward increased mobility and safety goals. In fact, this page can provide the operational health status of the deployed CV equipment (operational indicators), the amount of data generated by vehicles (application activation indicators), and key mobility and safety performance measures related to the deployment of V2V and V2I applications (performance indicators). The measurement dashboard page (Figure 2.9) provides more detailed information about the CV fleet, the status of each RSU, BSM-based performance measures, and a comprehensive assessment of the V2V and V2I warnings deployed. This page can be viewed only by administrators, the THEA CV Pilot team, and US DOT analysts. Finally, the last feature of the dashboard is the visual animation of every single event that triggers a warning in the study area (Figure 2.11). When a user selects a particular warning on the map, more options appear to display more detailed information.

Based on the THEA CV Pilotproject insights, TDOT can consider such a dashboard to provide near real-time measures internally, manage the system more precisely, and provide the public with access to CAV information.

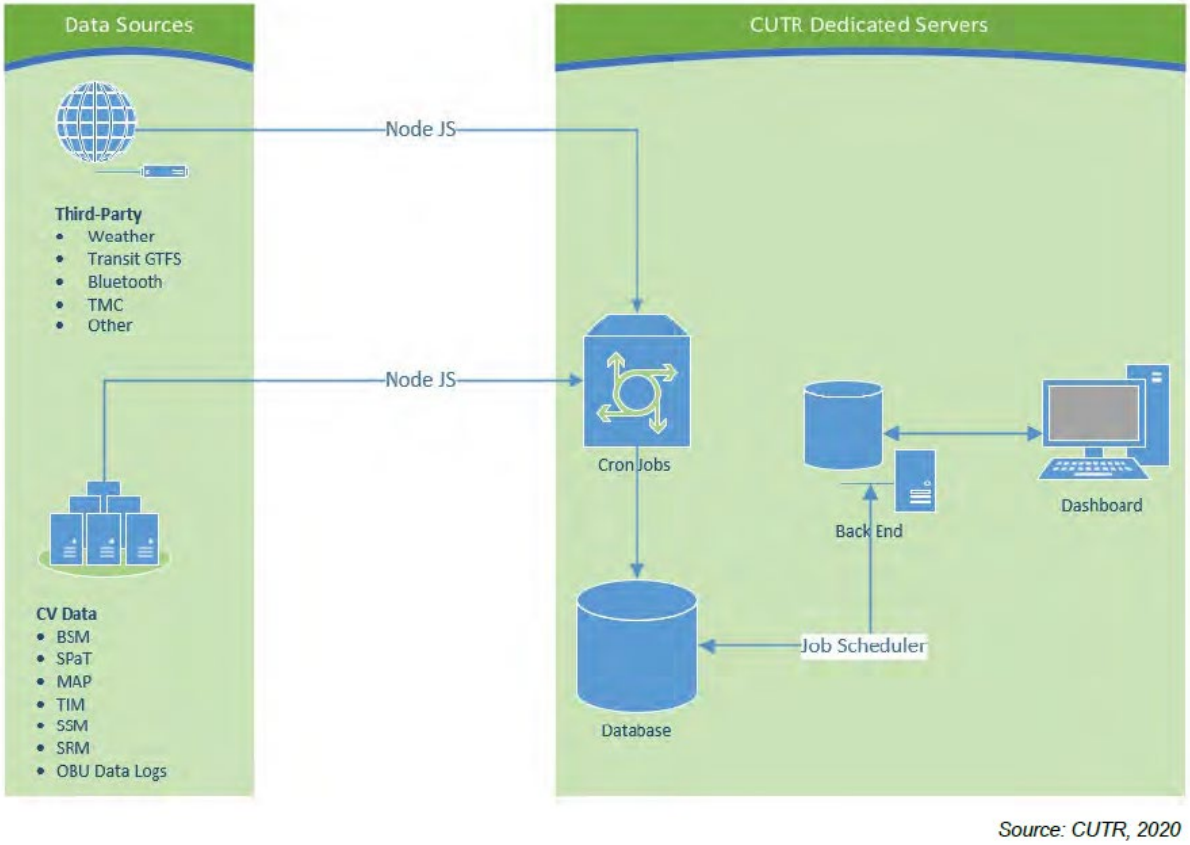


Figure 2.8 THEA CV Pilot Dashboard Components
Source: *THEA CV Pilot Performance Evaluation Dashboard, 2020*

Figure 2.9 Measurement Dashboard Page
Source: *THEA CV Pilot Performance Evaluation Dashboard, 2020*

Figure 2.1 Performance Dashboard Page
Source: *THEA CV Pilot Performance Evaluation Dashboard, 2020*

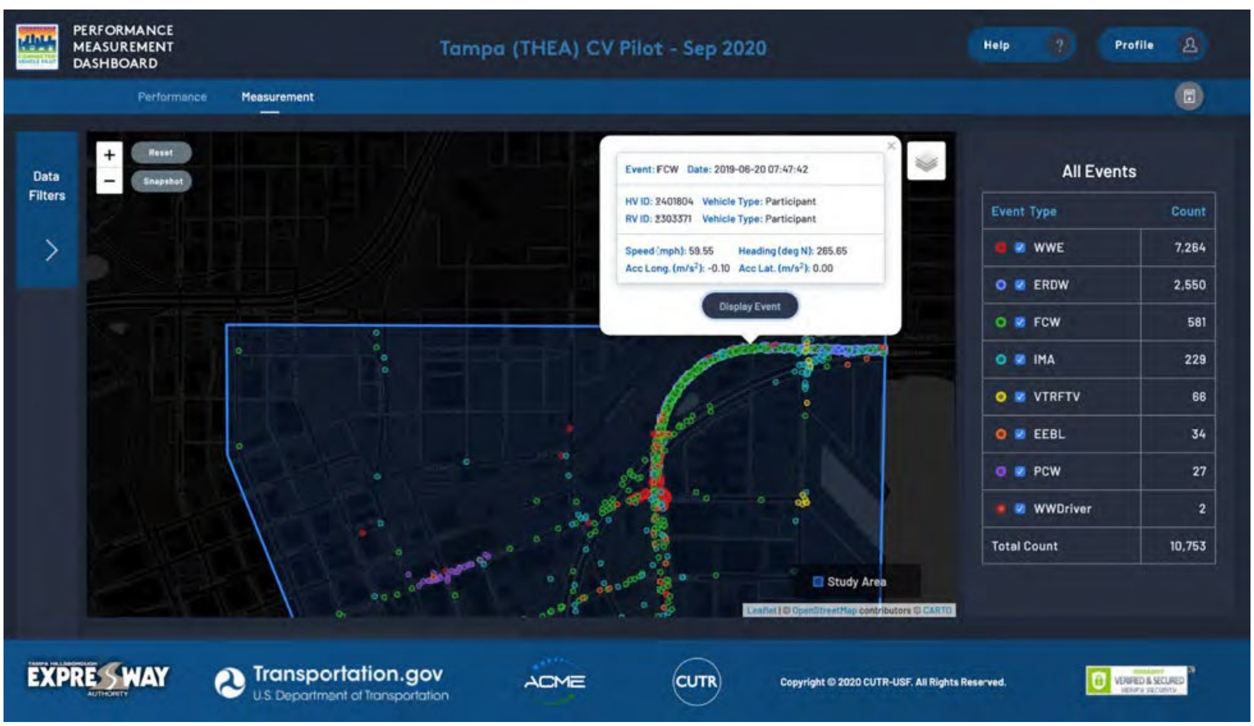


Figure 2.2 Warning Event Profile Option
Source: *THEA CV Pilot Performance Evaluation Dashboard, 2020*

2.6 Synthesizing findings for Tennessee: Lesson Learned

The following is a synthesis of the material presented about CAV data collection in this chapter.

- TDOT can collect CAV data and use CAV data dashboards to visualize the data and use it in near real-time to improve transportation system performance, including traffic flow, congestion, and incidents at the regional, corridor, and intersection levels. Note that the issue of TDOT not owning and operating traffic signals in Tennessee is critical and options in this regard are discussed in earlier project reports. Successful deployment of CAV technologies and the collection or use of associated data will require TDOT to form collaborative partnerships with local authorities (Infrastructure Owners and Operators or IOOs) to implement CAV initiatives, especially regarding traffic signals.
- TDOT should consider developing smart infrastructure strategies (e.g., high-resolution cameras by Gridsmart or Miovision) to collect and use non-CAV data in collaboration with partner agencies.
- TDOT can use the structure and categorization followed by the US DOT's ITS Data Hub Dashboard for consistency in CAV data management and sharing.

Chapter 3 Data Analytics for Infrastructure Technologies (Task 5)

3.1 Introduction

Investment in harnessing CAV data is critical for TDOT. This can substantially improve transportation system performance in Tennessee.

Data analytics and modeling are needed to use the data effectively. This can include visualization of the collected data to measure system performance in real-time and for tactical/strategic planning. The analytics include visualization of CAV data. Specifically, BSMs are data packets related to a vehicle's position, heading, speed, acceleration, state of control, and predicted path. These data can be transmitted from one vehicle to other vehicles via V2V and V2I communications. In a real-life application, they are analyzed by the receiving unit to determine the presence of hazardous situations and alert the driver of the host vehicle accordingly. Storing and analyzing these messages can provide insights into whether the alerts were given appropriately and if they were effective in avoiding hazardous situations. Similarly, TIM provides drivers with information about traffic incidents, major events, and even evacuations. These messages utilize V2I communications and are sent to vehicles by RSUs. Furthermore, SPaT messages contain data about the state of signal phases at an intersection and related information. SPaT messages are processed by vehicles to support driver/vehicle decision-making at an intersection, e.g., whether to stop or go at a signalized intersection. The point is that these data are analyzed to improve the transportation system's performance, e.g., in terms of safety and mobility, as well as these messages, can be analyzed for their effectiveness and harnessed more generally to improve system performance. Examples of harnessing the data are provided in this report. Modeling the data and applications of Artificial Intelligence (AI) have gained momentum in this realm.

Investments in harnessing data and applying AI can substantially improve transportation system performance. The research team has developed several case studies that demonstrate the application and harnessing of data from CAV infrastructure technologies. To demonstrate, the team has developed a framework to harness big data from CAVs (e.g., using CAV speeds, volatile driving, and time to collision) and analyzed recent CAV disengagement data. The research team has also provided information about applications of CAVs for multiple modes that include transit and pedestrians. A list of the studies is presented below. These are meant to guide TDOT in deploying the entire cyber-physical ecosystem for CAV technologies and smart infrastructure.

Vehicle to Infrastructure Applications

- Study 1: MLK Smart Corridor
- Study 2: Synthesis of V2I applications in THEA CV Pilot Project
- Study 3: Enhanced Pedestrian Collision Warning System (EPCW) for bus drivers

- Study 4: Cooperative Merging at Ramps

Vehicle to Vehicle Applications

- Study 1: Safety-critical Applications: Safe Pass Advisory
- Study 2: Platooning-Cooperative Adaptive Cruise Control
- Study 3: Pedestrian Crash Prevention Systems

Harnessing CAV Data

- Study 1: Driving volatility helps identify hazardous intersections
- Study 2: Easing in Automated Vehicles-Experimentation in mixed traffic
- Study 3: Predicting future crashes more accurately with CAV Data
- Study 4: Identifying hazards through Automated Vehicle disengagements

3.2 Study Design and Modeling Methods

Randomized and quasi-experimental designs can be used to study the impacts of specific technology applications. Experimental design is critical to evaluating the impact of a specific technology in a situation. One example is to recruit a panel of drivers in a testbed and then use panel data experimental design for drivers who experience alerts and warnings (or even control assists) from the V2V and V2I applications and those who do not receive warnings. Some of the equipped participants will receive warnings via a display in a randomized experiment, while others will not.

Usually, data analysis relies on information from RSUs and travel logs stored by OBUs. The data are transmitted and stored in databases that can be used for analysis. Usually, millions of V2V and V2I communication data points can be analyzed. The data can be harnessed for research on giving warnings to drivers and whether a tested technology correctly identified hazardous situations and appropriately issued warnings to drivers.

In transportation, mobility and safety benefits are evaluated, e.g., for safety applications, the analysis can include identifying false positives and false negatives for warnings. A before-after assessment with interrupted time series can be used for mobility applications.

There is an increasing array of statistical modeling tools available in data science. These can include linear regression, Poisson and negative binomial regressions, discrete choice models, cluster and factor analysis, random-effects models, and panel-data models. Lately, Bayesian models have gained popularity. These models are based on the Bayes theorem and integrate prior information with new information.

Analytics and modeling increasingly include AI-driven approaches that include random forest, neural networks, reinforcement learning, and natural language processing. AI is being applied to improve transportation system performance, e.g., traffic signal controller performance. With AI, more educated decisions by system managers can be made to reduce time spent by vehicles waiting to be serviced or time wasted giving green lights to empty roadways. Traffic signals can further benefit from cooperative actions with other traffic signal controllers. Traffic controllers can use reinforcement learning in AI applications, especially Dyna-Q learning methods, to improve traffic signal performance.

3.3 Vehicle-to-Infrastructure Applications

User service packages form the core of intelligent transportation system technologies that can be considered for deployment in Tennessee. This section presents a list of CAV technologies deployed relatively broadly nationwide as use-cases for consideration by TDOT. TABLE 1 illustrates a list of specific CAV technologies deployed and the types of vehicles that received the onboard units. Also, the table provides information about whether or not the US/State DOT did the test by partnering with local agencies. It is evident from the data that in some cases, OBUs were deployed on public vehicles, i.e., the devices were deployed on agency fleet vehicles, such as transit buses, HELP trucks, and Fire trucks. Use cases also illustrate that several use cases did not have local agency partners. Given this and the substantial activity in using public vehicles for OBUs, TDOT can consider deploying OBUs without partnering with local agencies, e.g., when deploying wrong-way-detection technologies on interstate ramps. In other cases, host agencies can be considered for partnerships, especially when deploying CV technologies related to traffic signals or public transportation.

TABLE 1: LIST OF DEPLOYED CAV TECHNOLOGIES FOR TDOT CONSIDERATION

Technology	*OBUs Deployment on*	*State DOT Partnership*
End of Ramp Deceleration Warning	Public Vehicles	Without a Partnering
Pedestrian Collision Warning	Public Vehicles	Partnering with local agencies
Transit Signal Priority	Agency Vehicles: Bus	Partnering with local agencies
Wrong-Way Entry	Public Vehicles	With and Without a Partnering
Intelligent Traffic Signal System	Public Vehicles	Partnering with local agencies
Bus Pedestrian Collision Warning System	Agency Vehicles: Bus	Partnering with local agencies
Cooperative Merging at Ramps	Public Vehicles	Without a Partnering
Cooperative Adaptive Cruise Control	Public Vehicles	Without a Partnering
Video/Audio Sensors	No OBUs	With and Without a Partnering

Study 1: MLK Smart Corridor

The City of Chattanooga is a pioneer in urban renewal and sustainable development. It is internationally recognized as one of the most innovative smart cities, partly because of the contributions of the University of Tennessee at Chattanooga (UTC) and Chattanooga's locally owned electric distribution and communication provider, Electric Power Board (EPB). In 2009, EPB deployed a 600-square mile fiber-optic network that provides up to 10 Gbps Internet service to over 87,000 households and businesses. The Center for Urban Informatics and Progress with initial internal investment from UTC and UT System, operational support from Chattanooga Department of Transportation, communications infrastructure from EPB, and design support from The Enterprise Center, has launched an urban testbed in downtown Chattanooga named

MLK Smart Corridor. This corridor covers about 1.5 miles and consists of 16 poles, each containing some combination and permutation of the following technologies. A wide array of sensors and communication devices are deployed at each intersection, as listed below.

- Internet of Things: Video Sensors; Audio Sensors; Air Quality Sensors
- Communication: DSRC Road-Side-Units and dual-mode DSCR and C-V2X units; LoRaWAN Gateways; Wifi Access Points; Software Defined Radio (SDR)
- Edge Computation: Industrial Computer; Raspberry Pi; Graphics Processing Unit Resources

MLK Smart Corridor is built modular and programmable to ensure additional sensors and capabilities can be augmented easily. The poles are connected to EPB's existing gigabit fiber network, allowing a backhaul for data transmission at the low latency and high throughput needed to make real-time decisions.

UTC developed a data platform that supports data collection, storage, processing, and monitoring of data generated on the MLK Smart Corridor. The core component of this infrastructure is a distributed event-driven architecture. This architecture is designed for high-availability, high-throughput, and low-latency operations. Systems and applications communicate with the Data Hub via the client application programming interfaces (APIs). Client applications publish data to dedicated streams organized by categories or data types. Applications or services can subscribe to those same streams to consume real-time data. UTC was developed using lightweight native libraries that efficiently expose functionality to integrate data into the system. Additionally, WebSockets or REST APIs can ingest or push data into the system.

One service that utilizes the real-time component of the data infrastructure is UTC's dashboard. The dashboard provides a single-source destination for monitoring the health of devices on the testbed and data generated by these devices and applications. Figure 3.1 shows the view of the dashboard for a single intersection. Metrics shown include volume by approach, air quality metrics, near-miss incidents, average pedestrian wait time, and percent arrivals on green and red. Data sources can be easily integrated into the dashboard using open-source connectors. Video streams are also monitored via the dashboard.

Data ingested can be analyzed in a variety of ways. Archived data stored in the cloud can be analyzed using standard SQL or NO-SQL queries. Real-time data can be processed using the Streams module, enabling users to process real-time data based on single-message logic. For example, connected vehicle data received and generated on the RSUs are ingested into this infrastructure. SPaT, MAP and BSM messages are currently collected and stored for processing. The method used to ingest CAV data consists of a push-based architecture where the RSU is configured to push these message types to a pre-configured IP address and port number over the User Datagram Protocol communication protocol. This data is pushed into the cloud and analyzed every hour. UTC generates meta-data for every message type and RSU location.

MLK Smart Corridor deployment results provide a blueprint for future cities that wish to adopt the smart city concept. The testbed has been continuously ingesting data for extended durations (years). A data platform is deployed to provide a standard means of producing and consuming data to and from the infrastructure. This platform allows new systems and technologies to be easily integrated into the ecosystem. The system features low-latency and high-throughput support, which was a key design constraint from the beginning. With the open platform design implemented, researchers and developers can test new technologies in a live urban environment. This ability has already been used, allowing researchers to produce and ingest data in real-time, test new hardware, and access data from new hardware or algorithms with ease. This design has the potential to become a real-time dashboard for citizens to improve their day-to-day health, mobility, and transportation.

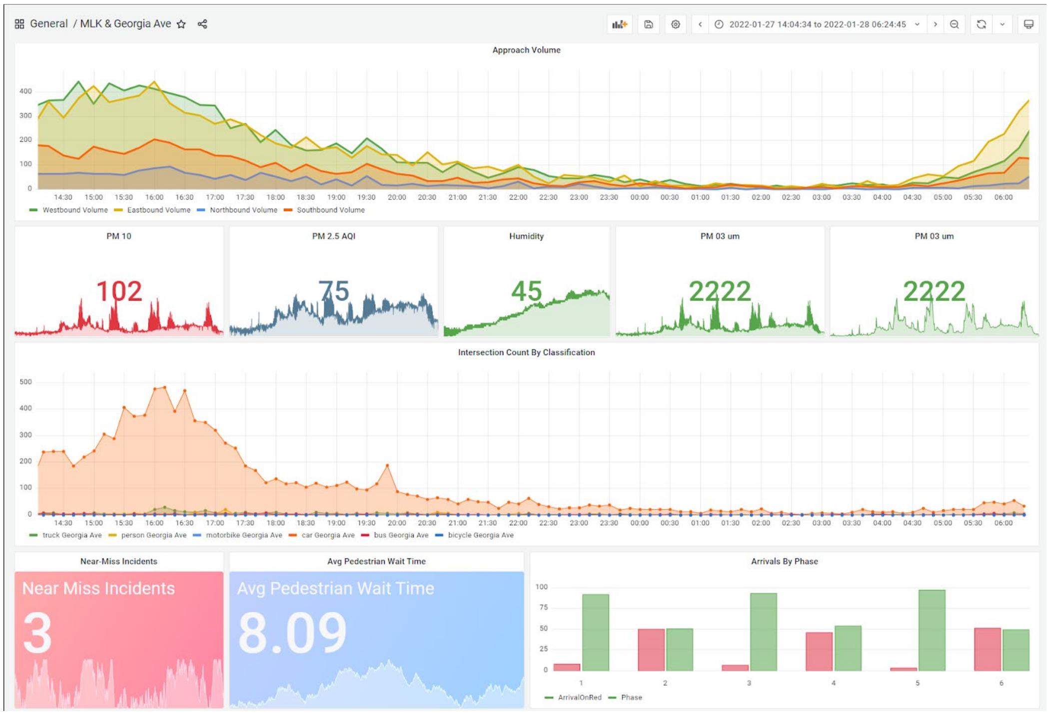


Figure 3.1 MLK Smart Corridor Data Dashboard and Visualization

In the future, the real-time data currently being collected at the testbed will be used for a simulation environment. These simulations will mimic the current state of the conditions on the testbed. The data will also be used to create cyber-physical systems to support the transportation network. Localized intersection processing applications will trigger detections based on the data currently being generated by the computer vision application. The lane, speed, and predicted trajectory of inbound vehicles will be used to optimize ITS systems and reduce congestion [3, 4].

