TTI: 0-7131

Leveraging Artificial Intelligence (AI) Techniques to Detect, Forecast, and Manage Freeway Congestion: Technical Report

Technical Report 0-7131-R1

Cooperative Research Program

TEXAS A&M TRANSPORTATION INSTITUTE
COLLEGE STATION, TEXAS

sponsored by the
Federal Highway Administration and the
Texas Department of Transportation
https://tti.tamu.edu/documents/0-7131-R1.pdf

Technical Report Documentation Page

1. Report No. FHWA/TX-23/0-7131-R1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle LEVERAGING ARTIFICIAL INTELLIGENCE (AI) TECHNIQUES TO DETECT, FORECAST, AND MANAGE FREEWAY CONGESTION: TECHNICAL REPORT		5. Report Date Published: September 2023 6. Performing Organization Code
7. Author(s) Subasish Das, Ioannis Tsapakis, Md Nasim Khan, Jinli Liu, David Mills, Matt Miller, Kevin Balke, Jason Wu, Mehdi Azimi, and Yi Qi		8. Performing Organization Report No. Report 0-7131-R1
9. Performing Organization Name and Address Texas A&M Transportation Institute The Texas A&M University System, College Station, Texas 77843 Texas State University 601 University Drive, San Marcos, Texas 78666 Texas Southern University 3100 Cleburne St, Houston, Texas 77004		10. Work Unit No. (TRAIS) 11. Contract or Grant No. Project 0-7131
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office 125 E. 11th Street Austin, Texas 78701-2483		13. Type of Report and Period Covered Technical Report: September 2021–August 2023 14. Sponsoring Agency Code

15. Supplementary Notes
Project sponsored by the Texas Department of Transportation and the Federal Highway Administration.
Project Title: Leveraging Artificial Intelligence (AI) Techniques to Detect, Forecast, and Manage Freeway Congestion
URL: https://tti.tamu.edu/documents/0-7131-R1.pdf

16. Abstract
Enhancing the quality and efficiency of the Texas surface transportation system necessitates reliable predictions about the initiation and dispersion of prolonged congestion, as well as effective tracking of atypical events and their potential evolution. Artificial intelligence (AI) offers a unique avenue for achieving these goals, presenting an opportunity to accurately estimate congestion measures by utilizing data from various sources, including agency-owned sensors, third-party providers, and extensive enterprise databases. The Texas Department of Transportation 0-7131 project aimed to bridge the current research gap through the implementation of two main project phases. The first phase verified the reliability of commercial data sources for transportation planning and operations. The second phase focused on identifying the most effective AI models or algorithms to meet the Agency's needs based on specific use cases and data availability. Finally, the team developed a prototype decision support tool based on geographical information system technology. The findings provide valuable guidance for decision-makers to prioritize resources, allocate funding, and implement specific measures tailored to the unique characteristics and congestion patterns of different freeway segments.

17. Key Words Rural Roadways, Safety Evaluation, Safety Performance Functions, Decision Support Tool		18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service Alexandria, Virginia http://www.ntis.gov	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 170	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

LEVERAGING ARTIFICIAL INTELLIGENCE (AI) TECHNIQUES TO DETECT, FORECAST, AND MANAGE FREEWAY CONGESTION: TECHNICAL REPORT

by

Subasish Das, Ph.D.
Assistant Professor, Texas State University (TXST)

Ioannis Tsapakis, Ph.D.
Research Scientist, Texas A&M Transportation Institute (TTI)

Md Nasim Khan, Ph.D.
Post-doctoral Associate, TXST

Jinli Liu
Doctoral Research Assistant, TXST

David Mills
Doctoral Research Assistant, TXST

Matt Miller
Research Scientist, TTI

Kevin Balke, Ph.D., PE
Senior Research Engineer, TTI

Jason Wu, Ph.D.
Associate Research Scientist, TTI

Mehdi Azimi, Ph.D., PE
Assistant Professor, Texas Southern University (TSU)

Yi Qi, Ph.D.
Professor, TSU

Report 0-7131-R1
Project 0-7131
Project Title: Leveraging Artificial Intelligence (AI) Techniques to Detect, Forecast, and Manage Freeway Congestion

Sponsored by the
Texas Department of Transportation
and the
Federal Highway Administration

Published: September 2023

TEXAS A&M TRANSPORTATION INSTITUTE
College Station, Texas 77843-3135

DISCLAIMER

This research was sponsored by the Texas Department of Transportation (TxDOT) and the Federal Highway Administration (FHWA). The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of FHWA or TxDOT. This report does not constitute a standard, specification, or regulation.

This report is not intended for construction, bidding, or permit purposes. The principal investigator of the project was Ioannis Tsapakis, and Subasish Das served as the co-principal investigator.

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers’ names appear herein solely because they are considered essential to the object of this report.

ACKNOWLEDGMENTS

This research was sponsored by TxDOT and FHWA. The research team members from Texas A&M Transportation Institute (TTI), Texas State University (TXST), and Texas Southern University (TSU) would like to thank TxDOT staff James Kuhr, Eddie Espinoza, and David Freidenfeld for providing valuable help, information, data, and advice throughout this project. TTI also gratefully acknowledges the support and assistance provided by TxDOT project manager Joanne Steele.

The research team members met with numerous other individuals at TxDOT to gather and/or complement data and information needed for the analysis. The research team would also like to thank Valerie Vierkant (TTI), Juan Cruz Gonzalez (TTI), and Sandip Acharya (TSU) for gathering and processing various data sets that were analyzed in this study.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF ACRONYMS, ABBREVIATIONS, AND TERMS

AADT	Annual Average Daily Traffic
ADS	Automated Driving System
ADT	Average Daily Traffic
AI	Artificial Intelligence
AMS	Analysis, modeling, and simulation
ANN	Artificial Neural Network
API	Application Programming Interface
AV	Autonomous Vehicles
BI	Buffer Index
BTI	Buffer Time Index
Caltrans	California Department of Transportation
CB	Catboost
CBG	Census Block Group
CBR	Cost-Benefit Ratio
CBSA	Core Based Statistical Area
CCTV	Closed-Circuit Television
CMV	Commercial Motor Vehicle
CNAM	Congestion Needs Analysis Model
CRIS	Crash Records Information System
CV	Connected Vehicle
DODT	Department of Transportation and Development
DOR	Department of Revenue
DOT	Department of Transportation
DRIVE-Net	Digital Roadway Interactive Visualization and Evaluation Network
DRL	Deep Reinforcement Learning
DSS	Decision Support System
DTA	Dynamic Traffic Assignment
dvmt	Daily Vehicle Miles Traveled
FHWA	Federal Highway Administration
FID	Feature Identifier
GB	Gradient Boosting