Study 2: Synthesis of V2I applications in THEA Pilot Project

The CV Pilot test being conducted in Tampa, Florida, has generated several datasets from interactions between vehicles (via OBUs) and between vehicles and infrastructure (OBU/RSU interaction). Vehicles traveling or operating generated BSM and alert data, which were collected by roadside units and transferred over the air to THEA's secured master server. The following synthesizes the application of V2I devices deployed in the THEA CV Pilot project.

End of Ramp Deceleration Warning

One of the main concerns about CAVs is whether they can discern the appropriate stopping sight distance, especially when a queue of vehicles exists. One of the V2I applications could be attributed to solving this issue and improving safety, which is implemented in THEA CV Pilot. The

End of Ramp Deceleration Warning (ERDW) calculates the geolocation of vehicle queues based on the longest lane queue length computed by the Intelligent Traffic Signal System (I-SIG) application to inform drivers about queue length in the roadside unit (Figure 3.2). An Infrastructure Sensor Message (ISM) is generated when a vehicle passes a traditional vehicle detector. I-SIG applies the ISM to enhance its queue length estimation. The Reversible Express Lanes are divided into one or more speed zones. Based on the end of queue location, the RSU sends a Traveler Information Message (TIM) with the recommended speed for each zone based on the appropriate safe stopping distance. During phase 3 of the deployment, Siemens mobility the following queue length estimation solution was provided to improve queue length estimation. The multinomial Intelligent Traffic Signal System queue length estimation is replaced using queue length measurement derived from the participants' BSMs, which enhances the reliability of the whole system. The results indicate that TDOT can implement ERDW in the future to solve safety issues about CAVs stopping sight distance and follow the new procedure of THEA CV Pilot to estimate queue length.

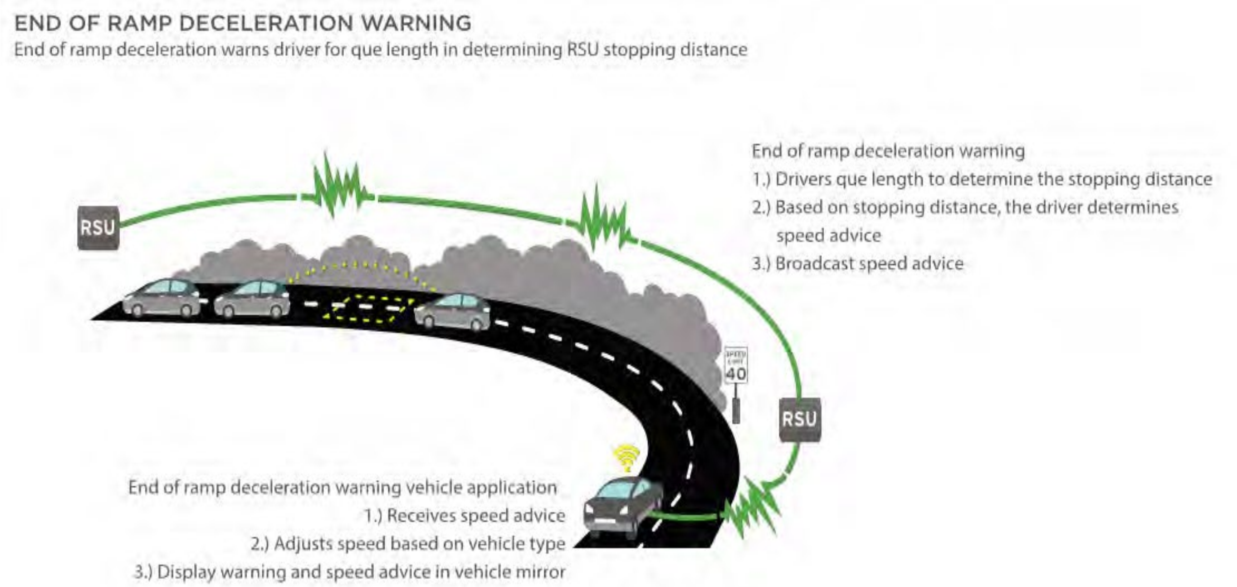


Figure 3.2 End of Ramp Deceleration Warning application
Source: *System Architecture Document, Publication FHWA-JPO-17-459, 2018*

Pedestrian Collision Warning

Pedestrians and vehicles are inseparable parts of the transportation system. One of the critical aspects of diffusing CAVs is that they could finally obtain the ability to detect pedestrians completely. The main goal of Pedestrian Collision Warning (PCW) is to improve pedestrian safety. This technology was implemented in Hillsborough County in THEA CV Pilot, which serves as a use case. Initially, two LiDAR sensors were installed, which can transform the information into the Pedestrian Safety Message (PSM). Then, PSMs are sent to the OBUs through DSRC. By doing so, drivers of OBU-equipped vehicles cab be aware when a collision condition with a pedestrian in the crosswalk exists (Figure 3.3). However, after implementing this technology, it was concluded that the operational reliability of the LiDAR sensors was not adequate. Hence, the LiDAR sensors were replaced with video and thermal imaging sensors. The thermal imaging sensors detect heat radiated as infrared light by pedestrians, bicyclists, and internal combustion vehicles. The video imaging sensors detect light reflected from objects, pedestrians, bicyclists, and vehicles. With these changes, the PCW system started full operation. TDOT's Office of Public Transportation promotes public transportation by providing financial and technical assistance to transit agencies in Tennessee. They also perform transit planning and operations assistance. The implications of this use case for TDOT are to consider partnering with transit agencies and local agencies (IOOs) interested in implementing PCW technology and associated video and thermal imaging sensors.

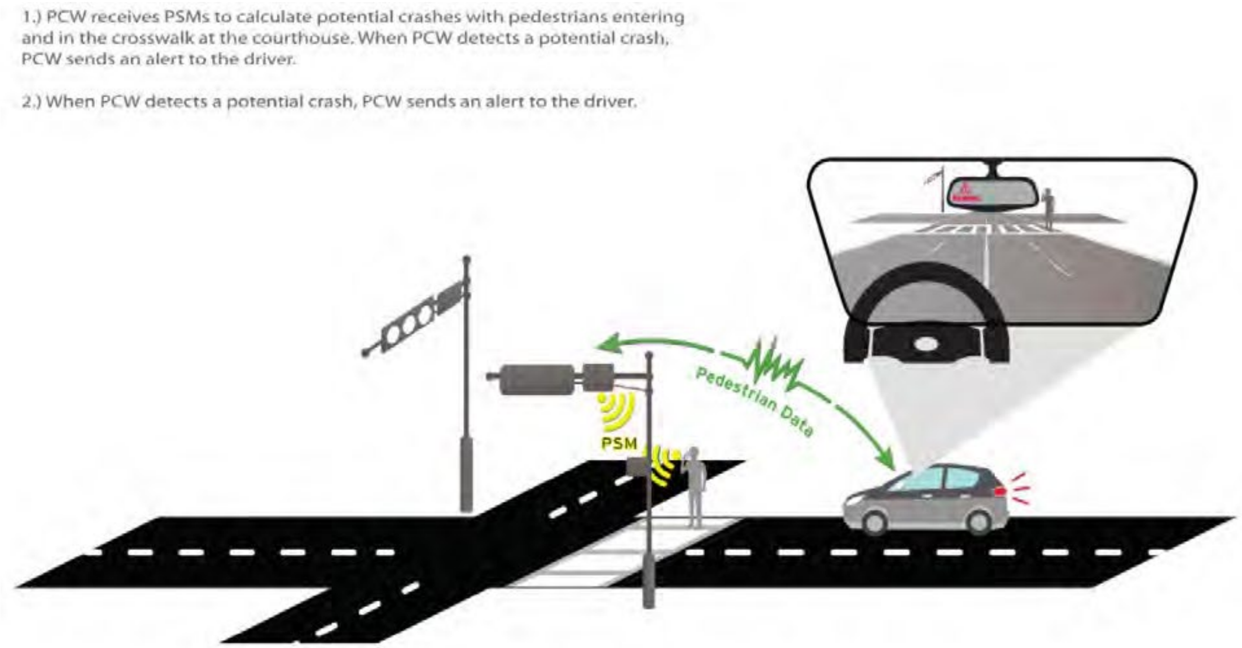


Figure 3.3 Pedestrian Collision Warning application
Source: *System Architecture Document, Publication FHWA-JPO-17-459, 2018*

Transit Signal Priority

A key user service is priority detection among different vehicle types at signalized intersections. One of the most practical applications of V2I technology is the application of transit signal priority (TSP). If a transit bus is behind schedule, then priority can be requested and granted to the bus at intersections. This can substantially reduce delays at an intersection for transit vehicles by using communication between a Signal Request Message from the transit server and the RSUs. Figure 3.4 and Figure 3.5 demonstrate the functional overview and flows of the application. In large urban areas served by public transit, TDOT can consider encouraging the use of transit signal priority with local partners. Because the priority of different vehicle types is a critical issue in the operation of urban transportation systems, TDOT can consider TSP in partnership with local agencies (IOOs), where TDOT provides the funding to deploy TSP technology and the local agencies operate and maintain TSP technologies.

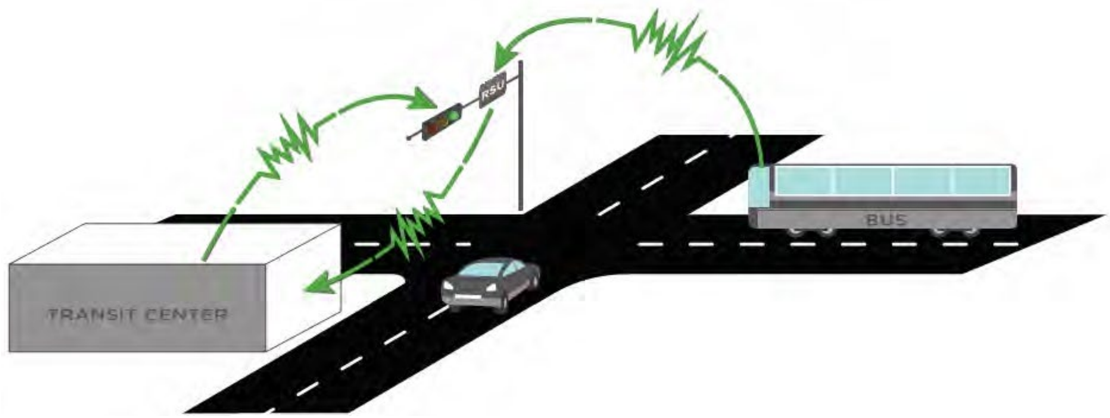


Figure 3.4 Transit Signal Priority overview
Source: *System Architecture Document, Publication FHWA-JPO-17-459, 2020*

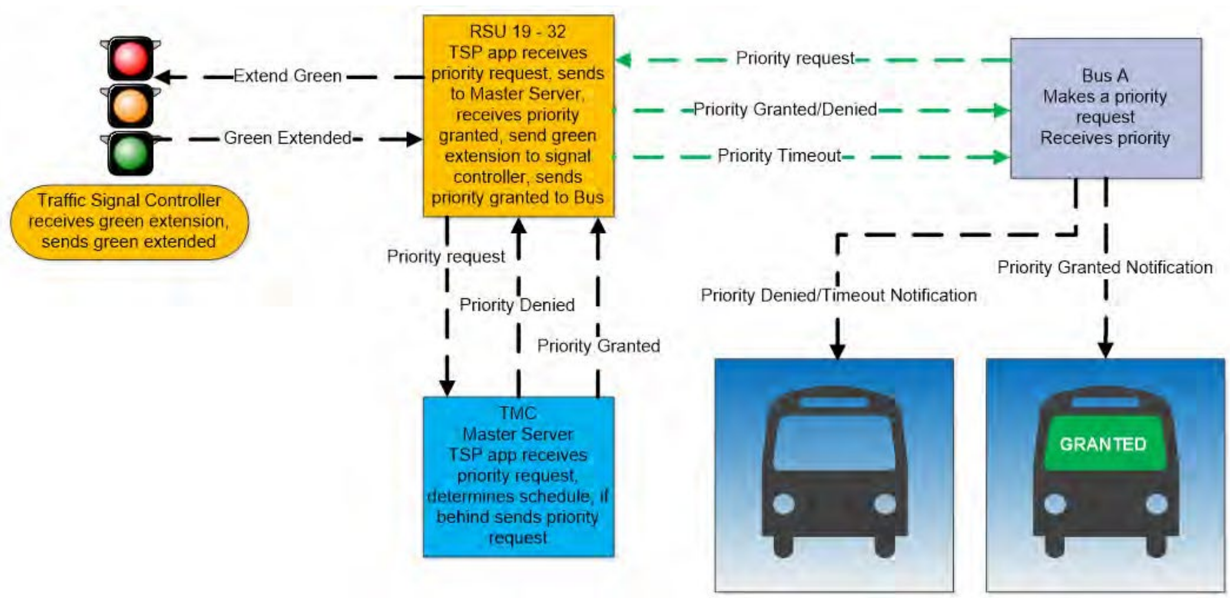


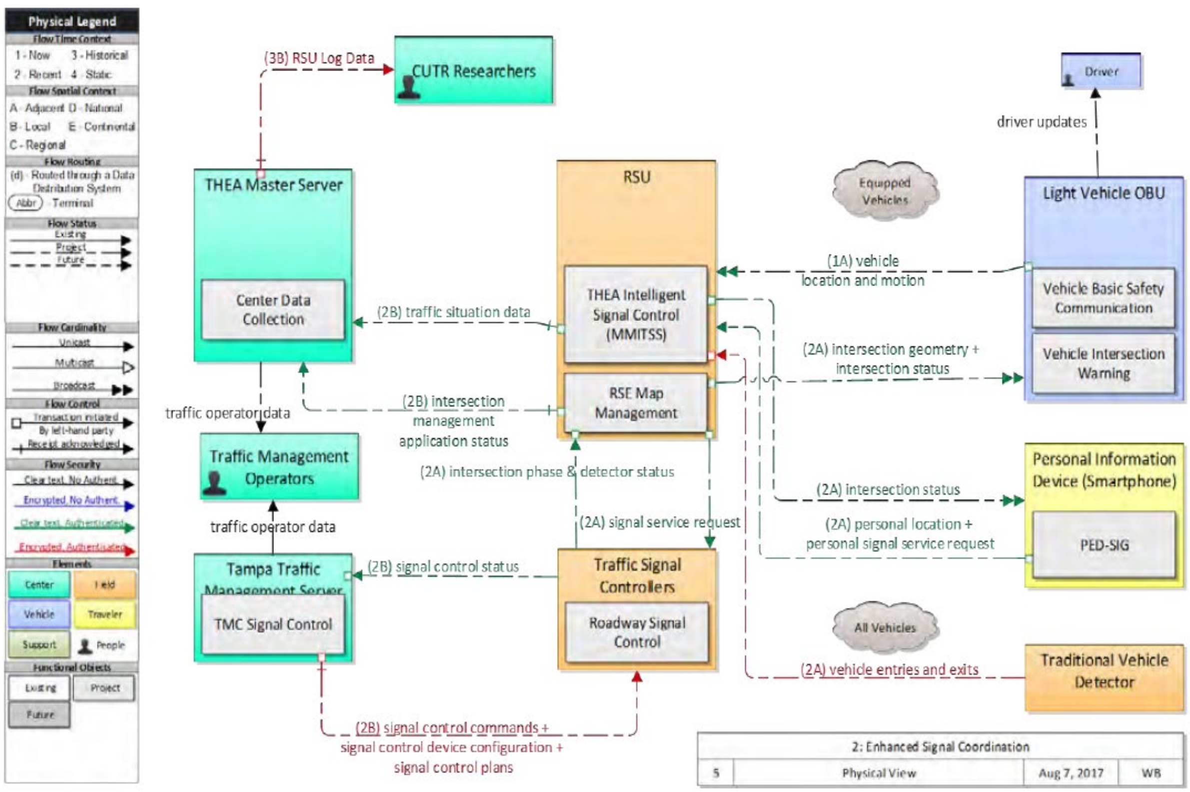
Figure 3.5 Transit Signal Priority functional flows
Source: *System Architecture Document, Publication FHWA-JPO-17-459, 2020*

Wrong-Way Entry

Wrong-way entry (WWE) of vehicles is a significant problem on roadways, with potentially dire consequences, reflected in head-on collisions. CAVs can also face similar situations in the future and comprehensive plans should be considered before testing CAV vehicles in pilots to address this issue. According to the THEA CV pilot, the WWE application warns OBU-equipped vehicles approaching an RSU-equipped intersection when the vehicles are not traveling in the allowed direction (and in the wrong way). The WWE application has multiple driver warning levels recorded with the same warning type (WWE) in the OBU data logs, including DO NOT ENTER, WRONG-WAY, NO TRAVEL LANE, and WRONG-WAY VEHICLE. The OBU analyzes vehicles' trajectory, speed, and allowed movements and then determines the appropriate message to be displayed. TDOT can consider this V2I application in a testbed environment. More comprehensive implementation of the technology will require many OBUs to be deployed in vehicles and partnerships between TDOT and traffic signal owners and operators.

Intelligent Traffic Signal System (I-SIG)

Reduction of delay at intersections is critical for large and medium-sized urban areas. According to THEA CV Pilot, the I-SIG transforms information into a Multimodal Intelligent Traffic Signal System to estimate queue length at intersections and other traffic delay measures to improve traffic progression in the relevant roadway sections (Figure 3.6). Field testing in a use case has shown that the most reliable queue length prediction can be made with the I-SIG application. Along with IOO partners, TDOT can consider testing and deploying I-SIG technology at intersections in suitable conditions [5].



Source: System Architecture Document, Publication FHWA-JPO-17-459, 2018

Figure 3.6 Traffic Progression Physical architecture

Study 3: Enhanced Pedestrian Collision Warning System (EPCW) for bus drivers

Among all road users, pedestrians are the most vulnerable as they are exposed and are difficult to observe while driving, irrespective of the time of the day [6]. According to the National Highway Traffic Safety Administration (NHTSA), “6283 pedestrians were killed in traffic crashes in 2018, of which 17% occurred at intersections [7].” The emergence of connectivity and automation technology indicates a reduction in overall crash risks in vehicles [8]. Therefore, it is crucial to harness the data from CAV pilot projects and explore their benefits to the drivers and road users at intersections. Thus, the objectives are: 1- To harness data from connected transit buses to understand the vulnerability of the pedestrians at intersections; 2- To explore the trend of EPCW alerts at different types of intersections; and 3- To identify alert zones within the intersections where bus drivers should be more careful to avoid bus-involved pedestrian crashes.

The data used for this research is BSM and alert data from the “Enhanced Transit Safety Retrofit Package (E-TRP)” project collected between February-August 2018. The study area is Cuyahoga County, Cleveland, Ohio. Three key cutting-edge technologies are used in this project. First, a DSRC system is used for V2I communication. Second, a High-precision Global Navigation Satellite System is used for tracking connected vehicles. Third, Forward-Looking Infrared cameras are used for detecting pedestrians. The installations on the 24 transit buses are called on-board transit vehicle-based subsystems, and the installations at the three pedestrian crossings are called infrastructure-based subsystems [9], as shown in Figure 3.7. The CV application used in this study is called Enhanced Pedestrian in Crossing Warning (EPCW), which alerts the driver of the connected bus about pedestrians at the curb and crosswalk. Other collected information includes position (latitude/longitude), timestamp, and speed information of the connected buses (*28*). Alerts are also classified under events, an incident for which an alert is stored when given to the driver. By identifying alert zones and observing the trend of a safety surrogate measure (SSM) over the study period, the efficacy of EPCW alerts in improving pedestrian safety is examined.

Alert Zone identification

To identify alert zones, the distance (meters) between the location of each type of pedestrian-based alert and the center of the intersection is measured in GIS. The average distance of all types of pedestrian alerts is considered the radius of the alert zones.

A unique SSM called “Mean Time Difference to Intersection” (MTDI) is developed. It denotes the mean time a connected bus would take to reach the intersection from the moment of receiving “WarnAlert” at an intersection, considering that the bus would be moving at the speed at which “WarnAlert” was received. The equation to calculate MTDI is as follows:

MTDI = time to reach intersection from the moment of receiving the “WarnAlert” (TI)/Frequency of “WarnAlert” in the intersection (F)

Figure 3.7 Onboard Subsystem (top) and Infrastructure Subsystem (bottom) (28)

It was observed that the location of receiving pedestrian-based alerts varied depending on the bus stop location. The midblock crossing received only “FarsidePed” and “NearsidePed” alerts, whereas the other two intersections received all four types of pedestrian-based alerts. Based on the alert frequency, the pedestrians were found most vulnerable at the signalized intersection of this study, followed by unsignalized and midblock intersections. However, when connectivity is deployed, the situation might reverse. As a result of receiving safety alerts, signalized intersections can become the safest intersections for pedestrians, followed by unsignalized and midblock intersections. Next, Alert zones were identified within a radius of 43.31 meters, 31.13 meters, and 51.75 meters from the center of the signalized, unsignalized, and midblock crossing intersections, respectively. Presumably, if transit drivers are more vigilant when entering and driving through these alert zones, then they may be able to avoid collisions with pedestrians. Finally, the novel SSM MTDI indicated that the unsignalized and midblock crossing intersections were safer than the signalized intersections for pedestrians even after the application of connectivity. The research report on the E-TRP pilot project found that connected bus drivers’ response to EPCW alerts increased by 16% over the six-month study period, and average drivers’ reaction time to the EPCW warning situation decreased by 18% [9]. These SSMs indicate that connectivity can help bus drivers to react more quickly over time and increase pedestrian safety.

TDOT and partner agencies can consider installing such connectivity systems in transit buses (OBUs) and critical intersections (RSUs) along their routes. They can collect BSMs and alerts and then analyze the data to identify alert zones for each intersection. TDOT can share this information (alert zones for intersections) with the transit drivers to identify pedestrians and drive safely in and around hazardous intersections. Checking the MTDI of the intersections over time can help TDOT and partner agencies identify dangerous intersections in urban areas of Tennessee and implement safety countermeasures. Overall, the EPCW system can improve the safety of pedestrians at hazardous intersections from bus-involved pedestrian crashes [10].

Study 4: Cooperative Merging at Ramps

Real-world CAV testing helps developers recognize and fix system limitations and drawbacks, not just on individual vehicles but across fleets [11, 12]. Cooperative Adaptive Cruise Control (CACC) has the potential to improve system performance by reducing human driving tasks [13, 14]; cooperative merging scenarios requiring V2I communication show promising results based on data from the CARMA2 project. The system runs on a Linux computer inside a vehicle, and the computer interacts with the vehicle's devices and microcontrollers via the vehicle's controller area network. The computer also interfaces with the OBU, which functions as a two-way radio for DSRC connected with other vehicles and the infrastructure. Moreover, the computer interacts with after-market sensors, such as radars. This study provides new knowledge about potential safety and environmental improvements from vehicle connectivity, which will be useful for researchers, developers, and implementers. The concept of driving volatility and time-to-collision are utilized to quantify the safety performance and fuel consumption and emissions are calculated to measure the environmental impacts.

Driving Volatility

Volatility measures are used to quantify driving variation. Volatility measures try to capture variations in longitudinal control of the vehicle. To this end, these measures can be applied to speed, acceleration/deceleration, and vehicular jerk.

Time-to-Collision Measures

Time-to-Collision (TTC) is a surrogate safety measure that is generally defined as "the duration of time before two objects collide with initial certain conditions" [15]. This measure has been used vastly to assess the risk of rear-end collision.

Fuel Consumption and Emissions

The model proposed by Kamal et al. [16, 17] is used to calculate fuel consumption. Their proposed equation takes advantage of the relation between speed, acceleration, and fuel consumption. The vehicle-specific power microscopic model estimates emissions regarding vehicle second-by-second speed, acceleration, and terrain gradient [18].

This study takes advantage of analyzing a unique dataset collected during US DOT's CARMA program [19]. The data represent a proof-of-concept vehicle platooning based on the ACC and CACC applications. The dataset consists of a scenario including the cooperative merge scenario, as shown in Figure 3.8. The scenario contains the variables of speed, acceleration/deceleration, and the position of the vehicles in a fleet of five vehicles, including a lead vehicle, a merging vehicle, and three following vehicles.

Results show that vehicles equipped with CACC substantially reduce driving volatility as a safety measure in a five-vehicle-platoon. The cooperative merging system through V2I communication reduces the volatility of the merging vehicle by 6.2% compared to a manually driven merging vehicle. The results reveal that the amount of fuel consumption for the merging vehicle slightly increases by 0.54% compared to the manually driven mode. Regarding the environmental impact, the merging vehicle in the CACC mode emits slightly more gas (4.1%). Fuel consumption and emissions can increase due to the ability of the CACC algorithm to deal with higher acceleration limits to form a stable platoon. This increase can be improved through an adjustment to the CACC control algorithm for accelerations that allows for the energy benefits during acceleration.

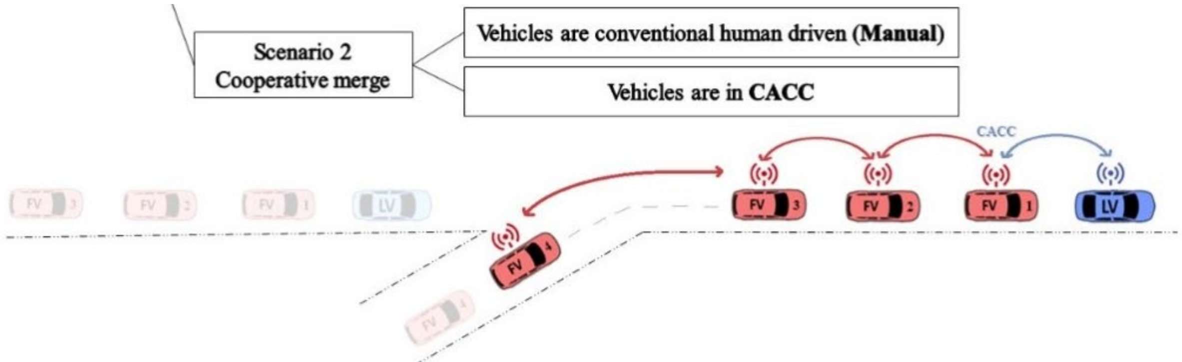


Figure 3.8 Definition and visualization of the cooperative merging scenario

V2V and V2I connectivity in CACC systems may lead to substantial improvements in sustainable mobility. There is a need for more empirical evidence regarding the impacts of automation on wide-scale deployment. The methods used in this study are repeatable; the results reduce the uncertainty of vehicle technologies in terms of their impacts. To further improve the confidence of these results, it will be necessary to collect even more extensive samples of data. Since the CARMA Platform allows automated vehicles (AVs) to interact and cooperate with infrastructure (such as on-ramp merging infrastructure) and other vehicles via communication under a federal project, it may be inevitable for TDOT to prepare its infrastructure to accommodate such technologies. Therefore, these technologies can be deployed on a large scale in Tennessee [20].

3.4 Vehicle-to-Vehicle Applications

Study 1: Safety-critical Applications: Safe Pass Advisory

Vehicle-to-vehicle (V2V) communication promises to help reduce vehicle collisions and increase traffic flow stability. Fast communication allows vehicles in a highly mobile and complex network to send and receive safety messages. However, many factors can cause a safety-critical automotive application to be unreliable due to communication failures. While the reliability of V2V communication has been a subject of study by several researchers, there are still open questions regarding how the placement of the DSRC devices (inside or outside the host vehicle), the vehicle's interior elements, and the differences in altitude can affect the V2V communications. This study provides experimental testing data and analyses to quantify the impacts of relative vehicle speeds, altitude differences between vehicles, and interior obstacles on V2V communication range and reliability for opposite traffic in Tennessee's city and rural highway environments. A theoretical model was first developed to evaluate the feasibility of implementing the "Safe Pass Advisory" application. Furthermore, efforts were made to formulate the constraints of DSRC communication ranges and the minimum duration for seamless communication using kinematic equations. Then a series of field experiments were conducted with two vehicles approaching from the opposite direction on an interstate freeway (I-26) to estimate the actual communication duration and range available between opposite traffic to validate the theoretical model.

The research team discusses how these results can adversely affect the design parameters of safety-critical applications by considering the V2V application "Safe Pass Advisory" on two-lane rural highways. Freeway experiment results indicate that the communication range and connection period between two vehicles driving in opposite directions are more than doubled if the OBU is mounted on the rooftop instead of inside the vehicle. If the OBU is placed inside the vehicle, then the communication range in the backward direction is less than half of the range achieved in the forward direction. This phenomenon occurs because it only goes through the windshield when the signal propagates in a forward direction. On the contrary, when the signal propagates in the reverse direction, it must find its way through backseats and trunk/cargo spaces, struggling with more obstacles. However, if the OBU (or just the DSRC antenna) is mounted on the roof, then the effective communication range is almost the same in every direction surrounding the vehicle. The experimental results from the city environment provided some crucial information about how even with a rooftop OBU, the V2V communication is obstructed by changes in altitude within a straight road segment. Changes in altitude translate into additional constraints for designing a safe pass advisory application involving two-lane rural highways.

Based on the analyses, a typical truck-passing maneuver on a two-lane U.S. highway with a speed limit of 55 mph takes at least 12.16 s and a minimum of 712 m (2336 feet), considering the maximum acceleration for that speed. From the experimental results, two oncoming vehicles traveling at 55 mph can only start communicating when they are a maximum of 466 m away. However, with a rooftop OBU, two vehicles can start communicating at a maximum of 800 m apart, making it possible to implement the safe pass advisory application. Multi-hop V2V communication increases the potential for actual implementation.

In summary, the application of this study is as follows:

1. Development of mathematical analysis to understand the required V2V communication parameters and constraints about a DSRC-based "safe pass advisory" application for two-lane rural highways.
2. Experimental quantification of V2V communication ranges and connectivity periods between two DSRC-equipped vehicles approaching each other from opposite directions with various speeds for both freeway and city environments. The data obtained from this experimentation helped determine the time and distance constraints for the proposed "safe pass advisory" application.
3. Investigation of the impacts of varying altitudes on V2V communication reliability and its implications for safety-critical applications.
4. Evaluation of vehicle-interior obstacles and OBU placement impacts on the reliability, range, and duration of V2V communications, both in the forward and reverse directions. The collected experimental data provided insights for utilizing multi-hop V2V communication to overcome the limitations of the "safe pass advisory" application.

The implication of this study is for TDOT to consider disseminating the results to the freight industry that operates in Tennessee to increase their safety and awareness. Deployment of this V2V technology in trucks, i.e., to receive "safe pass advisory" messages, can decrease potential truck-involved crashes when passing on two-lane two-way rural roadways in Tennessee [21].

Study 2: Platooning-Cooperative Adaptive Cruise Control

Advanced driver-assistance systems (ADAS), such as ACC and CACC, may lead to substantial improvements in traffic networks. However, the impacts of these technologies are uncertain and the amount of improvements in safety, energy consumption, and emission reduction is not well-known. On-road testing and early deployments are critical to enhancing the performance of vehicle automation. Safe deployment of ADAS is crucial and predominant to the U.S. DOT's line of action. Real-world AV testing helps developers recognize and fix system limitations and drawbacks, not just on individual vehicles but across fleets [11, 12]. ACC and CACC have the potential to influence traffic safety, energy, and emission-related issues by reducing human driving tasks [13, 14]. At low levels of automation, the ACC controller manages the brake and throttle to control and adjust the vehicle speed based on the lead vehicle speed. ACC capabilities can be further enhanced with the addition of V2V communication, leading to the CACC [22]. This study aims to provide new knowledge about potential safety and environmental improvements from vehicle automation, which will be helpful for researchers, developers, and implementers.

The concept of driving volatility and time-to-collision are utilized to quantify the safety performance and fuel consumption and emissions are calculated to measure the environmental impacts. These measures are discussed in detail in [20].

This study takes advantage of analyzing a unique dataset collected during US DOT's CARMA program [19]. The data represent a proof-of-concept vehicle platooning based on the ACC and CACC applications. The dataset consists of a scenario including the vehicle platooning scenario, as shown in Figure 3.9. The scenario contains the variables of speed, acceleration/deceleration, and the position of the vehicles in a fleet of five vehicles, including a lead vehicle and four following vehicles.

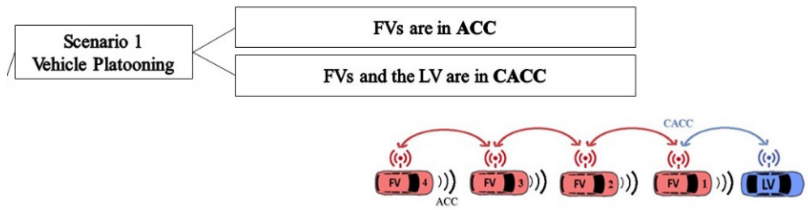


Figure 3.9 Definition and visualization of the platooning scenario

Results show that vehicles equipped with CACC substantially reduce driving volatility as a safety measure in a five-vehicle-platoon from 13.6% to 29% compared with the ACC-equipped vehicles. As one of the main features of the CACC system and V2V communication, it is expected that the automatic synchronization of longitudinal movements of a string of vehicles will reduce the driving volatility. Also, CACC reduces the risk of rear-end collision by increasing the minimum values of TTC near 100% for the vehicles in the fleet. One of the reasons the CACC system can improve vehicle- and system-level fuel efficiency and reduce emissions is to control the magnitudes of the vehicle's acceleration and the variation of acceleration over time. The results show that the CACC technology reduces the overall fuel consumption and emissions in a five-car fleet by 3.7% ranging from 0.5% to 6.7%, compared with ACC. Regarding the environmental impact, the reduction in total emissions ranges from 3.1% to 4.9% and, on average, 3.7% in the first scenario.

The methods used in this study are repeatable; the results reduce the uncertainty of ACC and CACC platooning technologies in terms of their impacts. To further improve the confidence of these results, it will be necessary to collect even more extensive samples of data (e.g., larger vehicle platoons). The state of Tennessee can collect more data in real-world or field tests to expand the scenarios examined in different mixed traffic environments and conditions, e.g., adverse weather, different roadway classifications, and edge cases, to have more generalizable results [20].

Study 3: Pedestrian Crash Prevention Systems

Over the past few years, the number of fatalities and severe injuries of vulnerable road users, particularly pedestrians, has risen substantially. Clearly, the safe mobility of pedestrians is critical in the transportation system. Technology can help reduce vehicle-pedestrian crashes, fatalities, and injuries. Emerging technologies such as pedestrian crash prevention (PCP) systems utilized in on-road vehicles have the potential to mitigate pedestrian crash severity or prevent crashes. However, the reliability and effectiveness of these technologies have remained uncertain. This study contributes toward understanding the effectiveness of PCP systems utilized in on-road vehicles with a low level of automation by investigating two crossing scenarios and one longitudinal scenario. The Insurance Institute for Highway Safety field test data from 2018 to 2021 is harnessed, where several on-road vehicles and their PCP systems are evaluated in terms of safety. The large-scale experimental dataset comprises 3125 tests of 92 vehicles with different sizes, makes, and models, as shown in the study framework in Figure 3.10.

The empirical results indicate that in hazardous pedestrian-vehicle conflict situations, the performance of PCP systems has improved in recent years. The test data shows that some pedestrians were undetected in some tests, but on average, in 70% of the tests, the PCP systems avoided pedestrian crashes. However, for the occurred crashes, PCP systems, on average, were able to mitigate impact speeds of more than 50%. This could translate to substantial reductions in injury and fatality risk in real-life situations. Through rigorous analysis, the associations of critical factors in the studied scenarios and the performance of PCP systems are explored and discussed in this study. The modeling results show that increasing the maximum deceleration rate of the PCP system and lower weight of vehicles can significantly improve the performance of the PCP system by decreasing the speed at impact with pedestrians. The average maximum deceleration utilized in PCP systems has increased from 7.48 m/s^2 in 2018 to 9.36 m/s^2 in 2021. This can be one of the reasons behind the improvement of PCP systems in recent years.

The results reduce the uncertainty of pedestrian crash prevention technologies regarding their impacts. To further improve the confidence of these results, it will be necessary to collect data in real-world or field tests to expand the scenarios examined in different mixed traffic environments and conditions, e.g., adverse weather conditions, low illumination, and edge cases where the system cannot perform perfectly, to have more generalizable results.

A clear understanding of the impact and the situations where the PCP system cannot perform well can help traffic safety practitioners and professionals in Tennessee prepare roadways for a large-scale deployment of these technologies. Identifying the situations where the PCP system cannot perform well and implementing safety countermeasures can maximize the benefit of PCP systems and reduce severe pedestrian crashes [23].

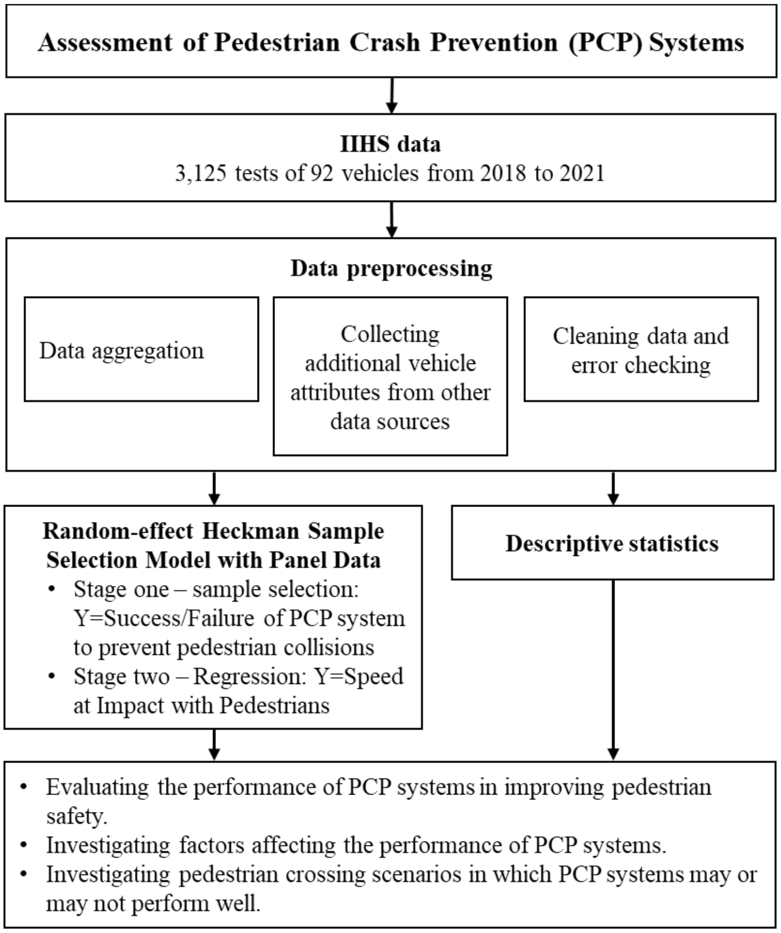


Figure 3.10 Pedestrian Crash Prevention System Evaluation Framework

3.5 Harnessing CAV and non-CAV Data

Study 1: Driving volatility helps identify hazardous intersections

Connected and automated vehicles have enabled researchers to use big data to develop new metrics that can enhance transportation safety. The emergence of such big data coupled with the computational power of modern computers has enabled researchers to obtain a deeper understanding of instantaneous driving behavior by applying the concept of "driving volatility" to quantify variations in driving behavior. This study uses a methodology to quantify variations in vehicular movements utilizing longitudinal and lateral volatilities and proactively studies the impact of instantaneous driving behavior on the type of crashes at intersections. More than 125 million Basic Safety Message data transmitted between more than 2800 connected vehicles were analyzed and integrated with historical crash and road inventory data at 167 intersections in Ann Arbor, Michigan, USA. Figure 3.11 shows the CAV data process steps.