GBDT	Gradient Boosting Decision Tree
GIS	Geographic Information System
GPS	Global Positioning System
HCM	Highway Capacity Manual
HOV	High Occupancy Vehicle
HTTR	Historical Travel Time Ratio
HVR	Historical Volume Ratio
IDPMs	Incident Detection and Prediction Models
IOHMM	Input-Output Hidden Markov Model
IoT	Internet Of Things
ITS	Intelligent Transportation System
KABC/KABCO	Injury Scale for Fatal (K), Incapacitating Injury (A), Non-Incapacitating Injury (B), Possible Injury (C), And No Injury (O) Crashes
KML	Keyhole Markup Language
KNN	K-Nearest Neighbor
LOS	Level of Service
LOTTR	Level of Travel Time Reliability
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAE	Mean Average Error
MAPE	Mean Absolute Percentage Error
MDD	Mobile Device Data
ML	Machine Learning
mph	Miles Per Hour
MSE	Mean Squared Error
MT3I	Maximum Throughput Travel Time Index
MV	Multiple Vehicle
MVDS	Microwave Vehicle Detection System
NCHRP	National Cooperative Highway Research Program
NHS	National Highway System
NLP	Natural Language Processing
NN	Neural Network
NPMRDS	National Performance Management Research Data Set

NPV	Net Present Value
OBU	On-Board Unit
OD	Origin Destination
PDO	Property Damage Only
PDP	Partial Dependence Plot
PeMS	Performance Measurement System
PSL	Posted Speed Limit
PTC	Percent of Travel Under Congestion
PTI	Planning Time Index
PVTT	Percent Variation of Travel Time
RF	Random Forest
RHiNO	Road-Highway Inventory Network Offload
RL	Reinforcement Learning
RMSE	Root Mean Squared Error
RTMC	Real-Time Monitoring and Control
Saas	Software-As-A-Service
SHAP	Shapley Additive Explanations
SIS	Strategic Intermodal System
STD	Standard Deviation
SV	Single Vehicle
SVM	Support Vector Machine
SVR	Support Vector Regression
TC	Transportation Cabinet
TCI	Texas Congestion Index
TD	Transportation Department
TFA	Traffic Flow Analysis
TMAS	Travel Monitoring Analysis System
TMC	Traffic Management Center
TRIMARC	Traffic Response and Incident Management Assisting the River Cities
TSMO	Transportation Systems Management and Operations
TSU	Texas Southern University
TTAve	Average Travel Time
TTI	Texas A&M Transportation Institute

TTRMS	Travel Time Reliability Measurement System
TxDOT	Texas Department of Transportation
TXST	Texas State University
USDOT	United States Department of Transportation
V/C	Volume to capacity ratio.
VHD	Vehicle Hours of Delay
VHT	Vehicle Hours Traveled
VMT	Vehicle Miles Traveled
VOR	Value of Research
VSF	Volume/Service Flow Ratio
VTrans	Vermont Agency of Transportation

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

To improve the quality and effectiveness of the Texas surface transportation system, it is important to be able to predict where and when prolonged congestion will start and how it will spread, as well as to track atypical events and estimate their evolution. Artificial intelligence (AI) approaches provide a unique opportunity to estimate precise congestion measures by utilizing data from agency-owned sensors, third-party providers, and big enterprise data.

In this study, researchers at Texas A&M Transportation Institute (TTI), working with researchers at Texas State University (TXST) and Texas Southern University (TSU), aimed to mitigate the current research gap by conducting two major project phases. These researchers are defined as the *TTI team* in this report. The first phase confirmed the validity of commercial data sources for planning and operations, while the second phase involved understanding which AI models/algorithms are the most suitable for addressing Texas Department of Transportation (TxDOT) needs based on desirable use cases and data availability. The TTI team determined the suitable congestion measures and developed the AI models and associated workflows to determine whether it is sustainable to train, test, and validate the AI techniques. Moreover, the TTI team achieved its research goals by conducting a comprehensive analysis; documenting the commercial big data platforms, datasets, and appropriate AI algorithms; and creating a robust prototype tool to foster return on investment and reduce freeway congestion.

1.2 PROJECT GOAL AND RESEARCH TASKS

The TTI team outlined three goals for this study, which are summarized as follows:

- Determine the available big data platforms and big datasets that are adaptable for existing datasets.
- Reduce freeway congestion on the selected congested freeways using the insights gleaned from AI-based models.
- Minimize the amount of time and resources required to reduce freeway congestion.

In order to achieve the project goals, TTI conducted five major tasks, summarized as follows:

- **Review of Big Data Platforms and Innovative Dataset:** To perform the sub-tasks in this task, the TTI team conducted an extensive review of the existing literature and data platforms/sources, such as commercial big datasets and big data platforms; explored robust AI algorithms to perform big data modeling in the field of congestion reduction; determined crucial factors (including, but not limited to, land use, weather, facility environment, and demographic information) associated with freeway congestion; explored AI and explainable/trustworthy AI frameworks for prototype application; and selected noteworthy practices by other state departments of transportation (DOTs) or agencies.
- **Big Data Validation and Novel Application Identification:** The TTI team explored the available big data platforms (identified in Task 2) to determine the best-fit platforms and datasets for the application of AI in detecting, forecasting, and managing freeway

congestion. Since a need for joining multiple datasets exists, the TTI team examined the suitability of the existing datasets. Then, the TTI team developed a data product (P3) by providing detailed documentation of the validation procedures.

- **Key Performance Metrics and AI-based Big Data Modeling:** In this task, the TTI team formulated key performance metrics to provide TxDOT with better situational awareness of traffic conditions along the corridor to enable more informed decisions on what strategies to deploy to optimize corridor management. Team members developed AI techniques using explainable AI methods to make the results interpretable for the users.
- **Prototype Application Development:** Based on the modeling and exploratory results in Task 4, the TTI team developed an interactive map-based prototype decision support tool (https://txdot.shinyapps.io/0_7131/) that can estimate and visually illustrate the status of freeway congestion. The decision support tool can visually illustrate the congestion profile of the roadway segments, ranked from high congestion to low congestion.
- **Guideline Development:** The TTI team developed a guideline document that synthesizes the key findings and recommendations from the research.

1.3 REPORT ORGANIZATION

The remaining chapters of this report cover the following:

- **Chapter 2: Information Review**—an overview of big data providers, associated innovative datasets, and AI platforms for freeway congestion reduction.
- **Chapter 3: Survey**—the development of a survey questionnaire to learn more about state DOT practices in big data usage in freeway congestion reduction.
- **Chapter 4: Big Data Validation—**review of novel big data applications within traffic management, traffic forecasting and modeling, and real-time reliability monitoring-oriented systems.
- **Chapter 5: Novel Application Identification**—identification of general trends in congestion performance measures, the application of AI strategies and big data sources in traffic demand modeling, and the use of predictive analytics for real-time travel time reliability monitoring of nonrecurring congestion.
- **Chapter 6: Key Performance Metrics**—suggestions of several congestion performance measures and their associated analysis procedures.
- **Chapter 7: Forecasting Using Artificial Intelligence**—forecasting of suitable congestion measures within Texas.
- **Chapter 8: Guidelines and Specifications**—development of a geographic information system (GIS)-based prototype decision support tool that can estimate and visually illustrate the status of freeway congestion.
- **Chapter 9: Conclusions and Recommendations**—summary of the results of this project.