Given that driving volatility represents vehicular movement and control, it is expected that erratic longitudinal/lateral movements increase the risk of a crash. In order to capture variations in vehicle control and movement, the research team quantified and used 30 measures of driving volatility by using speed, longitudinal and lateral acceleration, and yaw rate. Rigorous statistical models, including fixed parameters, random parameters, and geographically weighted Poisson regressions, were developed.

The results revealed that controlling for intersection geometry and traffic exposure and accounting for unobserved factors, variations in longitudinal control of the vehicle (longitudinal volatility) are highly correlated with the frequency of rear-end crashes. Intersections with high variations in the longitudinal movement are prone to have higher rear-end crash rates. Referring to sideswipe and angle crashes, along with speed and longitudinal volatility, lateral volatility is substantially correlated with the frequency of crashes. When it comes to head-on crashes, speed, longitudinal and lateral acceleration volatilities are highly associated with the frequency of crashes. Intersections with high lateral volatility have a higher risk of head-on collisions due to the risk of deviation from the centerline leading to a head-on crash.

Collection of BSM data and utilizing the developed methodology and volatility measures can help TDOT proactively identify hotspot intersections where crashes are low, but the longitudinal/lateral driving volatility is high. The reason that drivers exhibit higher driving volatility levels when passing these intersections can be analyzed to come up with potential countermeasures that could reduce volatility and, consequently, crash risk [24].

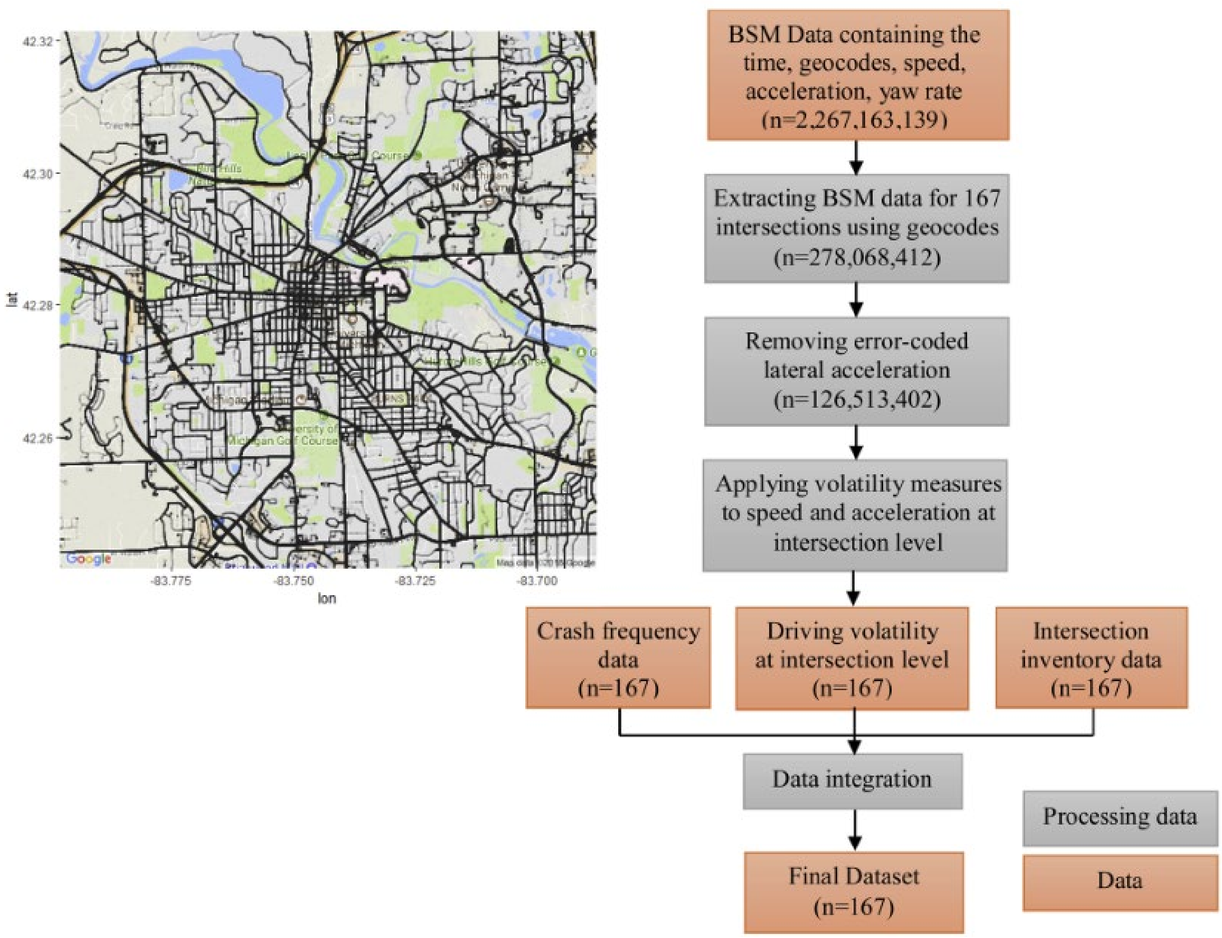


Figure 3.11 Create Map from BSM Data (left), Data Preparation Framework (right)

Study 2: Easing in Automated Vehicles (AVs)-Experimentation in mixed traffic

The introduction of AVs into the transportation network is expected to improve system performance, but AVs' impacts on mixed traffic streams are lightly studied. As AV's market penetration increases, the interactions between conventional vehicles and AVs are inevitable but by no means clear. This study aims to create new knowledge by quantifying the behavioral changes caused when conventional human-driven vehicles follow AVs and investigating the impact of these changes (if any) on safety and the environment.

This study analyzes data obtained from a field experiment by Texas A&M University to evaluate AVs' effects on the behavior of a following human driver. The dataset comprises nine drivers that attempted to follow 5-speed profiles, with two scenarios per profile. In scenario one, a human-driven vehicle follows an AV that implements a human driver speed profile (base). In scenario two, the human-driven vehicle follows an AV that executes an AV speed profile. In order to evaluate the safety, these scenarios are compared using TTC and several other driving volatility measures. Likewise, fuel consumption and emissions are used to investigate environmental impacts. The framework of the study is shown in Figure 3.12.

Overall, the results show that AVs in mixed traffic streams can induce behavioral changes in conventional vehicle drivers, with some beneficial effects on safety and the environment. On average, a driver who follows an AV exhibits lower driving volatility in speed and acceleration, representing more stable traffic flow behavior and lower crash risk. The analysis showed a remarkable improvement in TTC due to the notably better speed adjustments of the following vehicle (i.e., lower differences in speeds between the lead and following vehicles) in the second scenario [25]. Furthermore, human-driven vehicles were found to consume less fuel and produce fewer emissions on average when following an AV.

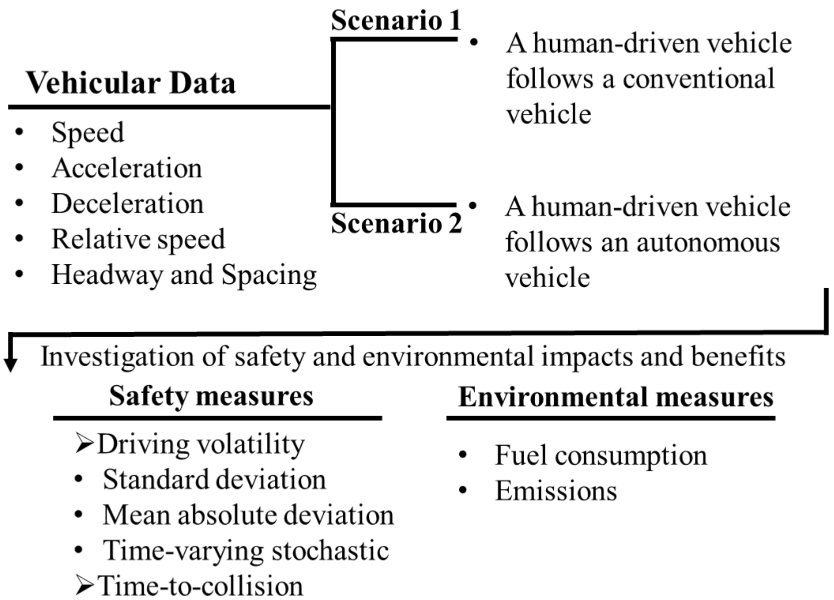


Figure 3.12 Framework of the Pedestrian Crash Prevention Evaluation Study

Since the interactions between conventional vehicles and high-level AVs in the near future are inevitable, TDOT and partner agencies can prepare for it by reducing the uncertainty of impacts of AVs in mixed traffic by testing AVs in different scenarios in mixed traffic [25].

Study 3: Predicting future crashes more accurately with CAV Data

Every year, about 40 percent of the crashes in the US are related to intersections. To deal with such crashes, Safety Performance Functions (SPFs) are vital elements of the predictive methods in the Highway Safety Manual. The predictions of crash frequencies and potential reductions due to countermeasures are based on exposure and geometric variables. However, the role of driving behavior factors, e.g., hard accelerations and decelerations, which can lead to crashes, are not explicitly specified in SPFs. One way to capture driving behavior is to harness connected vehicle data and quantify performance at intersections in driving volatility measures. Studies have found driving volatility to be associated with risk and safety-critical events. Therefore, volatility can serve as a surrogate for driving behavior. This study incorporates driving volatility measures in the development of SPFs for four-leg signalized intersections. The Safety Pilot Model Deployment data containing over 125 million BSMs generated by over 2,800 CVs are harnessed and linked with the crash, traffic, and geometric data belonging to 102 signalized intersections in Ann Arbor, Michigan. The framework of the study is shown in Figure 3.13.

The results show that incorporating driving volatility measures in the intersection SPFs substantially improves the goodness-of-fit and predictive performance of the models. Also, the best results were obtained by applying Bayesian hierarchical Negative Binomial Models in which the spatial correlation between the signalized intersections is considered. The results of this study can have implications for practitioners and transportation agencies.

Collection of BSM data and utilizing the volatility measures in SPFs can help TDOT predict crashes more accurately. This can help TDOT develop potential countermeasures that could reduce volatility and, consequently, crash risk.

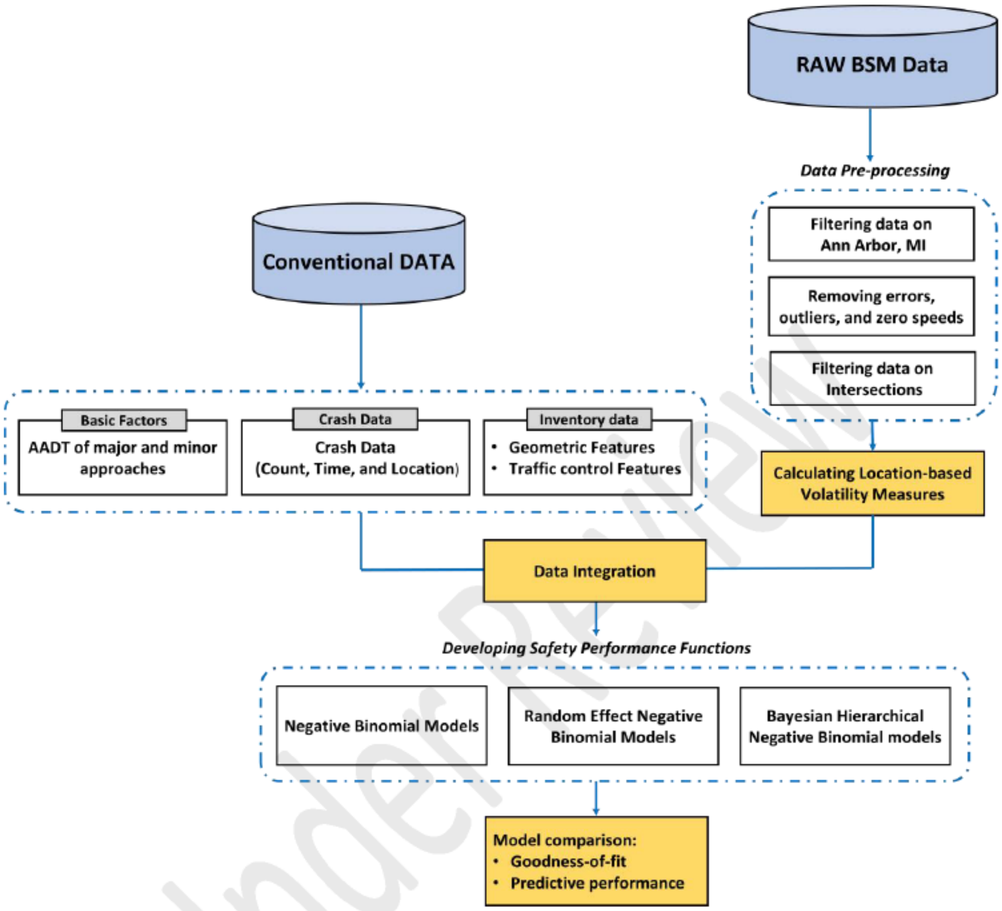


Figure 3.13 Methodological Approach of the Study

Study 4: Identifying hazards through Automated Vehicle disengagements

As automated vehicles diffuse through the system, manufacturers continue to test their automated driving system (ADS) capabilities in complex real-world environments. California's Automated Vehicle Tester Program, run by the Department of Motor Vehicles (DMV), provides valuable data on disengagements in higher-level AVs. This provides the opportunity to develop a more comprehensive understanding of the ADS safety performance through the California DMV disengagement/crash reports. This study comprehensively examines the safety performances (159,840 disengagements, 124 crashes, and 3,669,472 automated vehicle miles traveled by the manufacturers) documented since the inauguration of the testing program. The reported disengagements were categorized as control discrepancy, environmental conditions and other road users, hardware and software discrepancy, perception discrepancy, planning discrepancy, and operator takeover shown in Figure 3.14.

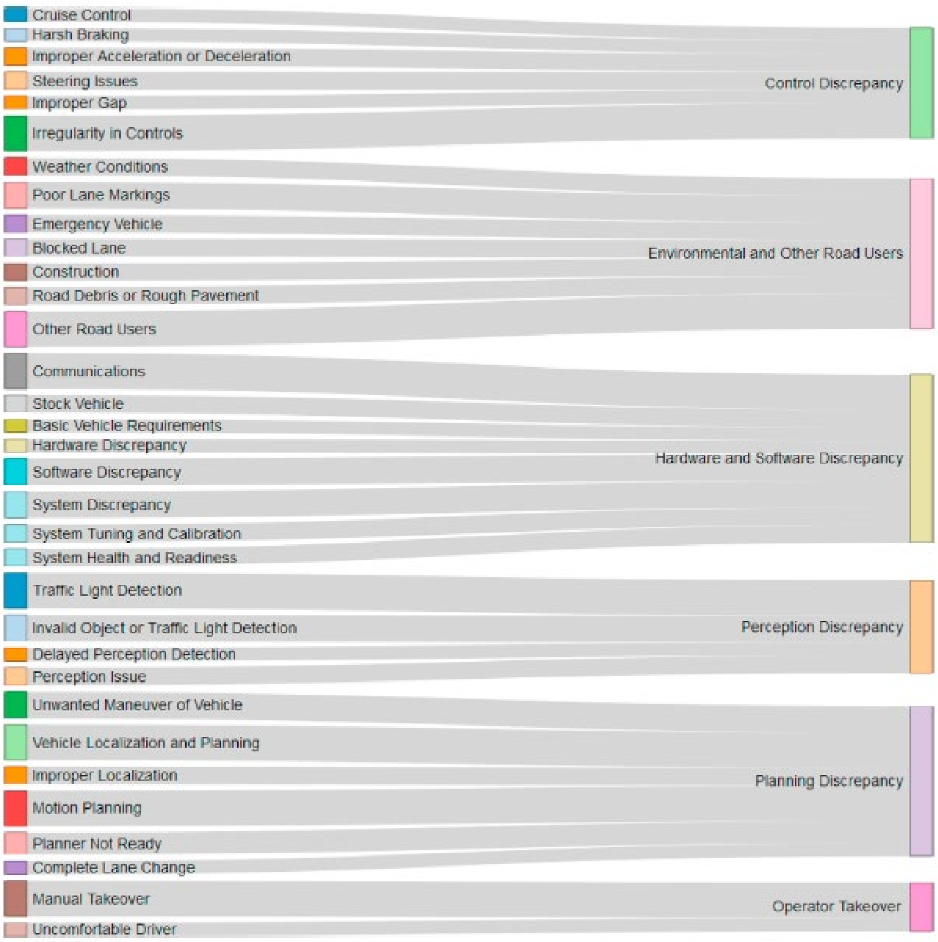


Figure 3.14 Categorization of Disengagements Causes

An applicable subset of disengagements was then used to identify and quantify the 5 W's of these safety-critical events: who (disengagement initiator), when (the maturity of the ADS), where (location of disengagement), and what/why (the facts causing the disengagement). The disengagement initiator, whether the ADS or human operator, is linked with contributing factors, such as the location, disengagement cause, and ADS testing maturity through a random parameter binary logit model that captured unobserved heterogeneity. Results reveal that compared to freeways and interstates, the ADS has a lower likelihood of initiating disengagement on streets and roads compared to the human operator. Likewise, software, hardware, and planning discrepancies are associated with the ADS initiating disengagement. As the ADS testing maturity advances in months, the probability of the disengagement being initiated by the ADS marginally increases compared to human-initiated. Overall, the study contributes by understanding the factors associated with disengagements and exploring their implications for automated systems.

There is uncertainty about how AVs will perform in situations never seen before. As these vehicles currently have relatively small market penetration, there is a low probability of these fringe cases, but as the market penetration grows and more high-level automation vehicles join the fleet, the seldom cases will transform into more frequent occurrences. Analysis of AV disengagement data can provide insights into the performance of AVs in the transportation network in the near future that can help TDOT prepare roadways for a large-scale deployment of AVs [26].

3.6 Smart infrastructure: Traffic Signal Control Strategies Using AI

Many individuals may find themselves waiting in traffic while a traffic light system is servicing a lane with little to no traffic present. While many traffic engineers will attempt to regulate and minimize the average exposure of "green idling" in any traffic network, it is not always guaranteed to work and may not always account for fringe or unusual traffic behaviors. Through machine learning and decentralized negotiations, traffic signals may be optimized to reduce the number of green idling and reduce vehicles' average waiting time.

In general, most traffic signal controllers are programed as either static controllers or actuated controllers. Static controllers are generic traffic signal controllers that follow a fixed sequence of signal phases to manage traffic buildup at an intersection. These controllers are typically applied to dense or urban environments, where traffic flow is uniform and vehicles are being serviced. Actuated controllers are traffic signal controllers that utilize additional sensors to improve the service rates of vehicles by minimizing green idling. Actuated controllers reduce green idling by using sensors, such as induction loops, to monitor the flow within a given range to determine if traffic is continuing or nonexistent. If traffic flow is nonexistent, then an actuated controller may reduce the duration of a signal phase due to a lack of urgency, allowing for another signal phase to be applied that may service more vehicles than the current phase. Actuated controllers may be applied in arterial lanes or other intersections that have sensors integrated and have enough discrepancy in traffic waves that there may be periods where no vehicles arrive and waste time that could be spent servicing other lanes. In general, the rate at which traffic signal timing is calibrated is through a large degree of on-site observations and a degree of engineering judgment. Traffic engineers may count the number of vehicles that arrive at a given intersection at a given period and assess traffic density and build a time sequence that may optimize service rates for a given period.