CHAPTER 2: INFORMATION REVIEW

2.1 INTRODUCTION

This chapter provides a synthesis of big data providers, associated innovative datasets, and AI platforms for freeway congestion reduction. To develop the synthesis, the TTI team gathered and reviewed relevant documentation such as websites, state DOT websites, big data platforms, journal articles, research reports, guidebooks, and handbooks.

2.2 BIG DATA PROVIDERS

In the last decade, many transportation agencies have begun using data that are collected, aggregated, and resold by commercial third-party big data providers. These providers generate passive data since the data are created as a result or byproduct of other processes. Some of the major technologies behind the passive data sources are Bluetooth devices searching for receivers to connect with; automotive drivers and other road users such as motorcyclists, bicyclists, and e-scooter riders either using or turning on a global positioning system (GPS); cellular phones connected with network towers; and individual smartphone users with their location service turned on. Individual users reveal their position in space at a defined point in time via any of these technologies, and big data providers trace and aggregate these data points and sell these data to the agencies.

The information available in these data are different than the data from active data collection methodologies and systems used by the DOTs. Those active data collection efforts include manual traffic counts or automatic counters, travel time studies using probe vehicles, household travel surveys, and vehicle intercept surveys. The major issue with active data collection is the coverage and frequency. Active data collection is very costly, and the data collection efforts are very limited. By leveraging passive data, it is possible to supplement or even replace the conventional active data collection efforts.

The TTI team conducted an internet search of available big data providers and AI platforms that were found to be in use or within potential use of public sector transportation agencies and state DOTs. The TTI team used two primary factors in the review of the platforms: (a) whether they provide an AI analytical and predictive modeling component, and (b) whether they have spatial and temporal data visualizations on the platform or a big data set based in real time and applicable within a transportation systems management and operations setting for traffic detection. Table 1 lists 14 vendor services containing predictive AI analytics and/or data visualization components that may be applicable to planning and operations functions within a big data model. Additionally, seven AI platforms were identified that have the potential to support big data automation routines and workflow processes in traffic management congestion.

Most of the 14 services use algorithms to fuse various mixtures of mobile, traffic signal sensor, and connected vehicle (CV) data as documentation for historic traffic congestion patterns and root sources to predict likely geospatial changes in traffic congestion over a given route or segment in the short, medium, and long term. The larger majority offer a dashboard or cloud-based data visualization service to view the transportation network, with about half focusing on municipal clients and the other half working with municipal, regional, and state clients. These algorithms include look-ahead predictive algorithms—in the case of INRIX XD™, reinforcement learning (RL) and machine learning (ML)—and deep learning algorithms, as in the case of DeepDrive, TrafficLink/Miovision, and UrbanLogiq. Workflow output processes from these services include application programming interface (API), internal intelligent transportation system (ITS) infrastructure automation routines, and cloud-based dashboards.

Table 1. Big Data Providers, Products, and Clients.

Platforms	Vendor	Solution	AI Analytics/ Planning	Data visualizations/ Operations	Clients	Link
Surtrac	Rapid Flow Technologies	Optimizing traffic signals; route casting collecting origin destination (OD) and volume from mobile or CV sources; predictive modeling.	AI Analytics uses video, radar, and loop sensor feeds along with RouteCast connections.	Real-time dashboard available with API outputs.	Oriented to municipalities—City of Pittsburgh and state agencies among clients.	https://www.rapidflowtech.com/surtrac
DeepDrive	University of California, Berkeley	Adaptive traffic signal control.	RL algorithm method that uses simulated traffic, not real-world data.	None.	No clients are listed.	https://deepdrive.berkeley.edu/project/adaptive-traffic-signal-control-based-deep-reinforcement-learning
PTV Optima	PTV group	Congestion detection and traffic prediction.	AI analytic forecasting applying data fusion tailored to available sources (INRIX XD, local signal sensors, etc.).	Modular AI data fusion platform for real-time congestion management and response planning.	No U.S. clients	https://www.ptvgroup.com/en/solutions/products/ptv-optima/

Platforms	Vendor	Solution	AI Analytics/ Planning	Data visualizations/ Operations	Clients	Link
TrafficLink and Scout	Miovision	Temporary and permanent roadside unit detection and counts system.	ML algorithm to identify congestion issues based on historic and real-time traffic signal-based multimodal volume data, turning movement counts, bike/ped. counts. No freeway mentions. No modeling/fore-casting.	Real-time visual platform for incident detection and congestion management.	Municipal deployments in City of Austin, TX, Waterloo, ON	https://miovision.com/trafficlink
Waycare Platform	Waycare	Modular predictive AI analytics platform fusing real-time historic traffic data, speed sensor data, and aggregated CV data.	Planning for transportation systems management and operations (TSMO) through forecasting system crashes.	Integrates with Traffic Management Center (TMC) for real-time platform-driven solutions in safety service patrol communications to TMCs and CV data ingestion that reveals queue formations and potential incidents as they occur.	Regional Transportation Commission of Southern Nevada, California DOT, Nevada DOT	https://www.rekor.ai/waycare

Platforms	Vendor	Solution	AI Analytics/ Planning	Data visualizations/ Operations	Clients	Link
StreetLight Insight	StreetLight Data	Fusing mobile-source data through AI and CUBIQ/SDKs-based apps (recently changing).	AI and big data visualization platform of historic mobile phone-based trip counts.	Not applicable.	DOTs: Arizona, California, Florida, Iowa, Kansas, Louisiana, Maine, Maryland, New York, Virginia, Washington, West Virginia, Wisconsin	https://www.streetlightdata.com/how-it-works/
Strava Metro	Strava	Representative sample of bicycle and pedestrian data for modeling, forecasting, and planning purposes.	Downloadable, geo-referenced bicycle and pedestrian forecasting and planning data in a web platform and visualization dashboard.	Not applicable.	Nebraska DOT	https://metro.strava.com/

Platforms	Vendor	Solution	AI Analytics/ Planning	Data visualizations/ Operations	Clients	Link
Waze for Cities	Waze	Event detection and archived data set with speed, volume, and nonrecurring incidents from 2017 on.	Archived data provided through Google Cloud with BigQuery data warehouse tool and data visualization tool (Data Studio).	Data streaming to TMC and 511 sites for traffic incident detection.	Alabama DOT, Georgia DOT, TxDOT, Florida DOT, California Department of Transportation (Caltrans), District of Columbia, Wisconsin DOT, Tennessee DOT, Iowa DOT, Kentucky Transportation Cabinet (TC), Louisiana Department of Transportation and Development (DODT), Maine DOT, Massachusetts DOT, Nebraska DOT, New Hampshire DOT, Oregon DOT, Pennsylvania DOT, Utah DOT, Virginia DOT	https://wazeopedia.waze.com/wiki/USA/Waze_for_Cities

Platforms	Vendor	Solution	AI Analytics/ Planning	Data visualizations/ Operations	Clients	Link
XD	INRIX	Real-time and predictive traffic data service and platform.	Traffic speed data download along with immediate forecast estimations (not capable of accounting for long-term volume trends).	INRIX XD platform provides data in real time to various TSMO functions for detection and response in operations.	Connecticut DOT, Indiana DOT, Iowa DOT, Arizona DOT, Louisiana DOTD, Maryland DOT, Michigan DOT, Nebraska Department of Revenue (DOR), New Jersey DOT, New Mexico DOT, Oklahoma DOT, Pennsylvania DOT, Rhode Island DOT, South Dakota DOT, Utah DOT, Wisconsin DOT	https://inrix.com/wp-content/uploads/2016/09/INRIX-XD-Monitoring-Brochure.pdf
Comma.ai	Comma.ai	Edge-computing AI on-board units (OBU) to convert level 1 automated driving systems (ADS) vehicles to level 2.	No data platform or related analytics products available.	No data platform or related outputs are available on the market.	None	https://comma.ai
Aimsun	Aimsun	Predictive modeling and AI platform.	Predictive modeling.	Real-time traffic management and data visualizations.	North Carolina DOT, Florida DOT, New York City	https://www.aimsun.com/real-time-transportation-management/

Platforms	Vendor	Solution	AI Analytics/ Planning	Data visualizations/ Operations	Clients	Link
UrbanLogiq	UrbanLogiq	Data fusion service with cloud platform.	Application of ML and deep learning AI analytics for forecasts and projections	Cloud platform delivers data visualizations for planning and traffic analysis purposes- not real-time traffic detection.	TxDOT	https://urbanlogiq.com/traffic/
ClearMobility™ Platform	Iteris, Inc.	Cloud-based platform with real-time data fusion engine, and Software-as-a-Service (SaaS) or API delivery methods.	AI analytics.	Traffic operations and real-time detection alongside data visualizations that can be exported to third-party applications.	Montana, Minnesota, Nebraska, North Dakota, South Carolina, West Virginia, Wisconsin	https://www.iteris.com/clearmobilityplatform
Secure Data Commons	United States Department of Transportation (USDOT)	Archive and data analytics web platform with raw and curated data sets from Waze and other partners.	Planning and forecasting analytics.	Not applicable.		https://www.transportation.gov/data/secure/conducting-analysis

2.3 AI PLATFORMS

AI has the potential to solve problems that are hard for traditional methods to address, and various AI methods have achieved state-of-the-art performances in speech recognition, visual object recognition, object detection, and many other domains. Some AI-based methods even surpass human-level performance on some specific problems. Because of increasing population, number of vehicles, and mobility demands, improving the safety, efficiency, and sustainability of the transportation system remains a challenge. Subsequently, traditional methods may not be able to fully address these issues. To overcome these issues, an increasing number of studies have applied AI-based methods to solve complicated transportation problems, including traffic signal control, traffic prediction, microscopic traffic modeling, and autonomous driving. In this section, a brief review of recent studies that utilized AI-based methods is presented (Wang et al., 2019).

Well-designed platforms or systems possess the capability to effectively leverage vast transportation datasets and advanced AI methods. This section provides a succinct review of existing transportation data/AI platforms aimed at mitigating traffic congestion. For the detection and prediction of traffic congestion, PTV Optima stands out by offering the ability to forecast traffic conditions for the upcoming hour (PTV, 2022). Furthermore, the Miovision TrafficLink platform was specifically developed to empower traffic engineers in creating more responsive and efficient traffic networks (Miovision, 2020). In the realm of predictive insights and proactive traffic management optimization, Waycare is at the forefront, shaping the future of urban mobility. It achieves this by enabling cities to gain complete control over their roadways through the utilization of in-vehicle data and municipal traffic information (Waycare, 2022). Meanwhile, Aimsun has established itself as a leader in traffic prediction software and services (Siemens, 2022). With fully integrated software packages, Aimsun complements the Intelligent Transportation Systems (ITS) portfolio by simulating future traffic flows. This supports both offline strategic transportation planning and real-time mobility management. Adding to this array of platforms, the Digital Roadway Interactive Visualization and Evaluation Network (DRIVE-Net) serves as a transportation data storage, management, and visualization platform. This platform facilitates extensive online data sharing, visualization, modeling, and analysis capabilities (Ma et al., 2011). In addition to the aforementioned systems, the URBANLOGIQ platform plays a pivotal role by aggregating diverse datasets, including traffic counts, weather data, infrastructure details, and crash statistics. This empowers cities to gain a comprehensive understanding of movement patterns and congestion dynamics (URBANLOGIQ, 2022).

The seven AI platforms listed in Table 2 help to illustrate some of the hardware and software architecture and system requirements for traffic planning and operations. For example, roadside and signal-based sensor systems in the field have the potential to be updated with edge-computing Internet of Things (IoT) components across a variety of system platforms (e.g., TensorFlow, Intel, NVIDIA) that can upgrade limited real-time traffic data into expansive big data that are capable of generating additional insights and traffic management automation routines. Beyond roadside units, mobile-based data can be incorporated into edge-computing infrastructure-based developments through the use of open, modular AI platform networks such as ONNX that are capable of fusing edge-computing, mobile, and CV data into a big data model supporting both operations and planning functions. In this way, AI platforms driven by a system engineering plan and clear systems needs requirements can take a modular approach and combine both mobile and infrastructure-based components.

Much of the heavier lifting in big data systems development and deployment, based on internet findings, comes from expanding the use of big data from planning functions into congestion detection and traffic operations functions. As a result, many AI platform services out of the seven provide a catalog of potential applications as templates that may work in the realm of roadway and multimodal traffic management congestion along with TSMO automation in traveler information service distribution, traffic incident management, and real-time ITS optimization and fleet management systems.

Table 2. Key AI Platforms.