A prior experiment was performed by applying reinforcement learning, a form of machine learning that uses historical information to train and improve its performance over time. Generally, a reinforcement learning program requires specific values and tables to work properly, such as Q-values, states, actions, rewards, and exploitation ratio. States represent the values that the reinforcement learning program can interpret. An action is a choice that a reinforcement learning program can decide to take and impact the environment. The rewards serve as a matrix to represent the reward or punishment for having a machine learning program reach a specific state by taking a specific at a previous state. The Q-values represent the approximated long-term value of taking action at a given state, based on prior experiences or pre-initialized Q-values. The exploitation ratio is a small chance that an action will not be taken based on which will result in the optimal q-value, but rather select a random action to assess potentially unexplored options that may be more beneficial than the currently utilized state-action pair. The discount factor plays a part in how Q-values may be updated with repeated iterations and the discount factor can determine how many future iterations will influence the current Q-value. The Learning rate also considers how much of the Q-value may be updated by recent rewards and possible future reward opportunities.

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha\left(R + \gamma\left(\max\left(Q(s',a)\right)\right)\right)$$

- Q: Station-action value or utility
- S: Current state
- S’: Resultant State
- A: Action
- α: Learning Rate
- γ: Discount Factor

While it is not ideal for a traffic controller to spend large amounts of time learning that certain combinations are sub-optimal, there are alternative methods to train and update Q-values with synthetic scenarios based on simulations and the likelihood of outcomes. Dyna-Q learning is a form of Q-learning that utilizes hypothetical scenarios to train and update Q-values. In Dyna-Q learning, a series of transition probability equations are utilized to imitate the likelihood of various state-action pairs creating a resultant state. If the transition probabilities closely resemble the actual likelihood, then it can allow for the Q-values to be updated faster without requiring real actions to be wasted by making sub-optimal decisions. In the case of a traffic signal controller being operated by a reinforcement learning program, an intersection making a subpar decision could cause more and more vehicles to backup and cause further issues for the intersection.

For the design of reinforcement learning for a traffic signal controller, the state is represented as either full or empty, with full representing when a queue has built up to the point that an action servicing the lane would not result in any green idling time. While states can be more complex than a binary value, this may be ideal if a state consists of multiple sub-states, causing the number of possible Q-values to increase exponentially. Each intersection's actions reflect the number of possible traffic flow configurations that a conventional traffic signal controller can apply. Rewards are selected to reflect the total accumulated delay experienced by all vehicles not being serviced as a result of an action. The Q-values are generated unilaterally and can be supplemented by Dyna-Q learning. Since the rewards are always negative, the Q values will always be a negative number with an upper limit at zero, and the intersection, if taking the choice to exploit the Q-values, will select the highest value available.

In addition to the behavior of a single intersection, if close enough, multiple intersections can benefit from communicating with each other to understand their environment better and make more optimal behavior. While it is ideal for a single computer to process all possible interactions, the computational power necessary to process each decision for each intersection may be unrealistic. However, an alternative is to make every intersection its own entity that only needs to calculate a portion of the entire environment, spreading the computational load across the entire system. This can be achieved by having each intersection decide the likelihood of their actions through Q-values for the current state. These values are then shared between intersections and provide relevant information for further Dyna-Q transitions by modifying those respective likelihoods, updating the Q-values again, and sharing the updated probabilities between each other again. This process can continue until either a timeout occurs, and the subsequent decision needs to be made by an intersection in the real world, or the changes in likelihood values between intersections have become so low that further negotiations are negligible. The testing environment used to perform the experiment was built using a recreation of a grid traffic network containing 20 traffic signal controllers in the downtown district of Springfield, Illinois (Figure 3.15). No left turns were accounted for to reduce the complexity of the Q-values and training, and all traffic is driving through only.

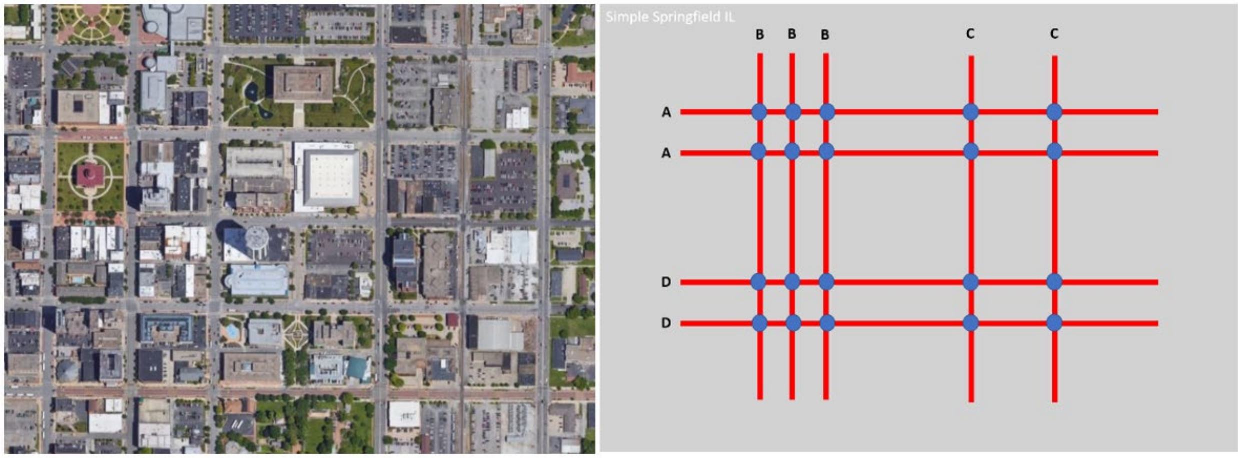


Figure 3.15 Case Study of a grid traffic network containing 20 traffic signal controllers in the downtown district of Springfield, Illinois

Recent results have demonstrated that traffic flow improved, reducing idle time and reducing CO2 emissions generated by vehicles waiting to be serviced (Figure 3.16). Currently, there is ongoing research to develop and expand this concept to be utilized along a major arterial laneway and to incorporate more complex features and traffic behavior. Preliminary results from the Shallowford corridor show some interesting differences between the actuated and RL-controlled traffic signals. Most importantly, note that the delay reduction through westbound traffic combined with the high traffic volume in that direction can potentially lead to better overall performance than the actuated controller. Most importantly, this can offer the groundwork for incorporating traffic signals working in conjunction with CAV platoons to increase performance mutually.

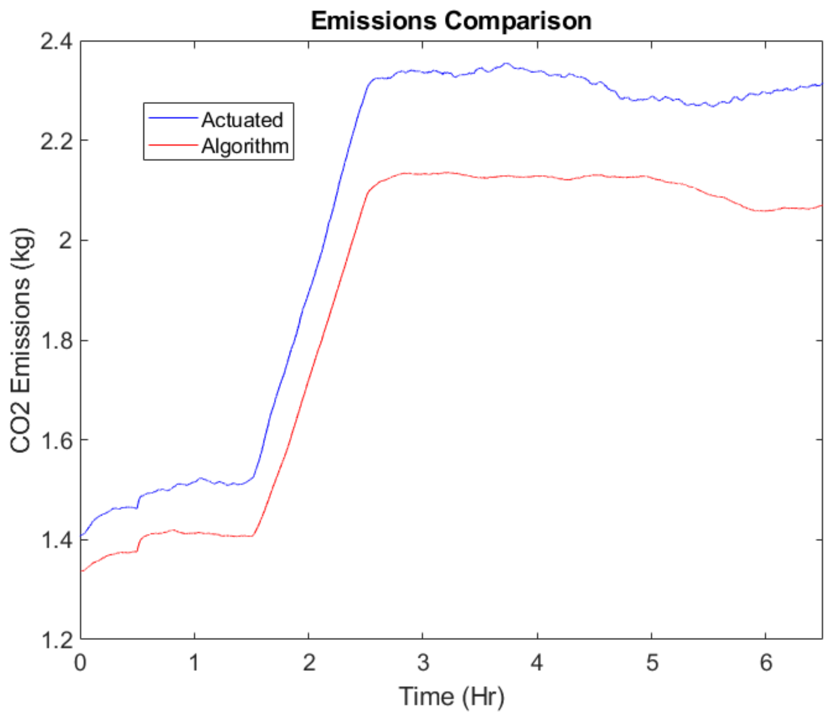


Figure 3.16 Time versus Emissions in the Studied Scenarios

3.7 Synthesizing findings for Tennessee: Lesson Learned

The following is a synthesis of the material presented about data analytics in this chapter.

- TDOT can list and prioritize CAV applications for different parts of Tennessee's network, e.g., highways, local roads, intersections, and different modes and users. Then TDOT can develop a realistic plan for expanding smart infrastructure deployments through CAV applications.
- TDOT and partner agencies can pilot test OBUs in the transit buses (to receive EPCW) and freight trucks (e.g., to receive Safe Pass Advisory) and RSUs at the critical or high-risk hazardous locations to enable V2I and V2V communication that can increase bus and truck mobility and safety.
- Gathering, managing, archiving, and sharing CAV BSM and alert data can be very valuable. Such data can be used to detect hazardous intersections, leading to proactive countermeasures, e.g., locations where the frequency of crashes is low but the longitudinal/lateral driving volatility is high.
- TDOT can implement applications (such as End of Ramp Deceleration Warning and signalized intersection applications) in the future to address safety issues.

Chapter 4 Simulations for Connected and Automated Vehicle Technology (Task 6)

4.1 Introduction

> *Several simulation tools such as SUMU and CARLA are available for envisioning CAV scenarios, sensitivity testing. and identification of edge cases in Tennessee.*

An essential set of tools for CAVs is simulations. Several simulation tools are available for envisioning CAV scenarios, sensitivity testing, and identification of edge cases. Simulations can range from 1) using tools such as SUMO and CARLA for insights about CAV performance at the levels of transportation network or vehicle sensors (LiDAR, radar, and cameras), 2) hardware-in-the-loop simulations, e.g., the Rototest driving simulator for a realistic representation of vehicle (drivetrain) components, 3) multi-user virtual reality simulators for understanding driver behavior at different levels of automation and connectivity, and 4) digital twins to represent a real-time digital counterpart of an operating transportation system. Simulations can provide a system or vehicle-level testing and analysis of vehicle sensors and components. Together, the tools can be viewed as "virtual testbeds" for developing and testing emerging technologies. Moreover, the toolsets can be integrated (e.g., combining SUMO and CARLA) to expand and enhance their capabilities. Generally, simulations are needed as part of the CAV eco-system because they can envision future strategic planning scenarios. Scenarios include mixtures of conventional vehicles and high-level automated vehicles merging at on-ramps and intersections. Also, "edge-cases" can be explored where extreme situations for CAVs can be anticipated and addressed proactively. Case studies of simulations are provided in this report, e.g., studies using SUMO for anticipation of future safety and CARLA to identify edge cases, and the digital twin using a representation of transportation systems in Chattanooga, Tennessee. The highlighted work represents a collaboration between The University of Tennessee and Oak Ridge National Laboratory.

The capabilities of the platforms mentioned are explained further:

- **SUMO software for simulation.** To perform vehicle-level micro-level simulation, the Simulation of Urban Mobility or SUMO platform can be utilized. The simulation can be calibrated with real-world CAV data to study the mobility and safety impacts of CAVs. The team presents a case study that has modeled interactions of different penetration rates of CAVs with conventional vehicles in mixed traffic. A supplemental tool VENTOS can be used with SUMO to model ACC and CACC technologies.
- **CARLA software simulation.** To model sensor-level micro-level behavior of CAVs and their sensors and interactions of AVs with conventional vehicles, especially in a crash or near-crash situations, CARLA software can be used. The simulation can model several AV sensors such as LiDAR, radar, and cameras mounted on a CAV.

4.2 SUMO simulation-How future CAV scenarios are represented

Regarding the experimentation and development of simulations, potential simulation tools include the open-source program-Simulations of Urban MObility (SUMO). SUMO was initially developed at the German Aerospace Center and allowed for the inclusion of pedestrians and specialized traffic. It also allows for utilizing a large set of tools to create specialized scenarios. The program is generated through a series of XML files that account for a specific feature or asset for a single traffic simulation.

The first file to be created is the network document to create the entire traffic roadway network. This file can allow for the roadways to be created with specific positions, lengths, speed limits, lane numbers, and vehicle clearances assigned in the settings. While more complex traffic networks can make writing a document difficult, a graphic user interface generates the same information. Additionally, the SUMO program will include a special program that allows individuals to extract map data from a mapping program, such as Google Maps, to recreate a local traffic environment while also creating a handful of generic vehicle flows as a demonstration of SUMO's capacity.

The next file to be created is the routing file, which generates vehicles, assigns their behavior, and determines what directions they will travel through. While many settings can be set to default values, one value that must be established is to assign where the vehicle enters the network and where the vehicle exits the network. Additional settings can change the definition of the vehicle, which can further change driver behavior or what roadways may be preferred. These vehicle definitions also include interpreting a vehicle as a pedestrian, as SUMO will essentially treat pedestrians as a type of vehicle with an exclusive roadway network mainly in parallel with the established roadway network. SUMO also includes selecting either a singular vehicle or a recurring flow of vehicles. However, any vehicle generated within a flow will all share the same settings, but these settings can include variation and deviation from posted speed limits or average reaction time.

A secondary set of files that can be added are the additional files. Additional files represent a series of unique and niche features and settings that can be added to the network. These additional files include the generation of traffic signal controllers at selected intersections. These include options to control the sequence of signal phases, their duration, and whether the intersection follows a static or actuated controller; if actuated, additional settings allow control of the local sensors, the minimum green time, and the maximum green time for each phase. Alternatively, Additional files can also generate sensors and other devices that can record and report information following the completion of the simulation. These files can include sensors such as induction loops. While these programs can be recorded into a closing document, there are alternative methods for recording information from the SUMO network and can even allow the simulation to be modified while in progress.

SUMO is capable of working with a software function called the Traffic Control Interface (TracI) and allows for information from an active SUMO simulation to be collected and the option to alter the behavior of numerous components of the simulation at any specific time. One popular interface to run TracI is through any Python IDE, but it can also be run through MATLAB by using special programs. The TracI program allows for numerous changes to the traffic environment, including modifying the parameters of a traffic signal controller by changing its current phase or prolonging the current duration, or modifying vehicles to accelerate or change their path-planning route. Several of the ongoing experiments have relied heavily on the application of TracI to use SUMO for prediction and machine learning programs fully.

When developing the experiment for a decentralized traffic signal control network, TracI was used to collect data from pre-designed induction loops located around the traffic signals. These signals could report back the number of vehicles measured within a given time period to MATLAB to generate the state of traffic at the intersection. Using the state information, the action can be enforced from the reinforcement learning program by directly changing the traffic phases and duration of each traffic light to reflect the intended nature of each intersection. Additionally, TracI was also applied to each vehicle to record the waiting time of vehicles and the total amount of emissions generated by vehicles throughout the simulation's duration.

The queue prediction experiment also utilizes SUMO and TracI programming to a large degree. The total duration and traffic flow over a complete 24-hour period were generated into SUMO by using a handful of MATLAB scripts, mainly to automate the entire process. The TracI aspect of the simulation is also relevant for recording the service rate of vehicles by identifying which vehicles that have been previously recorded along a roadway no longer exist, indicating that a vehicle had left the roadway and was serviced. TracI was most notably used to identify and count the flow of traffic in a roadway that had queued; in this case, it was assigned as queued by observing the velocity of each vehicle along a roadway and selecting vehicles as queued when their velocity had decreased to about half of the legal speed limit.

Future works are also planned to continue benefitting from the applications of SUMO and TracI. One future project is to further expand on the current intelligent traffic system experiments to respond and behave when exposed or accounting for vehicle platoons and how it may negotiate with the platoon to split itself to ensure a specific portion that the traffic light will service. At the same time, the second half may begin to decelerate and reduce fuel consumption further during this time. This can be performed as SUMO does include packages for platoons and CAVs.

4.3 Case Studies applying simulation

Study 1: The future of mixed traffic at intersections-SUMO simulations

This study investigates the safety impact of CAVs in mixed traffic with conventional vehicles at intersections. Analyzing real-world AV crashes in California revealed that rear-end crashes at intersections are the dominant crash type. Therefore, to enhance understanding of the future interactions between human-driven vehicles with CAVs at intersections, a simulation framework was developed to model the mixed traffic environment of Automated Vehicles (AV), cooperative AVs, and conventional human-driven vehicles. Adaptive Cruise Control (ACC) and cooperative ACC (CACC) models represent AV's driving behavior. In order to evaluate the predefined scenarios involving multiple degrees of automation, the intersection of Huron Parkway and Washtenaw Avenue in Ann Arbor, Michigan, was selected Figure 4.1.

This study explores system improvements due to automation and connectivity across CAV market penetration scenarios. ACC and CACC car following models are used to mimic the behavior of AVs and cooperative AVs. Real-world connected vehicle data are utilized to modify and tune the acceleration/deceleration regimes of the Wiedemann model. Next, the driving volatility concept capturing variability in vehicle speeds was utilized to calibrate the simulation to represent the safety performance of a real-world environment. Two surrogate safety measures are used to evaluate the safety performance of a representative intersection under different market penetration rates of CAVs shown below.

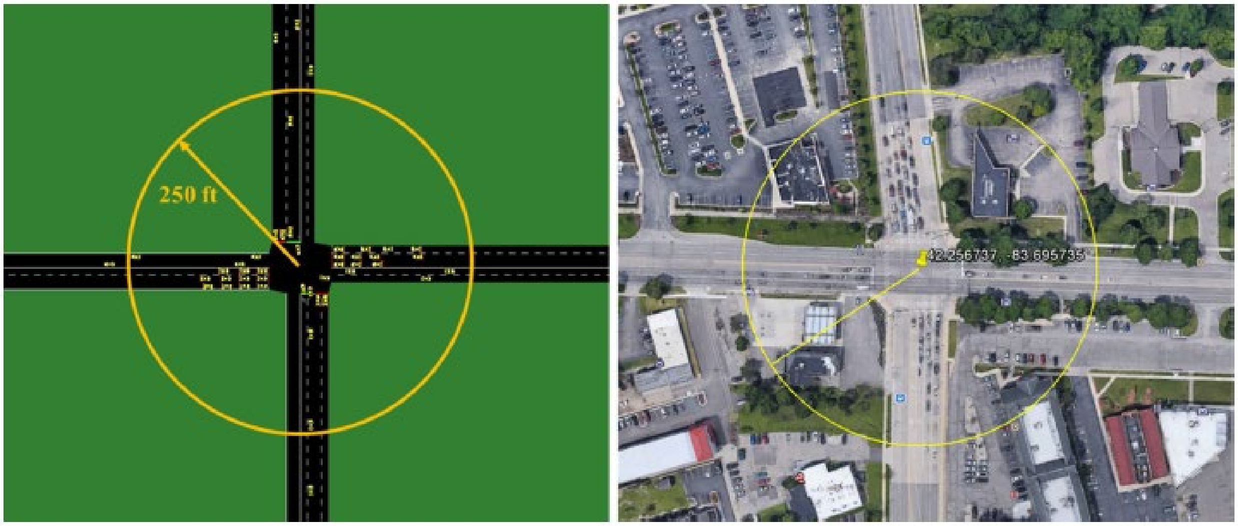


Figure 4.1 Intersection of Washtenaw Ave. and Huron Parkway in Ann Arbor, MI. The study area is shown by yellow boundaries

The number of longitudinal conflicts:

$$\mathrm{TTCi,t} = \frac{(x_{i-1,t} - x_{i,t}) - L_{i-1,t}}{(v_{i,t} - v_{i-1,t})}$$

here $x_{i-1,t}$ and $v_{i-1,t}$ are the positions and speed of the leader, x_i and v_i are the location and speed of the following vehicles, L_{i-1} is the vehicle length, and t refers to the time. A TTC lower than 0.5 seconds was considered a serious conflict in this study.