Platforms	Vendor	Solution	AI Analytics	Data Visualization	Link
C3.ai	C3.ai	Suite of AI applications, development platforms, and standards/specifications. No DOT or public sector transportation agencies listed as clients.	Reliability suite of applications provide asset-based historic, real-time, and predictive analytics.	Data visualization for current asset conditions and automated incident response enabler applications.	https://c3.ai/products/c3-ai-reliability/
NVIDIA Metropolis	NVIDIA	Application catalog, development framework, and AI platform for smart cities. Focuses on computer vision but may have congestion analysis applications within catalog.	Applications for predictive analytics across edge, mobile, and infrastructure.	Data visualization applications for traffic operations.	https://resources.nvidia.com/en-us-metropolis-smart-cities/gtcfall20-a21335
Wejo Neural Edge	Wejo	Third-party data aggregator for connected vehicle (CV) and autonomous vehicle (AV).	Predictive analytics.	Data visualizations.	https://www.wejo.com/press/wejo-announces-wejo-neural-edge-processing-platform-streamlining-connected-vehicle-data-and-driving-autonomous-vehicle-reality-forward
Waymo Open Dataset	Waymo	Limited (six non-Texas cities) processed but disparate open data set of high-resolution AV sensor data (lidar, camera, radar).	Planning analytics may be derived, but it is not prepared for use.	Data visualizations yes, but not for use in operations and detection.	https://waymo.com/open/data/motion/
Open	Open AI	Catalog of TensorFlow deep reinforcement learning (DRL) algorithm implementations.	Predictive analytics possible from DRL implementations.	Data visualizations and traffic detection automation potential from DRL implementations.	https://www.sciencedirect.com/science/article/pii/S2590198221001317

Platforms	Vendor	Solution	AI Analytics	Data Visualization	Link
Yunex Traffic Digital Labs	Yunex Traffic/Siemens Mindsphere	Mindsphere IOT PaaS AI platform using data fusion of infrastructure, fleet, and third-party traffic data for traffic management.	AI analytics.	Data visualization, API-driven operations response.	https://www.yunextraffic.com/global/en/portfolio/ai-and-digital-solutions/ai-based-traffic-and-city-mobility-solutions
ONNX.ai and Runtime	ONNX (developed by Facebook and Microsoft)	Open neural network (NN) AI model development and optimization standard with file conversion mechanism across file types from the plethora of proprietary edge systems, mobile systems, and computer vision systems.	AI architecture/groundwork for interoperable AI analytics and planning/forecasting applications relying on edge computing as well as data fusion needs.	AI architecture/groundwork for automation routines with data visualization in traffic operations and detection	https://onnx.ai/about.html

2.4 BIG DATA SOURCES

In particular, the review focused on studies conducted with Bluetooth receivers, GPS devices, and mobile device data (MDD) obtained through cellular networks and location-based services embedded in smartphone applications. The review considered the specific applications of the data products seen within the literature and the sample size reported or inferred within the reviewed studies and summarized the success or failure of the attempted applications. The review also contains a summary of the underlying technologies, including inferred and known sources of bias or incompleteness (see Table 3).

The collective efforts of the many researchers compiled in this document show that each technology has its own particular strengths and weaknesses. In general, Bluetooth excels in projects of a temporary nature, such as construction corridor planning, as well as targeted cordon studies that focus on the entry and exit of individuals within a selected perimeter. Bluetooth can be permanently installed and is often used in real-time applications, such as travel time prediction on highways. Bluetooth falls short in its ability to obtain the true origin and destination of individual trips. GPS technology is highly accurate and precise and therefore excels in studies that require individual route traces, such as network construction and travel time studies. Due to the lower penetration of GPS devices across an unbiased population, the technology may be unable to identify wider behavioral information. By contrast, the widespread proliferation of MDD promises a wider and less biased view of population travel patterns, although the aggregation of data from multiple sources and the coarse temporal resolution of these data limits the precision of studies conducted at small scales.

2.4.1 Bluetooth

Bluetooth is a short-range, radio-based communication protocol that permits pairs of authenticated devices to send out limited amounts of data to each other, with one device referred to as the transmitter and the other referred to as the receiver. Applications of this technology include wireless headphones that connect with mobile telephones or mobile telephones that connect with in-car audio and information systems. Bluetooth technology originated in the late 1990s, and it has been under continuous development ever since. Bluetooth 1.0 was released in 1999 with a maximum range of 33 ft. Currently, Bluetooth 5.0+ has an extreme maximum transmission range of around 780 ft based on the signal processing technology used, and individual Bluetooth devices may or may not have the power necessary to transmit over the maximum range (Macfarlane and Copley, 2020).

When enabled, Bluetooth devices are continually emitting and searching for Bluetooth signals. Bluetooth receiver units are set up by the DOT or its contractors, and transportation analysts have exploited this feature of Bluetooth technology to conduct many transportation studies.

2.4.2 Global Positioning System

GPS was originally developed as military technology used to find precise targets and aid military vehicles in navigation. Today, with the right combination of weather, atmospheric conditions, and receiver technology, it is possible for the public to have precise location information. By triangulating the signals reflected from multiple GPS satellites, GPS-enabled devices can accurately locate their positions anywhere on Earth independent of local infrastructure or data connectivity. This has made GPS an attractive technology for in-car position and navigation systems in consumer and commercial vehicles and for managing fleet operations (Macfarlane and Copley, 2020).

Commercial providers of GPS services such as TomTom, HERE, and INRIX can acquire the locations and timestamps of devices on their networks by tracking GPS equipment placed in vehicles and, in some cases, data collected through MDD. In some instances, cellular phone navigation apps use location-based services or other mobile device-locating features in addition to GPS information.

2.4.3 Mobile Device Data

In recent years, most people in the United States carry a smart phone device as a matter of habit, utility, and convenience. Some other smart wearable devices include smart watches, fitness trackers, and music players, which are usually connected to data networks to provide services to their users. For real-time services, these devices require location services to be turned on so that the users can get real-time updates about location, weather, and nearby interests. Even before the proliferation of these smart devices and location services, cellular phones sent data to and from towers that are in space. Over the last few years, many private big data vendors have improved techniques to collect, aggregate, and sell data from cellular towers and location-based services.

Table 3. Key Data Sources.

Data Source	Advantages	Disadvantages	Big Data Companies
Bluetooth	• Short-duration event data • Real-time estimates • Accurate relative to receivers	• Gaining traces through a cordon is difficult • Require hardware setup	• Blyncsy • BlueTOAD (Iteris)
GPS	• High precision • True OD data • Continuous ping rates • Average travel time (TTAve) and speed data • Separate travel time and speed data for passenger cars and trucks	• Low penetration rates • May cause 16 ft error in clear conditions • Fleet vehicle origins are often biased	• HERE Technologies • INRIX • TOM TOM • Wejo • OtoNomo • Waze
MDD	• True O-D data • Large sample sizes • More opportunity for tracking active modes and mode differentiation	• Lack of intermediate trace data. • Lower accuracy	• AirSage • Citilabs • Replica—Sidewalk Labs • StreetLight Data

2.5 REVIEW OF STATE DOT PRACTICES

Congestion is a major and challenging issue for urban transportation systems, and reducing congestion has become complex in regard to traffic analysis and the emerging growth of transportation data and technologies. In recent years, there has been growing interest among highway transportation agencies and state DOTs to supplement traditional traffic data collection with new data collection practices and to use those data to analyze, forecast, and reduce traffic congestion. The majority of state DOTs have started adopting data initiatives since the inception of the USDOT Safety Data Initiative and are launching pilot projects. In addition, a few states started creating their own databases to store and analyze transportation data, whereas other states utilized the products of big data vendors and service companies such as StreetLight, INRIX, Iteris, Waycare, and Wejo to create, store, and use traffic data.

A few states (for example, Massachusetts) use geographic information system (GIS)-based platforms to monitor traffic data in the form of Keyhole Markup Language (KML) layers and shapefiles, as well as Excel-based (CSV) files. Those public data include average daily traffic (ADT) and annual average daily traffic (AADT), which are based on traffic count stations. Other states, such as Indiana, maintain their own dataset and provide updates on the traffic data on a daily basis. Those data include ADT, AADT, speed, and volume, along with analytical graphs such as the traffic volume index and speed graphs. In some cases, they may include data related to CVs. The data obtained from those public databases, as well as private data vendors, can be in the format of raw data (which requires processing) or refined data.

To analyze traffic data for sustainable mobility and solve transportation-related problems, some states have—in addition to maintaining their own dataset—database platforms, whereas a few states have started initiatives to utilize external private platforms for big data in the refined format. Private companies such as StreetLight, INRIX, Iteris, Waycare, and Wejo collect and analyze raw data and provide the states with refined data (which are more convenient than raw data) for use by the agencies. These data are not available for free to the public. It should be noted that although private companies emphasize the reliability and high quality of the data they provide, it is recommended to validate the data using multiple techniques. Many states use both their own databases and the databases by vendors, which helps them to have a wide range of mobility and traffic data from pedestrian data, bicycle volume data, vehicular data, and weather data.

Figure 1 shows the state market distribution of big databases and platforms. The information was collected by a thorough review of the literature, searching the news for stories relevant to the topic of the research project, visiting the state highway agencies, and disseminating from the private vendor webpages. From the figure, it can be seen that INRIX, StreetLight, and Iteris are widely used by different states for their transportation and mobility data.

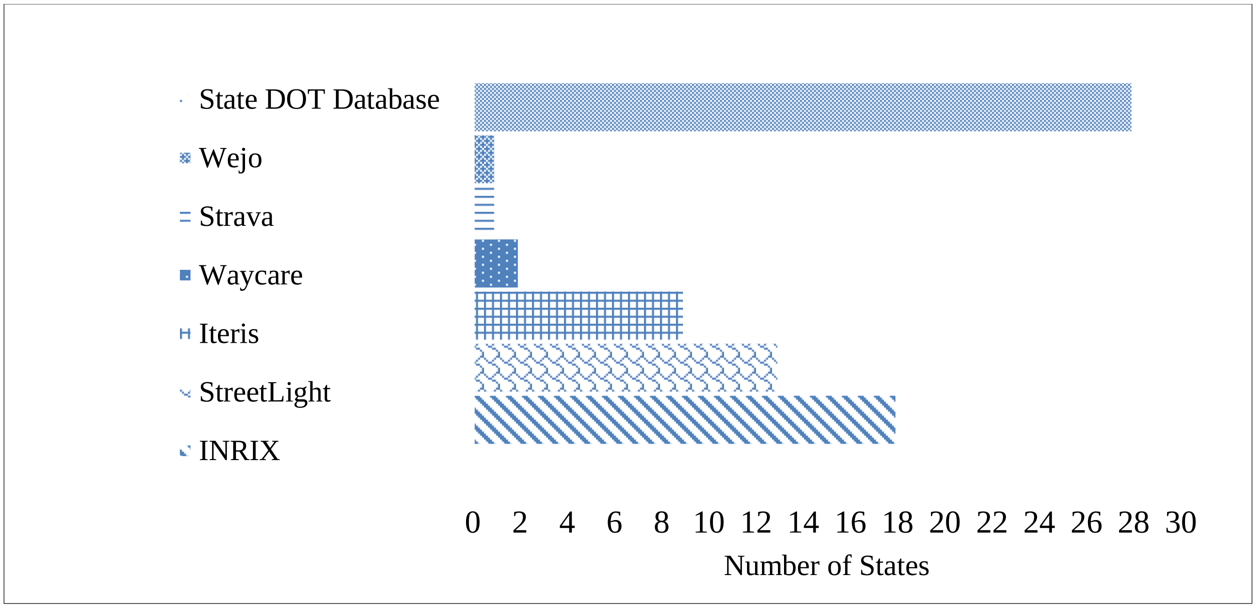


Figure 1. State Market Distribution of Big Databases and Platforms.

There are some states that use more than one database or platform. For example, Caltrans not only has its own database but also utilizes the data from the private vendors StreetLight and Waycare. Figure 2 shows the states that have more than one database or platform.

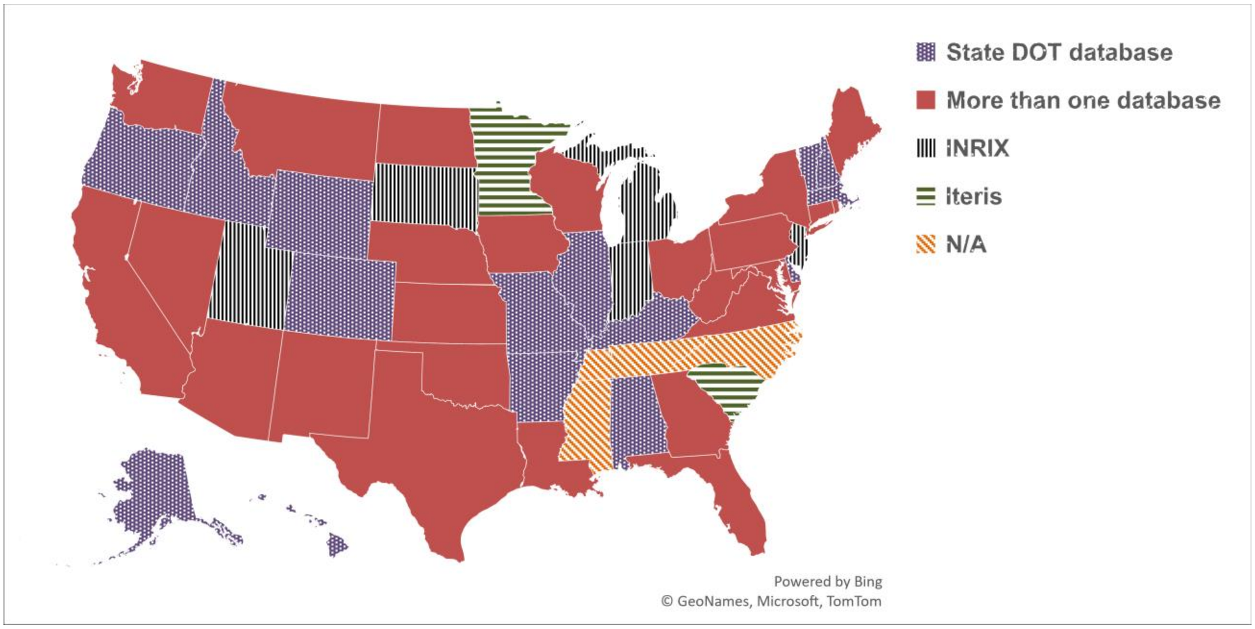


Figure 2. Usage of Big Data Providers by States.