Driving volatility:

$$Speed\,Volatility = \frac{k > Threshold}{n} * 100$$

where *k* is the number of times that observed speeds that lie beyond the defined threshold, and *n* is the total number of observations. The threshold is defined as

$$\mathrm{Threshold} = \bar{v} \pm 2 * \mathrm{S_{dev}}$$

where $\bar{v}$ is the average speed of the vehicles passing the intersection, and S_{dev} is the standard deviation of the observed speeds.

This study utilized SUMO open-source microsimulation software. The intersection of Huron Parkway and Washtenaw Avenue in Ann Arbor, Michigan, was selected to evaluate the pre-defined scenarios involving multiple degrees of automation. The research team has simulated 600 ft in each direction of the intersection, and intersection territory is defined as 250 ft from the intersection center, which the Highway Safety Manual recommends for the intersection safety analysis. The micro-simulation is calibrated to generate a reasonable approximation of real-world traffic conditions at the selected intersection. The following steps were taken:

- The Wiedemann car-following model was modified to generate more realistic acceleration and speed patterns.
- The simulation is calibrated by incorporating safety measures for conventional vehicles (base scenario).

To explore the safety impact of AVs in mixed traffic with conventional vehicles, this study considered two sets of scenarios. The first set of scenarios focuses on different market penetration rates of AVs with no coordination, following the ACC car-following model. The penetration rate of ACC vehicles ranges from zero to 100% in 10% increments. The second scenario considers the coordination of AVs, assuming that there is a V2V communication between AVs and the vehicles following CACC car-following model. At low levels of ACC market penetration, the safety improvements were marginal, but safety improved substantially with more than 40% ACC penetration. Additional safety improvements can be achieved more quickly through the addition of cooperation and connectivity through CACC. Furthermore, ACC/CACC vehicles improved mobility performance in terms of average speed and travel time at intersections. Figure 4.2 illustrates the number of conflicts observed at the simulated intersection (left figure) and speed volatility for ACC vehicles and CACC (right figure) vehicles in 10% increments of AV market penetration [27].

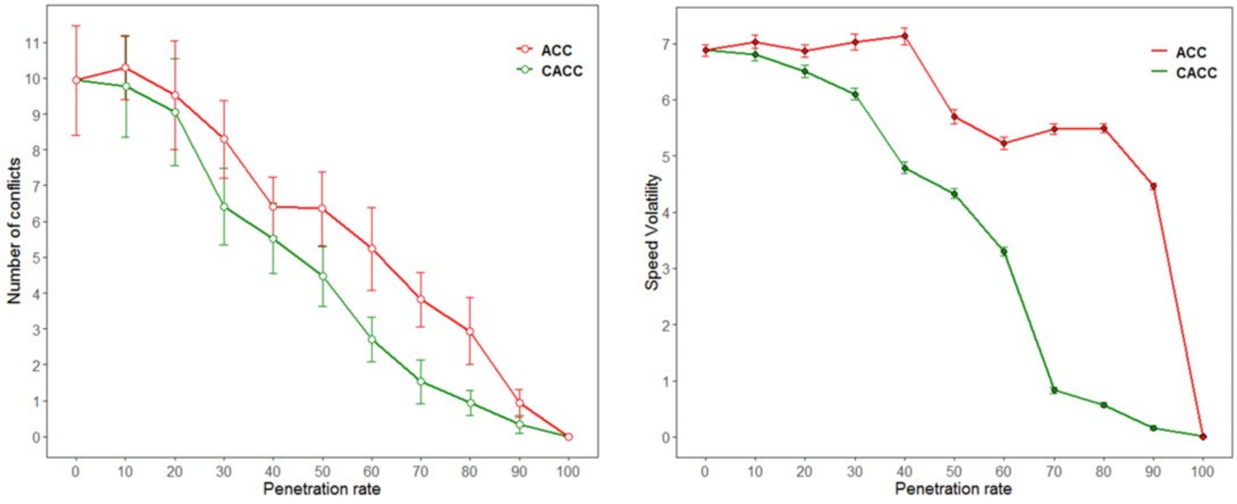


Figure 4.2 Number of conflicts observed at simulated intersection and confidence intervals in the study area, in 10% increments of AV market penetration (Left). Speed volatility at the simulated intersection for ACC and CACC vehicles (Right).

Study 2: Identifying edge cases-CARLA Simulations

As AVs are deployed across the world, it has become critically important to understand how these vehicles interact with each other and other conventional vehicles on the road. To achieve a deeper understanding of the safety implications for AVs, one such method is to analyze instances where AVs were involved in crashes. Unfortunately, this poses a steep challenge to crash-scene investigators. It is virtually impossible to fully understand the factors contributing to an AV-involved crash without considering the vehicle's perception and decision-making. Furthermore, a tremendous amount of data could provide insight into these currently unused crashes, as it also requires a deep understanding of the sensors and data management of the vehicle. The framework of AV data is shown in Figure 4.3.

To alleviate these problems, a data pipeline is proposed (Figure 4.4) that takes raw data from all onboard AV sensors such as LiDAR, radar, cameras, Inertial Measurement Units, and Global Positioning Systems. The data are processed into visual results that can be analyzed by crash scene investigators with no underlying knowledge of the vehicle's perception system. To demonstrate the utility of this pipeline, first, the latest information on AV crashes is analyzed that have occurred in California, and then select two crash scenarios that are analyzed in-depth using high-fidelity synthetic data generated from the automated vehicle simulator CARLA. As a demonstration, the scenarios were implemented in simulation to show that the raw data of the perception system can be interpreted to provide a deeper understanding of how the perception system or control system can fail in fringe case testing. The visualization and data analysis from these scenarios demonstrate the vast improvement in crash investigations obtained from utilizing state-of-the-art sensing and perception systems used on AVs.

As shown in Figure 4-5, a conventional vehicle was following behind a fully automated vehicle in this tested scenario. A pedestrian stepping in front of the automated vehicle caused a rear-end collision with the following conventional vehicle. Similarly, Figure **4.6** demonstrates an automated vehicle making a left turn at a four-way intersection, causing a collision with a vehicle entering the intersection from the right, causing a side-impact collision.

These simulations demonstrate the utility of defining and executing fringe case scenarios inside a simulator like CARLA. Not only are tests like this able to be carried out without endangering people or damaging equipment, but the virtual data allows the complete ability to interpret and diagnose the decision-making of an automated vehicle, demonstrated by the visual results. This study does not use data from BSMs or any V2V or V2X type communications data used for coordination between CAVs and smart infrastructures. Specific crashes involving multiple AVs might benefit from tapping into BSMs to study the pre-crash information exchanges and the corresponding responses from the two AVs. This study did not include this communication modality but was strictly restricted to data from AV sensor suites only. Future research can harness BSM data on vehicle kinematics. The study did analyze the latest information on AV crashes that have occurred in California.

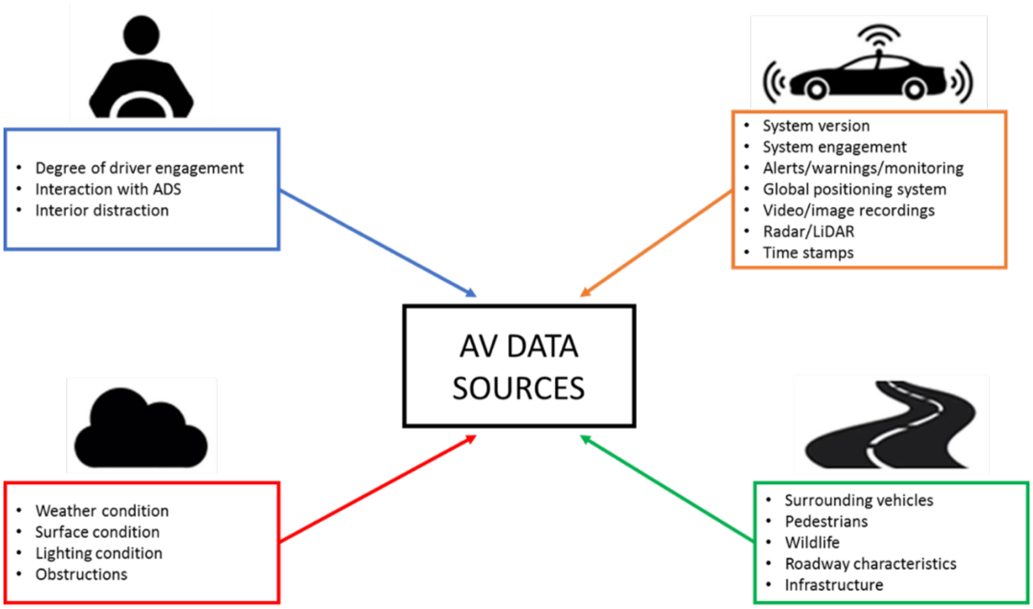


Figure 4.3 Framework to Incorporate AV Data into Crash Investigation Analysis

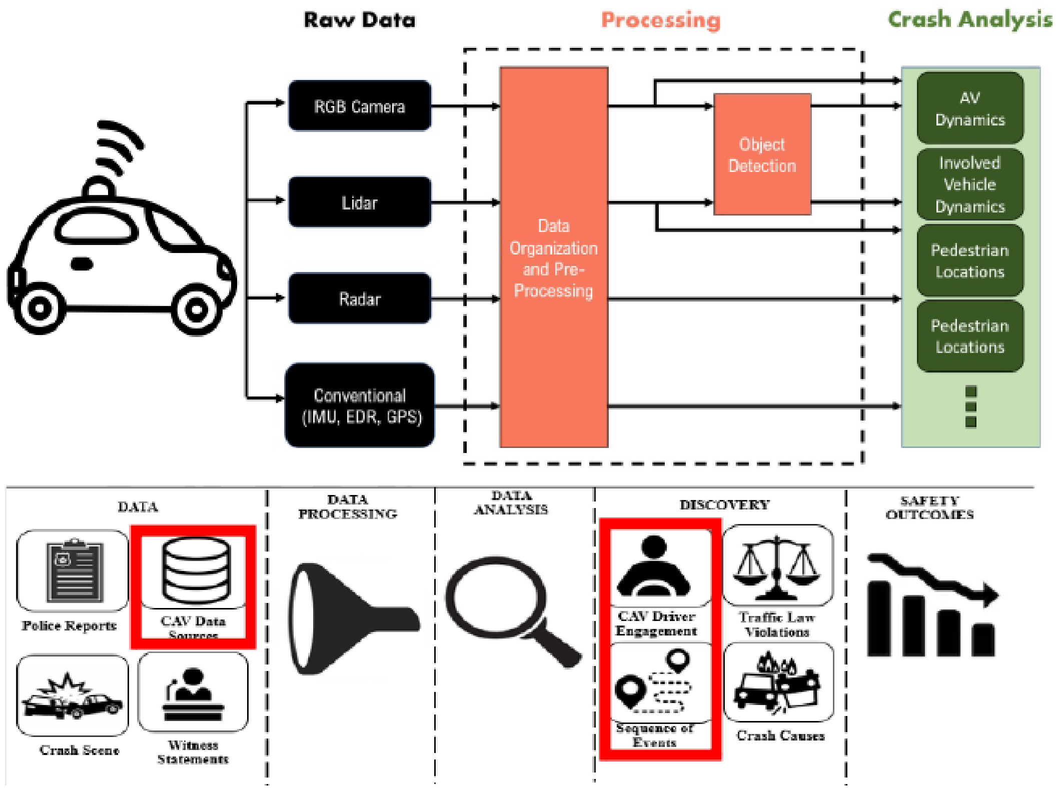


Figure 4.4 The proposed Data Pipeline. Raw Data Collection and Processing Take Place in a "Black Box", Effectively Removing the Need for Underlying Knowledge of Vehicle Sensors to Analyze Driving Behavior and Crash Reconstructions

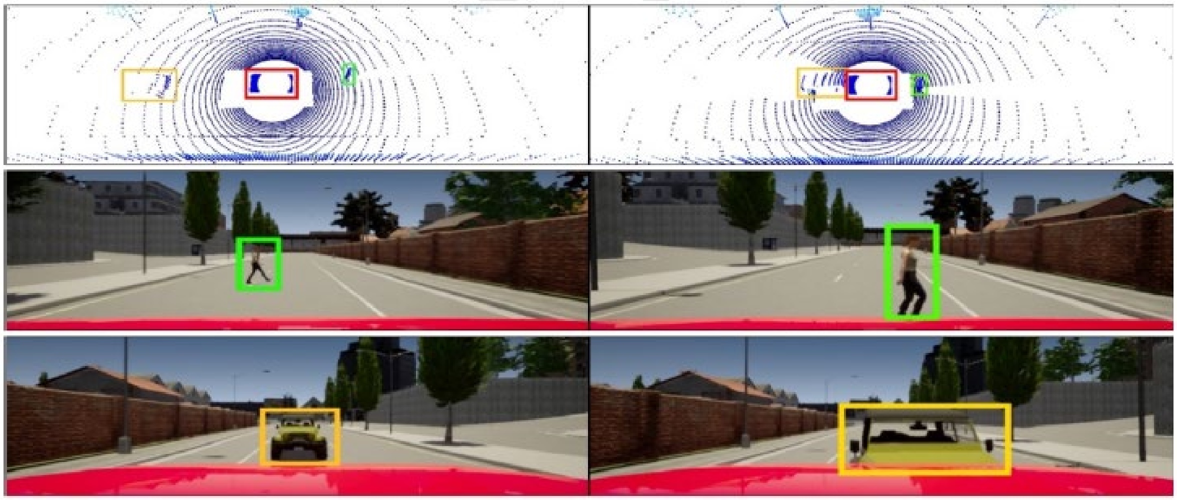


Figure 4.5 Visualization of CARLA sensors in CARLA Simulation #1. LiDAR and Radar (Top), front-facing camera (middle), rear-facing camera (bottom), 1.14 seconds before collision(left), at vehicle impact(right). Pedestrian (green), following vehicle (orange), and AV (red) are color-coded for all sensor modalities.

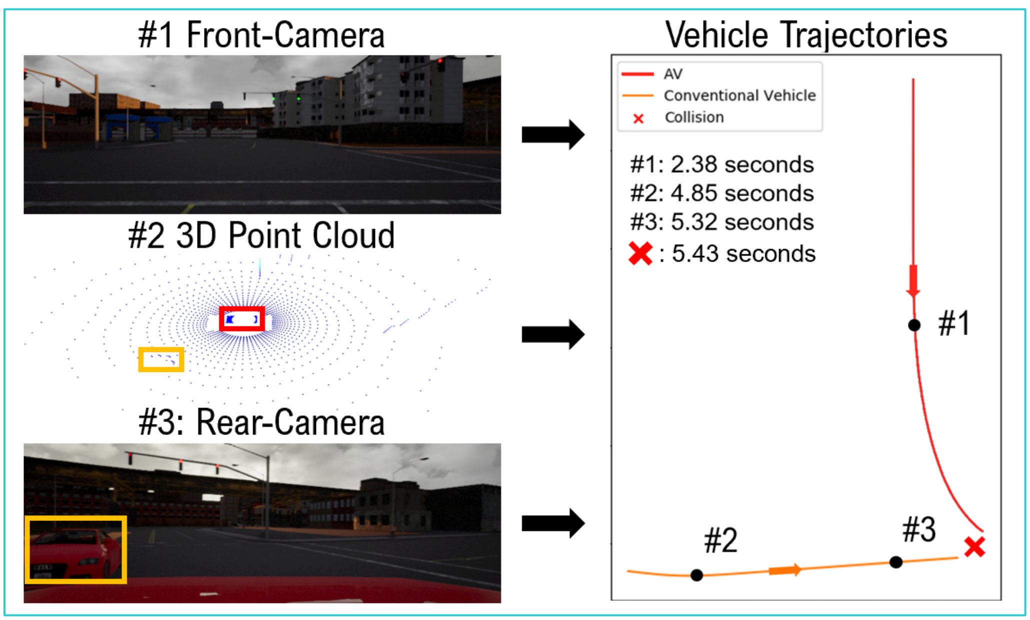


Figure 4.6 Visualization of CARLA Simulation #2. Times shown are in seconds following the start of data recording. Following vehicle (orange) and AV (red) are color-coded for all sensor modalities. Visual data collected (left) can be used to supplement scene dyna

Study 3: Cooperative on-ramp merging simulations of high-level automated vehicles

Vehicle merging is one of the leading causes of reduced traffic efficiency, increased risk of collision, and fuel consumption. Connected and automated vehicles (CAVs) can improve traffic efficiency, increase safety, and reduce the negative environmental impacts through effective communication and control. Therefore, to improve traffic efficiency and reduce fuel consumption in on-ramp scenarios, this study addresses the global and optimal coordination of the CAVs in a merging zone. Herein, a cooperative multi-player game-based optimization framework and an algorithm are presented to coordinate vehicles and achieve minimum values for the global pay-off conditions. Fuel consumption, passenger comfort, and travel time within the merging control zone were used as the pay-off conditions. After analyzing the characteristics of the merging control zone and selecting the appropriate control decision duration, multiplayer games were decomposed into multiple two-player games. An optimal merging strategy was derived from a payoff matrix, and minimum payoffs were predicted for several different potential strategies. The optimal trajectory corresponding to the predicted minimum payoffs was then utilized as the control law to coordinate the vehicles' merging. The proposed control scheme derives an optimal merging sequence and an optimal trajectory for each vehicle. The effectiveness of the proposed model is validated through simulation. The proposed controller is compared with two alternative methods to demonstrate its potential to reduce fuel consumption and travel time and improve passenger comfort and traffic efficiency [28].

4.4 Digital twin-application in Chattanooga

Camera systems (from Gridsmart) are a part of smart infrastructure, and they can provide a wealth of real-time data that can be visualized and analyzed to gain insights into traffic operation at the regional, corridor, and intersection levels. However, cameras can experience a handful of technical limitations as well. One of them is due to adverse weather conditions. Heavy rain, snow, or fog, can obscure the camera's field of view or lead to lower efficiency in the computer vision's performance. Moreover, the field-of-view of these fisheye cameras is also limited, with the efficiency rapidly dropping off near the edges of the visual field (Figure 4.7).

The combined sensing and analysis package offered by companies such as Gridsmart and Miovision offer turnkey solutions that require very little operator knowledge. However, these advantages have to be weighed against the considerable initial investment cost, potentially high maintenance cost, and lack of flexibility in deployment and operation. To some extent, the effectiveness of the camera network can be broadened by using innovative data collection, analysis, and simulation that leverage the camera data as well as the connectivity of intersections in developing models of traffic flow in parts of the road segment which are out of the visual envelope of the Gridsmart cameras. The Chattanooga case study is presented as an exemplar of how smart infrastructure can enhance systems performance, noting that these systems can further benefit from CAVs and their data where available.

This example of data collection and its impact on technologies is an ongoing experiment in which information derived from upstream sensors is used to estimate the size of traffic queues waiting at the downstream intersection. This process is only possible if vehicle data can be collected by sensors and can perform an important role in establishing a more intelligent and responsive machine learning strategy.

The process begins by establishing a rate at which collected data is broadcast between intersections. While it may be ideal that signals could share information every second, it may not be reliable due to the amount of communication that would be necessary and the possibility of data being lost. Depending on the frequency of updates, a roadway can be segmented into a series of update segments to represent the assumption that a vehicle traveling at the speed limit will transfer from one update segment to the next in the time between information updates. Figure 4.8 represents the flow of traffic, wherein vehicles would begin at Segment Gap 3 and would be in Segment Gap 2 when the next recorded traffic count is broadcast between intersections. The yellow Validation Sensor represents the small scope of observation that the Gridsmart cameras can see from their environment and offer a degree of validation and confirmation of flow. The smaller yellow strip represents the traffic counting from upstream sensors, such as another Gridsmart camera. The actual implementation of these "observation zones" can be seen in Figure 4.9, where the upstream and downstream observation areas are marked. Given that traffic may build up in the intervening region unobserved by the cameras over time, a prediction window is created to assess the number of vehicles that are expected to join a traffic queue within a time frame.