From the figure, it can be seen that a few states use only one database (e.g., Indiana, Michigan, and New Jersey) but the majority of states employ more than one database to analyze their traffic and forecast and reduce congestion. Table 4 lists the big databases used by the state DOTs for freeway congestion reduction.

Table 4. Big Databases Used by the State DOTs.

State	Region	Agency	Public Big Databases	Private Big Databases
Alabama	South	Alabama DOT	NPMRDS, TMAS, ALDOT Databases	NA
Alaska	West	Alaska DOT	NPMRDS, TMAS, Alaska DOT Databases	NA
Arizona	West	Arizona DOT	NPMRDS, TMAS, ADOT Databases	StreetLight Data
Arkansas	South	Arkansas DOT	NPMRDS, TMAS, iDriveArkansas, ARDOT Databases	NA
California	West	Caltrans	NPMRDS, TMAS, Caltrans Databases	StreetLight Data, Waycare
Colorado	West	Colorado DOT	NPMRDS, TMAS, CDOT Databases	NA
Connecticut	Northeast	Connecticut DOT	NPMRDS, TMAS, CTDOT Databases	INRIX
Delaware	Northeast	Delaware DOT	NPMRDS, TMAS, DelDOT Databases	NA
Florida	South	Florida DOT	NPMRDS, TMAS, FDOT Databases	StreetLight
Georgia	South	Georgia DOT	NPMRDS, TMAS, GDOT Databases	Iteris, Inc.
Hawaii	West	Hawaii DOT	NPMRDS, TMAS, HIDOT Databases	NA
Idaho	West	Idaho Transportation Department (TD)	NPMRDS, TMAS, ITD Databases	NA
Illinois	Midwest	Illinois DOT	NPMRDS, TMAS, IDOT Databases	NA
Indiana	Midwest	Indiana DOT	NPMRDS, TMAS, INDOT Databases	INRIX
Iowa	Midwest	Iowa DOT	NPMRDS, TMAS, IOWADOT Databases	NA
Kansas	Midwest	Kansas DOT	NPMRDS, TMAS, KDOT Databases	Iteris, Inc.
Kentucky	South	Kentucky TC	NPMRDS, TMAS, KYTC Databases	NA
Louisiana	South	Louisiana DOTD	NPMRDS, TMAS, La DOTD Databases	INRIX, Streetlight
Maine	Northeast	Maine DOT	NPMRDS, TMAS, MaineDOT Databases	NA
Maryland	Northeast	Maryland DOT	NPMRDS, TMAS, MDOT Databases	INRIX, StreetLight
Massachusetts	Northeast	Massachusetts DOT	NPMRDS, TMAS, MassDOT Databases	NA
Michigan	Midwest	Michigan DOT	NPMRDS, TMAS, MDOT Databases	INRIX

State	Region	Agency	Public Big Databases	Private Big Databases
Minnesota	Midwest	Minnesota DOT	NPMRDS, TMAS, MnDOT Databases	Iteris, Inc.
Mississippi	South	Mississippi DOT	NPMRDS, TMAS, MDOT Databases	NA
Missouri	Midwest	Missouri DOT	NPMRDS, TMAS, MoDot Databases	NA
Montana	West	Montana DOT	NPMRDS, TMAS, MDT Databases	Iteris, Inc.
Nebraska	Midwest	Nebraska DOT	NPMRDS, TMAS, NDOT Databases	INRIX, Strava, Iteris, Inc.
Nevada	West	Nevada DOT	NPMRDS, TMAS, NDOT Databases	NA
New Hampshire	Northeast	New Hampshire DOT	NPMRDS, TMAS, NHDOT Databases	NA
New Jersey	Northeast	New Jersey DOT	NPMRDS, TMAS, NJDOT Databases	INRIX
New Mexico	West	New Mexico DOT	NPMRDS, TMAS, NMDO Databases	INRIX
New York	Northeast	New York DOT	NPMRDS, TMAS, NYCDOT Databases	NA
North Carolina	South	North Carolina DOT	NPMRDS, TMAS, NCDOT Databases	NA
North Dakota	Midwest	North Dakota DOT	NPMRDS, TMAS, NDDOT Databases	Iteris, Inc.
Ohio	Midwest	Ohio DOT	NPMRDS, TMAS, ODOT Databases	NA
Oklahoma	South	Oklahoma DOT	NPMRDS, TMAS, ODOT Databases	INRIX
Oregon	West	Oregon DOT	NPMRDS, TMAS, ODOT Databases	NA
Pennsylvania	Northeast	Pennsylvania DOT	NPMRDS, TMAS, PennDOT Databases	INRIX
Rhode Island	Northeast	Rhode Island DOT	NPMRDS, TMAS, DOTRI Databases	INRIX
South Carolina	South	South Carolina DOT	NPMRDS, TMAS, SCDOT Databases	Iteris, Inc.
South Dakota	Midwest	South Dakota DOT	NPMRDS, TMAS, SDDOT Databases	INRIX
Tennessee	South	Tennessee DOT	NPMRDS, TMAS, TNDOT Databases	NA
Texas	South	TxDOT	NPMRDS, TMAS, TxDOT Databases	Waze, Wejo, INRIX, Strava
Utah	West	Utah DOT	NPMRDS, TMAS, UDOT Databases	INRIX
Vermont	Northeast	Vermont Agency of Transportation (VTrans)	NPMRDS, TMAS, VTRANS Databases	NA

State	Region	Agency	Public Big Databases	Private Big Databases
Virginia	South	Virginia DOT	NPMRDS, TMAS, VDOT Databases	NA
Washington	West	Washington State DOT	NPMRDS, TMAS, WSDOT Databases	NA
West Virginia	South	West Virginia DOT	NPMRDS, TMAS, WVDOT Databases	Iteris, Inc.
Wisconsin	Midwest	Wisconsin DOT	NPMRDS, TMAS, WisDOT Databases	Iteris, Inc.
Wyoming	West	Wyoming DOT	NPMRDS, TMAS, WYDOT Databases	NA

Note: NA = not available, meaning information "not available" from internet search. It does not necessarily mean that these DOTs do not use private big databases.