Figure 4.7 GridSmart camera field-of-view

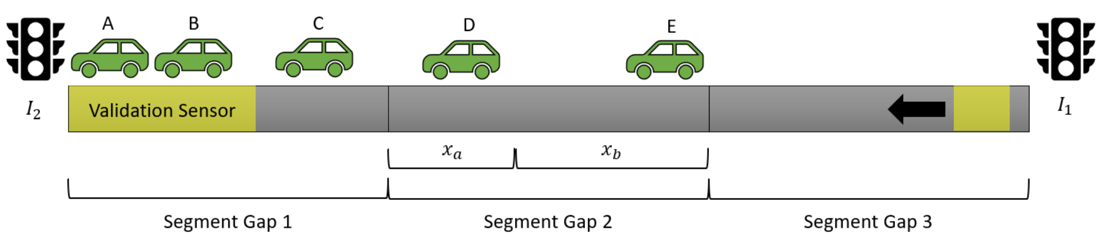


Figure 4.8 A schematic of queue estimation using probabilistic modeling

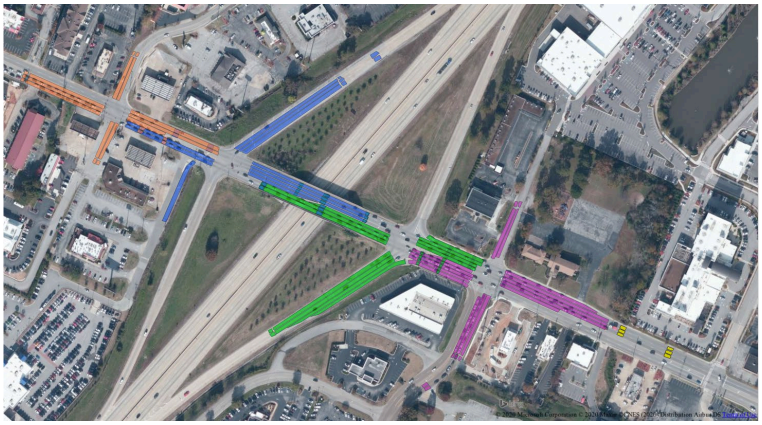


Figure 4.9 A relevant section of the Shallowford road corridor divided into observation zones

As a demonstration, with reference to Figure 4.8, vehicles A and B are already identified as part of a queue, and vehicle C is assumed to be joining the queue, with a probability model being implemented to calculate the likelihood of either vehicle D or E may be joining the queue.

In addition to assessing the flow of vehicles, special considerations will also be necessary to account for nuances in the definition of traffic queues. According to the FHWA, a vehicle is considered queueing when it decelerates to a slower speed or has come to a complete halt. By this definition, a vehicle that may pass through a traffic signal unhindered, mainly by never encountering a red light or other queued traffic, is not considered part of a queue because it never had to decelerate. This would require special adjustments to prior prediction windows, mainly by adding the distance required by a vehicle to brake due to time headway. When vehicles are not driven or operating in a CAV fashion, drivers will typically maintain an intervehicle distance that would allow the driver enough time to respond to a traffic shift, such as immediate braking, while ensuring the upstream driver does so not collide with the downstream driver.

To assess the potential viability of the queue prediction program, a portion of roadway in Chattanooga, TN, was imported into SUMO, a microscopic traffic simulator. This program supplemented vehicle count and signal timing data collected from Gridsmart cameras located along a major arterial lane. A linear roadway was selected for testing as no sensors were available at minor lanes between intersections and would potentially invalidate historical data. This data was collected through collaboration with Oak Ridge National Labs to investigate the potential benefits of intelligent traffic control to reduce congestion on a busy roadway. Figure 4.10 demonstrates what this roadway looks like in the SUMO simulation environment.

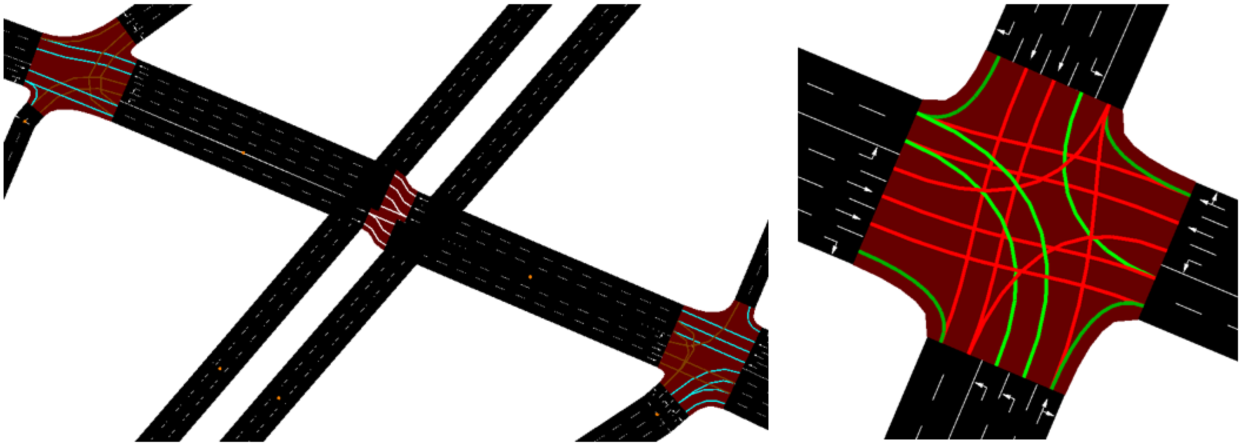


Figure 4.10 SUMO representation of the corresponding Shallowford section with all signal phases and actuated traffic control

To further test the performance of the prediction network, several tests were performed to gauge how well it may perform under various degrees of traffic flow. At this time, it has been tested on the traffic experienced on June 1st, 2021. Three tests (Figure **4.11**) have been performed to assess the accuracy of queue prediction between 6-7 am, 8-7 am, and 12-1 pm to assess how well the program may perform during light, moderate, and heavy traffic, respectively.

The data strongly suggests that the queue prediction algorithm can maintain relatively accurate queue predictions while not mistaking traffic flows during long green windows.

While it will be necessary to test this with multiple days recorded and simulated into SUMO for testing, and further calibrations to the prediction equation may be warranted, the recent results suggest that this program may be beneficial for providing information to a traffic program to estimate traffic flow within a marginal degree of error. Further testing with more complex traffic flows will also eventually be warranted as further traffic signal control experiments increase in complexity.

Demonstrating the queue prediction code for conventional vehicles' traffic flow shows the possibility of establishing intelligent traffic systems' capabilities to respond to heterogeneous traffic flows. This can allow CAV vehicles to perform more effectively in urban environments that may encounter several traffic signal controllers while being less likely to encounter delays that would otherwise reduce the benefits of fuel efficiency and reduce driving time for CAV platoons. While it is ideal for all vehicles to be capable of V2I and CAV functions, it is crucial for programs to be established to benefit CAV systems while accounting for conventional vehicles until enough CAV systems can fully penetrate the market and become as commonplace in automobiles as seatbelts.

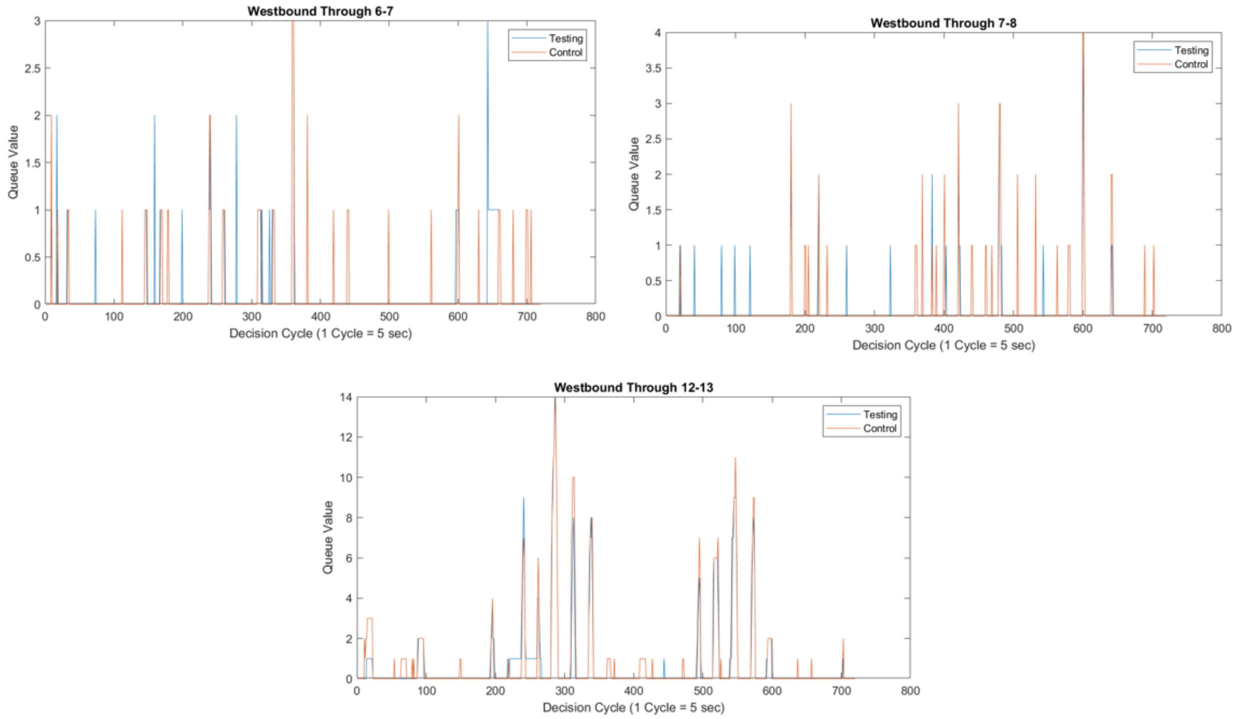


Figure 4.11 Ground-truth queue length and estimated queue lengths show a high level of accuracy in queue prediction using the probabilistic estimation method

4.5 Research on techniques in CAV simulations and AI applications

Currently, many manufacturers are marketing vehicles with the Society of Automotive Engineers (SAE) Level 2 or 3 automation and testing with more advanced vehicles on public roads (level 4). However, recent crashes of Automated Vehicles have raised critical questions about their safety. Are these vehicles with lower or higher automation safe enough to drive on public roads, and more fundamentally, how can their safety envelope be assessed? Currently, there is no consensus about whether testing should exist at the state or federal level, what functions should be tested, how independent testing should occur, and what constitutes safe thresholds.

In this report, specific attention is paid to the newer virtual component of this research. For virtual testing and validation to occur, complex systems must be in place to produce realistic vehicle behavior to the target behavior of physical vehicles on roadways. To that end, the focus was given to producing realistic synthetic data, which includes producing synthetic images and LiDAR point clouds that elicit realistic vehicle responses. Additionally, attention was given to the framework that orients the transition from synthetic driving to real driving and how those might be connected in a meaningful way. Primary points of emphasis in the latest iteration of this research have been:

- Quantification of domain shift error between real and synthetic images in the domain of object detection.
- Simulation of hand-picked fringe cases in a virtual simulation.
- The production of more realistic synthetic LiDAR point clouds using deep learning-based methods.

- The use and development of image-to-image translation techniques make synthetic imagery more realistic and palpable for AI-based driving techniques.

Domain Shift Quantification

Virtual simulation is a crucial ingredient to the exploratory fringe case testing and discovery that defines this research. Furthermore, simulation of vehicles is a long-standing tool that reduces costs and increases safety, more broadly speaking. To that end, the efficacy and realism of simulation are salient. Generally speaking, the realism of a simulated drive can be considered a form of domain shift [29, 30]. All aspects of driving, from perception, control, and the vehicle's dynamic response, should be preserved when driving behavior crosses from the real world to the synthetic domain. Due to the lack of physical vehicle testing, specific attention was paid to the perception systems of these vehicles and how to preserve functionality in the synthetic domain best. The goal is to identify the obstacles between the synthetic domain to the real domain. To this end, the KITTI dataset [31] and the vKITTI2 dataset [32] were primarily used as training data for testing. TABLE 2 describes the features of each dataset.

TABLE 2: SUMMARY OF THE KITTI AND VKITTI2 DATASETS USED

Dataset Feature	*KITTI*	*vKITTI2*
Training Images	7,481	4,200
Classes	Car, cyclist, pedestrian, tram, truck, van	Car, truck, van
Scenes	3	3
Unlabeled (Test) Set	Yes	No

For the synthetic-real domain comparison, the task of object detection was chosen, as it is the most fundamental challenge for the perception systems of camera-based vehicles. Object detection is a field in computer vision that can be considered one additional step beyond image classification. Instead of simply assigning a class to an entire image, object detection aims to localize that object in the image. An example is shown in Figure **4.12**.

Figure 4.12 An example of the YOLOv5 architecture applied to an image from the KITTI dataset. Each object in the image is assigned a location (box), as well as a probability associated with the classification.

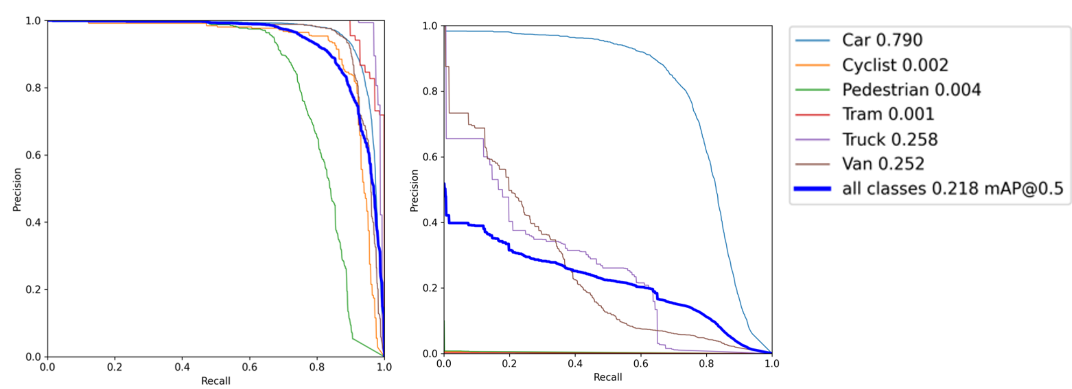
Currently, two primary architectures represent the state-of-the-art in object detection. R-CNN is a mature framework that repeatedly produces crops of the image to find all the objects in the scene. Various iterations have been made on this framework with attempts to produce architectures with more speed or higher accuracy while retaining this core concept [33]. The other primary technique used contrasts with R-CNN. Instead of repeatedly performing classification on different crops of an image, the You Only Look Once (YOLO) architecture only requires one pass of an image to provide detection for every object in the image [34]. Revealing specific details of implementation are beyond the scope of this report, but it is important to note that for every object in the scene, this network produces a four-dimensional bounding box $(X_{center}, Y_{center}, w, h)$ and a confidence score predicting the class of the object in that box, with each of N possible classes in the network $C_{1..N}$ having a value and all class predictions summing to 1. In this research, the open-source version of YOLO referred to as YOLOv5 is the detector used for all data collection [35].

The results of training two object detectors are shown in the precision-recall curves of Figure **4.13**, where a higher area under each curve is associated with higher accuracy in that class. In Figure 3a, an object detector was trained on KITTI training data and tested on the KITTI testing data. In Figure 3b, an object detector was trained on the vKITTI2 dataset and tested on the KITTI testing data. As can be seen, there is a substantial decrease in accuracy in the latter example in all classes with crossover examples. In the car class, the accuracy of the synthetic trained model demonstrated an accuracy decrease of approximately 30%.

Notably, there is a substantial change in the performance of two object detectors trained in two different domains. Considering the importance of the object detection objective in self-driving vehicles, this analysis highlights the need for effort to address this gap. This effort allows the further advancement of fringe case testing and validation techniques.

Edge case testing

As mentioned earlier in this report, edge/fringe case testing can be implemented using CARLA simulation. Such testing can show that the raw data of the perception system can be interpreted to provide a deeper understanding of the perception system or control system. This can, in turn, identify scenarios where these systems can fail. Hence fringe case testing is essential in identifying situations and locations that can be dangerous for CAVs.



(a) (b)

Figure 4.13 The testing results of two Yolov5 object detectors. (a) Trained on KITTI data, tested on KITTI data. (b) Trained on vKITTI2 data, tested on KITTI data.

Synthetic data generation

The problem of domain shift was defined and analyzed. There are a variety of techniques and methodologies for addressing domain shift to ensure simulated validation and testing bear a resemblance to their real-world counterparts. One such method is to modify the data directly. For instance, one could take real-world LiDAR data and model that data directly inside simulation to produce synthetic data. Alternatively, one could take camera images from inside the simulation and modify them to appear as if they came from the real world.

Keeping with the theme of high-fidelity simulation and useful validation, the task of data modeling and data modification was undertaken to explore opportunities for higher quality simulation data and more meaningful images that can potentially be taken from simulation and used for training and improvement of automated vehicles.

Synthetic LiDAR

Many CAV applications rely on LiDAR to produce mappings of the environment outside the vehicle, and the three-dimensional point clouds provided by LiDAR sensors are also superior to image-based sensors in estimating the depth of detected objects [36]. When LiDAR data are unavailable, additional information can be obtained from image-based depth maps in a mimicked form of LiDAR known as Pseudo-Lidar. Both forms are shown in Figure 4.14, and the success of 3D-based methods for detection and localization around LiDAR demonstrates the utility of framing data in the format of 3D point clouds.

In this work, conditional Generative Adversarial Networks (cGAN [37]) deep neural networks were used to replicate the presence and depth of LiDAR point clouds using RGB images as references. If a cGAN can capture the content of an image, that content interpretation can be transferred to simulation and the intricacies of LiDAR can be replicated without the need to replicate the physics of the LIDAR sensor. An example output of this model is shown in Figure 4.15.

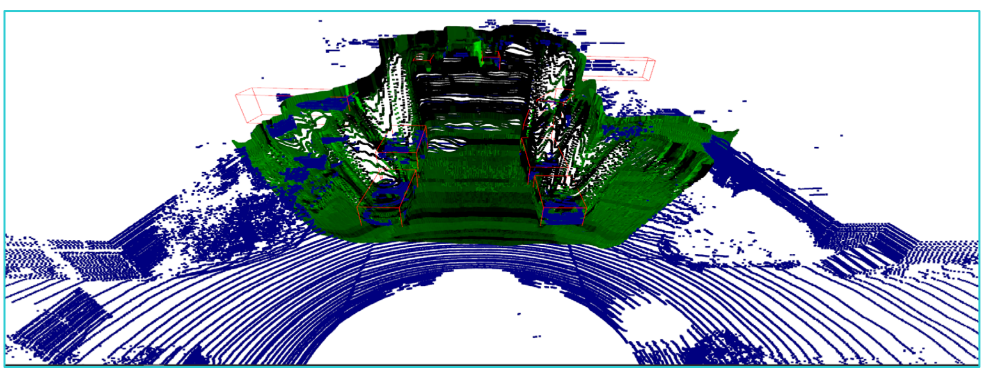


Figure 4.14 A LiDAR point cloud taken from a Velodyne LiDAR is shown in blue, containing 360 degrees of information around the vehicle. A Pseudo-LiDAR point cloud is shown in green, a point cloud where every pixel from a depth estimation is converted to 3D space.

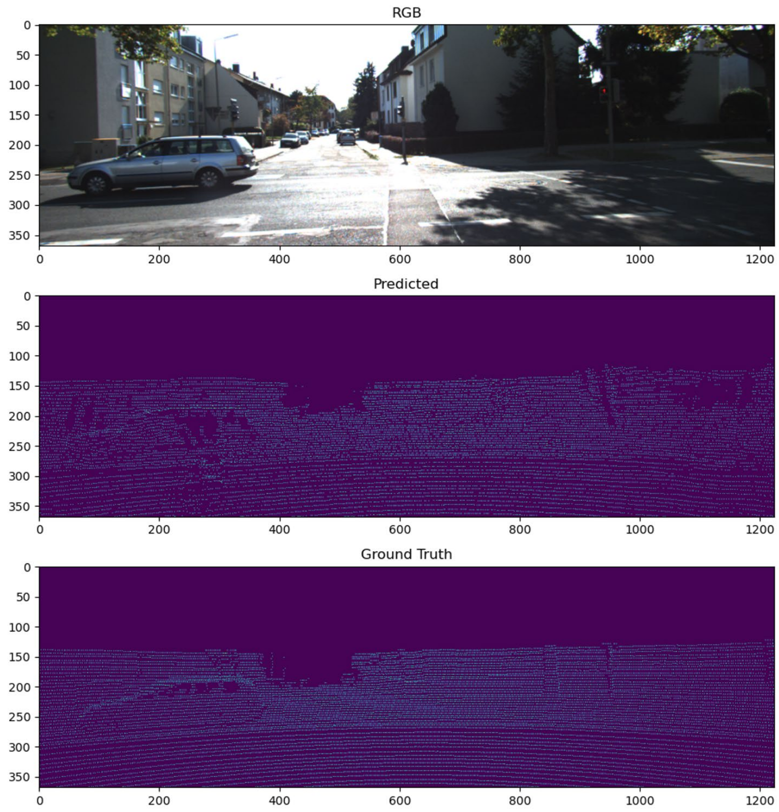


Figure 4.15 cGAN based Pseudo-LIDAR. The RGB image (row 1) is fed to a deep neural network that attempts to generate a realistic version (row 2) of the ground truth LiDAR projection (row 3). The network can learn the physical intricacies of LiDAR and replicate them without the need to reference the ground truth LiDAR point clouds directly.

Image-to-Image translation

Another form of low-level data manipulation that can facilitate the accuracy and realism of the simulation and the data that stems from that simulation is the use of image-to-image translation techniques. Much like the cGAN-based approach, generative adversarial networks are also beneficial at manipulating images on a pixel-by-pixel level, allowing the transformed image to maintain the content that could define a fringe case scenario and be more aligned with the opposite domain. This could be used to make simulated images look more realistic for training object detection algorithms, or it could be used to make real images look more like synthetic images so that systems can exploit known features of the real image like vehicle position in simulation. The CycleGAN framework was primarily used in this research [38], and an example of the synthetic-real transform in both directions can be observed in Figure 4-16.

Figure 4-16 Image-to-image translation methods are shown using the CycleGAN framework. (Top) Real to synthetic. (Bottom) Synthetic to real.

4.6 Synthesizing findings for Tennessee: Lesson Learned

The following is a synthesis of the material presented about CAV simulations in this chapter.

- TDOT can consider investing in research using simulation software (e.g., SUMO) to examine scenarios for the future, e.g., vehicle platoons meant to increase mobility and safety.
- TDOT can invest in a simulation framework to examine the future conditions in mixed traffic environments when CAVs and conventional human-driven vehicles use the same network.
- TDOT can consider developing an optimal merging sequence and optimal trajectory for vehicles entering critical ramps in Tennessee for minimum payoffs in terms of smoother flow and fewer delays, and safer movements in the merging control zone.
- TDOT can invest in testing queue prediction algorithms for different traffic flow scenarios on critical road networks in Tennessee by using simulations.

Chapter 5 Conclusions and Recommendations

Intelligent mobility technologies' research, development, and deployment are essential for improving transportation performance. It is critically important for any state DOT to lay the foundation for the smart technology infrastructure ecosystem. The infrastructure includes supporting 1) roadside and onboard devices for connected vehicles and new sensors, 2) selection of context-relevant applications/user services, 3) installing roadside cameras and dynamic message signs, 4) deploying fiber for fast communication of data, and 5) installing traffic control device improvements. Importantly, establishing the appropriate cyber-physical ecosystem is critical, which also entails the collection, processing, management/storage, and harnessing of CAV (Vehicle-to-Everything-V2X) communications data. For the operation of connected vehicles, such data are continuously being transferred (streamed) between roadside units and onboard units. The research team has worked on supporting TDOT's future efforts in terms of readiness for data collection, data analysis, and the use of simulation for emerging CAV technologies. Focusing on investments in smart infrastructure and intelligent mobility, actions and activities needed for supporting the CAV data collection, data analysis, modeling, and simulation efforts are provided. These are meant to assist in deploying the entire cyber-physical ecosystem for CAV technologies and smart infrastructure.

Focusing on smart infrastructure, the findings of investments in a CAV ecosystem are summarized in three areas:

Collection of CAV data. The whole CAV system is based on the fast movement of data over wireless networks, and hence a critical component of operating CAV systems is data collection. Data transfer in real-time enables 1) the applications and user services that improve traffic operations, 2) archived data helps improve planning and related models for the future, and 3) assists with an independent evaluation of emerging technologies. CAV data refers to the continuous streaming of BSMs, TIMs, SPaT messages, and logs of alerts or warnings, most of which are transmitted over wireless networks. For example, if alerts or warnings are given, then event logs can be created from BSM, TIM, and SPaT messages in a vehicle before and after the alert or warning was issued to the driver. Such data can be stored on ASDs at the time of collection and pushed Over-The-Air from the ASD to the roadside unit (RSU), from where it can be archived on a secure server. Notably, CAV data can be collected, archived, and harnessed in different ways. Details are provided about CAV and non-CAV data sources, data archival, processing, and sharing, with specific use case examples from Tennessee (MLK smart corridor and Shallowford road in Chattanooga) and around the country covering the implications for smart infrastructure technology deployments in the future.

Data analytics and modeling are needed to use the CAV data effectively. This can include visualizing the collected data to measure system performance in real-time and tactical/strategic planning. CAV data are increasingly being shared through dashboards, data hubs, and data lakes. The analytics include visualization of CAV data. Specifically, CAV user services such as red-light running alerts or curve-speed warnings use standardized BSMs, which are data packets related to a vehicle's position, heading, speed, acceleration, state of control, and predicted path. These data can be transmitted from one vehicle to another via V2V and V2I communications, collectively known as V2X communications. In a real-life application, they are analyzed by the receiving OBU to determine the presence of hazardous situations and alert the driver of the host vehicle accordingly. Storing and analyzing these messages can provide insights into whether the alerts were given appropriately and if they were effective in avoiding hazardous situations. Similarly, TIM provides drivers with information about traffic incidents, major events, and even evacuations. These messages typically utilize V2I communications and are sent to vehicles by RSUs. Furthermore, SPaT messages contain data about the state of signal phases at an intersection and related information. SPaT messages are processed by vehicles to support driver/vehicle decision-making at an intersection, e.g., whether to stop or go at a signalized intersection. The point is that these data are analyzed to improve the transportation system's performance, e.g., in terms of safety and mobility, as well as these messages can be analyzed for their effectiveness and harnessed more generally to improve system performance. Modeling the data and applications of AI have gained momentum in this realm.

- Case studies highlight the experience with V2I technologies in the Chattanooga MLK smart corridor, analysis of BSM and alert data from bus drivers with access to "Enhanced Pedestrian Collision Warning Systems," analysis of data on cooperative merging systems at on-ramps, and application of Artificial Intelligence techniques for smart traffic signal control strategies at intersections. New performance measures based on BSM data for safety (e.g., driving volatility and time to collision), energy, and emissions have also emerged.
- Case studies also feature experiences with specific CAV applications such as adaptive cruise control that utilized V2V technologies.
- Case studies further highlight how CAV data can be more generally harnessed for proactive planning without a specific CAV application or user service.

The application of a key set of tools for CAVs is simulations. Several simulation tools are available for envisioning CAV scenarios, sensitivity testing, and identification of edge cases. Simulations can range from 1) using tools such as SUMO and CARLA for insights about CAV performance at the levels of transportation network or vehicle sensors (LiDAR, radar, and cameras), 2) hardware-in-the-loop simulations, e.g., the Rototest driving simulator for a realistic representation of vehicle (drivetrain) components, 3) multi-user virtual reality simulators for understanding driver behavior at different levels of automation and connectivity, and 4) digital twins to represent a real-time digital counterpart of an operating transportation system. Simulations can provide a system or vehicle-level testing and analysis of vehicle sensors and components. Together, the tools can be viewed as "virtual testbeds" for developing and testing emerging technologies. Moreover, the toolsets can be integrated (e.g., combining SUMO and CARLA) to expand and enhance their capabilities. Generally, simulations are needed as part of the CAV ecosystem because they can envision future strategic planning scenarios, e.g., mixtures of conventional vehicles and CAVs, anticipate the operation of high-level automated vehicles' that are merging at on-ramps and intersections, as well as explore "edge-cases" where extreme situations can be anticipated and addressed proactively. Case studies of simulations are provided in this report, e.g., studies using SUMO to anticipate future safety and CARLA to identify edge cases, and the digital twin using a representation of the transportation system in Chattanooga, Tennessee. The highlighted work represents a collaboration between The University of Tennessee and Oak Ridge National Laboratory.

Associated with the selection of context-relevant connected vehicle user services is creating an effective ecosystem. A set of actions include the following:

- ***Invest in collecting CAV data.*** This entails developing a CAV data management system, given the large scale of such streaming data, and identifying the types of CAV data that can support core TDOT functions, including operations, maintenance, planning, and the required workforce for data collection and management. Data collection also comes with investments in cybersecurity, given the potential for adversarial attacks on the large-scale streaming data generated by CAVs. Notably, cybersecurity is a national challenge, and, in this regard, TDOT can follow the guidelines provided by NHTSA. Some of the best practices in cybersecurity in the automotive industry are gathered and discussed in the NHTSA cybersecurity best practices report [1]. Importantly, TDOT should consider developing CAV data sharing procedures within TDOT and a sharing policy with external partners that include other agencies, industry, research institutions, and the general public. Such policies can enhance traffic operations and freight supply chains and support smart city initiatives.
- ***Invest in CAV data analytics and modeling.*** Procedures are needed that fully utilize data from CAVs and other sources to successfully operate CAV user services and understand/improve transportation system performance. Data analytics, modeling, and artificial intelligence techniques are critical in designing highly efficient, safe, and sustainable transportation systems and providing smart mobility services to passengers and freight customers. TDOT should consider creating CAV data dashboards to monitor the performance of the transportation system and the deployed CAV technologies. Specifically, to manage data, TDOT can create and maintain a CAV data dashboard through centralized servers. Such dashboards can provide information that helps oversee operations and inventory and assists stakeholders in tracking resources and activities across the State. TDOT can emulate the connected data platform (CDP), similar to the Georgia DOT use case, to begin integrating diverse data sources. Specifically, CDP can overlay road inventory, WAZE data, CAV device information, highway patrol data, traffic, and crash data in a user-friendly interface.
- ***Innovative uses of CAV data.*** TDOT can use new data sources related to CAVs to support planning activities and assess modeling tools and the methodology they are applied to reflect future uncertainty about CAV adoption. This includes developing transportation models based on CAV data and other data (e.g., crowdsourcing) to accurately estimate and predict transportation system performance and develop proactive and multimodal transportation management plans. Another use of data is providing short-term traffic performance predictions and locating hazardous sites. The data can further be harnessed to improve traffic signal performance by incorporating new performance measures such as driving volatility of the CAV trajectories and using CAV data in high-uncertainty situations such as incidents and special events for lane recommendations and determining dynamic speed limits. TDOT's partner agencies can also use the data, such as Fire and Emergency Medical Services. Further, CAV data can fill data gaps for various functions provided by TDOT, e.g., by maintenance or environmental divisions. All the potential uses will require analysis of the CAV and related data, with some requiring research.
- ***Invest in simulations to create virtual testbeds and digital twins to enhance transportation system performance.*** More investments in "virtual testbeds" through simulation methodologies such as digital twins and the use of software SUMO and CARLA simulations can be valuable for CAV data integration and processing, anticipating future scenarios, doing sensitivity analysis, as well as identifying Tennessee-specific "edge" (fringe) cases.

Additionally, simulations can evaluate operational and planning strategies across large-scale networks. Notably, TDOT can further leverage modeling and simulation capabilities available in Tennessee through the universities and Oak Ridge National Laboratory. This can involve leveraging high-performance computing, data science, and advanced sensors and communications protocols to develop, test and deploy emerging technologies and algorithms for vehicle-to-everything communications (including, of course, the infrastructure and the grid) that enable applications for smart routing, smooth and safe traffic flow, and higher operational efficiency of the network. TDOT investments in applied research should be considered, e.g., using big data and machine learning to improve traffic signals' delay and safety performance in Tennessee or harnessing basic safety message data from CAV.

- ***Future research on data collection, processing, analysis, and dissemination.*** In terms of future CAV research, it is vital to invest in:
    - Developing sophisticated visualizations of CAV data. Specifically, TDOT can invest in creating a data visualization platform that will process real-time data and show different performance metrics. For instance, the visualization may include throughput, arrivals on green, progression ratio, and travel time index on signalized arterials.
    - Using modeling, artificial intelligence, and simulation capabilities based on data generated by CAVs and smart infrastructure enablers to enhance the diffusion of higher automation levels.
    - Accurately estimate and predict transportation system performance and develop proactive and multimodal transportation management systems.

References

1. NHTSA, *Cybersecurity Best Practices for the Safety of Modern Vehicles*. 2020: U.S. Department of Transportation.
2. Ahmed, M., A. Hoque, and A. Khattak, *Intersection Approach Advisory Through V2X Technology Using Signal Phase and Timing (SPaT) Information at Fixed-Time Signalized Intersection*, in *Transportation Research Board annual meeting*. 2018, National Academies: Washington, D.C.
3. Harris, A., J. Stovall, and M. Sartipi. *MLK smart corridor: An urban testbed for smart city applications*. in *2019 IEEE International Conference on Big Data (Big Data)*. 2019. IEEE.
4. Harris, A. and M. Sartipi. *Data integration platform for smart and connected cities*. in *Proceedings of the Fourth Workshop on International Science of Smart City Operations and Platforms Engineering*. 2019.
5. Concas, S., et al., *Connected Vehicle Pilot Deployment Program Performance Measurement and Evaluation–Tampa (THEA) CV Pilot Phase 3 Evaluation Report*. March 5, 2020, U.S. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/55818.
6. WHO, *Global status report on road safety 2013: supporting a decade of action: summary*. 2013, World Health Organization.
7. NHTSA, *National Highway Traffic Safety Administration, 2018 fatal motor vehicle crashes: Overview. Report No. DOT HS 812 826. National Center for Statistics and Analysis, Washington, DC, October 2019.*
8. Litman, T., *Autonomous vehicle implementation predictions*. 2017: Victoria Transport Policy Institute Victoria, Canada.
9. Valentine, D., et al., *Connected Vehicle (CV) Infrastructure–Urban Bus Operational Safety Platform Project Report*. 2019, United States. Federal Transit Administration. Office of Research.
10. Haque, A.M. and A.J. Khattak *Evaluating Pedestrian Safety in Bus-Pedestrian Interactions with the Application of Enhanced Pedestrian Collision Warning Systems: Analysis of Large-Scale Basic Safety Message Data*, in *Transportation Research Board Annual Meeting*. 2021: Washington D.C.
11. USDOT, *Preparing for the Future of Transportation: Automated Vehicles 3.0*. 2018, U.S. Department of Transportation.
12. NHTSA, *National Highway Traffic Safety Administration. Automated driving systems 2.0: A vision for safety*. 2017, Washington, DC: US Department of Transportation. p. 442.
13. Kidando, E., et al. *Traffic Operation and Safety Analysis on an Arterial Highway: Implications for Connected Vehicle Applications*. in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. 2018. IEEE.
14. Mahdavian, A., A. Shojaei, and A. Oloufa. *Assessing the Long-and Mid-Term Effects of Connected and Automated Vehicles on Highways' Traffic Flow and Capacity*. in *International Conference on Sustainable Infrastructure*. 2019.
15. Saffarzadeh, M., et al. *A general formulation for time-to-collision safety indicator*. in *Proceedings of the Institution of Civil Engineers-Transport*. 2013. Thomas Telford Ltd.
16. Kamal, M.A.S., et al., *Model predictive control of vehicles on urban roads for improved fuel economy.* IEEE Transactions on control systems technology, 2012. **21**(3): p. 831-841.
17. Kamal, M.A.S., et al., *Ecological vehicle control on roads with up-down slopes.* IEEE Transactions on Intelligent Transportation Systems, 2011. **12**(3): p. 783-794.
18. Frey, H., et al., *Methodology for developing modal emission rates for EPA's multi-scale motor vehicle & equipment emission system*. 2002, US Environmental Protection Agency, Ann Arbor, Michigan.

19. CARMA. *Cooperative Automation Research Mobility Applications*. [cited 2019 April]; Available from: https://cms7.fhwa.dot.gov/research/research-programs/operations/carma-overview.
20. Mahdinia, I., et al., *Safety, Energy, and Emissions Impacts of Adaptive Cruise Control and Cooperative Adaptive Cruise Control.* Transportation Research Record, 2020: p. 0361198120918572.
21. Hoque, M.A., et al., *The extent of reliability for vehicle-to-vehicle communication in safety critical applications: an experimental study.* Journal of Intelligent Transportation Systems, 2020. **24**(3): p. 264-278.
22. Shladover, S.E., D. Su, and X.-Y. Lu, *Impacts of cooperative adaptive cruise control on freeway traffic flow.* Transportation Research Record, 2012. **2324**(1): p. 63-70.
23. Mahdinia, I., A.J. Khattak, and A.M. Haque, *How Effective are Pedestrian Crash Prevention Systems in Improving Pedestrian Safety? Harnessing Large-Scale Experimental Data*, in *Transportation Research Board 101st Annual Meeting*. 2021: Washington D.C.
24. Arvin, R., M. Kamrani, and A.J. Khattak, *How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data.* Accident Analysis & Prevention, 2019. **127**: p. 118-133.
25. Mahdinia, I., et al., *Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles.* Accident Analysis & Prevention, 2021. **152**: p. 106006.
26. Boggs, A.M., R. Arvin, and A.J. Khattak, *Exploring the who, what, when, where, and why of automated vehicle disengagements.* Accident Analysis & Prevention, 2020. **136**: p. 105406.
27. Arvin, R., et al., *Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections.* Journal of Intelligent Transportation Systems, 2020. **25**(2): p. 170-187.
28. Jing, S., et al., *Cooperative game approach to optimal merging sequence and on-ramp merging control of connected and automated vehicles.* IEEE Transactions on Intelligent Transportation Systems, 2019. **20**(11): p. 4234-4244.
29. Turhan, B., *On the dataset shift problem in software engineering prediction models.* Empirical Software Engineering, 2012. **17**(1): p. 62-74.
30. Nikolenko, S.I., *Synthetic Data for Deep Learning.* arXiv:1909.11512 [cs], Sep. 2019, Accessed: Feb. 05, 2021. [Online]. Available: http://arxiv.org/abs/1909.11512, 2021.
31. Geiger, A., et al., *Vision meets robotics: The Kitti dataset.* The International Journal of Robotics Research, 2013. **32**(11): p. 1231-1237.
32. Cabon, Y., N. Murray, and M. Humenberger, *Virtual Kitti 2.* arXiv preprint arXiv:2001.10773, 2020.
33. Cheng, B., et al. *Revisiting RCNN: On awakening the classification power of faster RCNN*. in *Proceedings of the European conference on computer vision (ECCV)*. 2018.
34. Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, *Yolov4: Optimal speed and accuracy of object detection.* arXiv preprint arXiv:2004.10934, 2020.
35. Ultralytics/yolov5. *Accessed: Dec. 4, 2021. [Online]. Available: https://github.com/ultralytics/yolov5*. 2021.
36. Wang, Y., et al. *Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
37. Isola, P., et al. *Image-to-image translation with conditional adversarial networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

38. Zhu, J.-Y., et al. *Unpaired image-to-image translation using cycle-consistent adversarial networks*. in *Proceedings of the IEEE international conference on computer vision*. 2017.