2.6 CHAPTER SUMMARY

This chapter provided a comprehensive synthesis of big data providers, associated datasets, and AI platforms aimed reducing freeway congestion. The analysis was based on an extensive review of various documentation sources, such as websites, state DOT websites, research reports, journal articles, guidebooks, and handbooks. The chapter explored passive data generated by commercial third-party providers, such as Bluetooth devices, GPS-enabled devices, cellular phones, and smartphones with location services. It emphasized the advantages of passive data overactive data collection methodologies used by DOTs and highlights the potential benefits of using passive data to supplement or replace active data collection efforts. Last, it reviewed state DOT practices in adopting data initiatives and utilizing big databases and platforms to analyze and forecast traffic for effective freeway congestion reduction.

CHAPTER 3: SURVEY

3.1 INTRODUCTION

The TTI team developed a survey questionnaire to learn more about state DOT practices regarding big data usage in freeway congestion reduction. The survey contained the following sections: data collection, data analysis and methods, big data platforms, reports and guidance documents, lessons learned, and future needs. This survey had conditional logic built in to allow agencies to skip over nonapplicable questions. The final survey questionnaire is provided in Appendix A. The TTI team developed a list of 60 DOT contact persons as potential survey participants. The request to complete the survey was sent to the selected contact persons. As of February 26, 2022, the TTI team had received 14 responses. Out of these 14 responses, 6 responses were discarded due to incomplete submission. The agencies that completed the survey are listed below:

- Colorado DOT.
- Arizona DOT.
- North Carolina DOT.
- Wisconsin DOT.
- Georgia DOT.
- Pennsylvania DOT.
- Kansas DOT.
- New Hampshire DOT.

3.2 SURVEY ANALYSIS

Survey respondents were asked about the big data providers' data that they use the most to aid in freeway congestion reduction. The majority of the responding agencies used INRIX and Waze as their main sources. Some of the agencies have also started to use Streetlight and Wejo data (see Figure 3). Participant agencies also mentioned other data vendors such as HERE, Intelight/Maxview, WayCare, TomTom, and Drakewell.

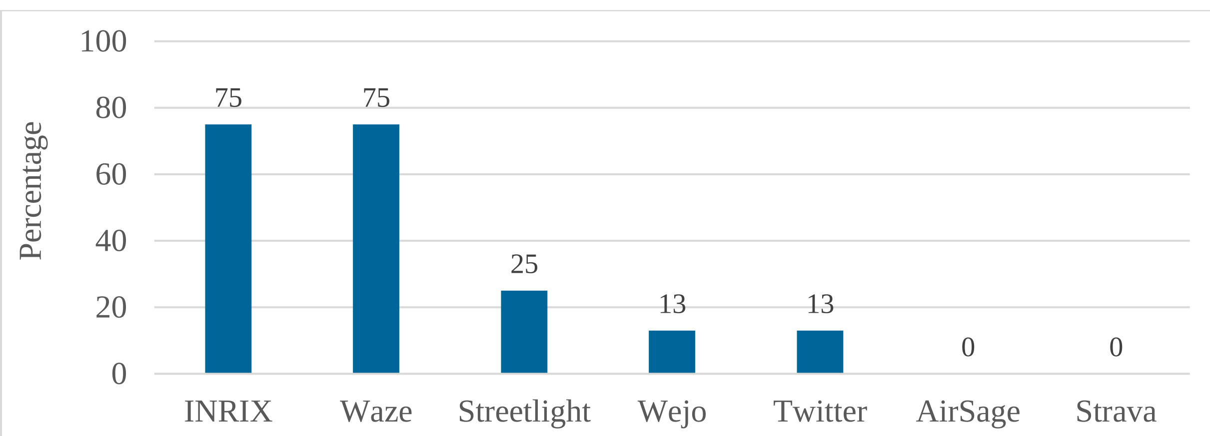


Figure 3. Big Data Providers.

Survey respondents were asked about the types of information they collect with the use of big data. All of the participant agencies used big data sources for the following four tasks (see Figure 4):

- Measuring travel time.
- Collecting real-time information.
- Managing incidents.
- Reducing incident clearance time.

Agencies also determine congestion measures and distribute emergency alerts and crisis information by using information from big data sources.

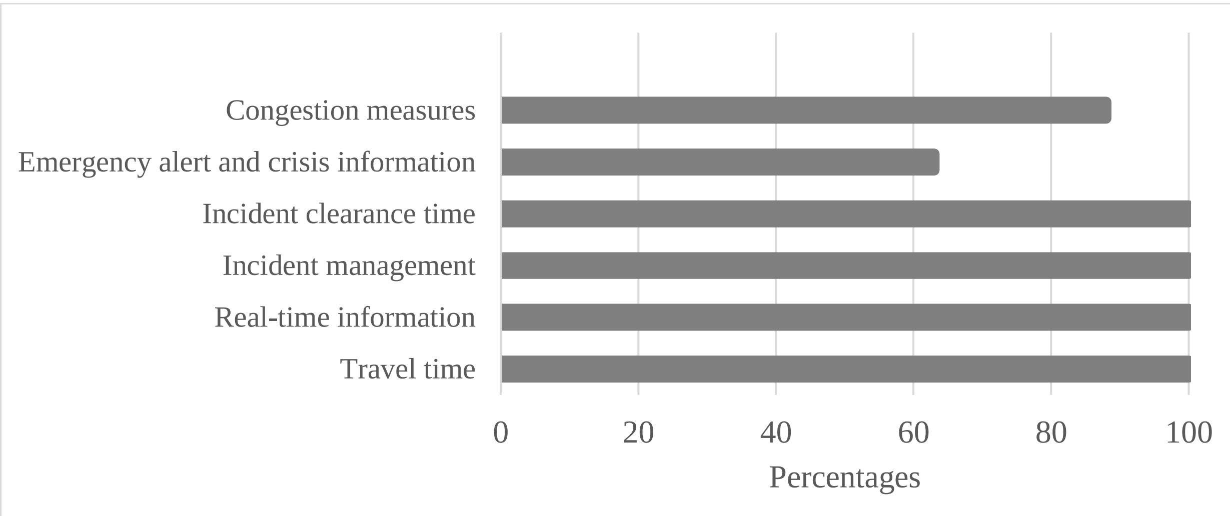


Figure 4. Type of Information Collected from Big Data Providers.

When asked about the types of information by different data vendors, the majority of the responses indicated that either INRIX or Waze data are mostly used for these tasks (see Table 5).

Table 5. Big Data Providers by Type of Information.

Type of Information	INRIX	AirSage	Waze	Streetlight	Strava	Others
Congestion measures	50	—	—	—	—	—
Travel time	50	—	—	—	—	—
Incident management	25	—	13	—	—	—
Incident clearance time	25	—	—	—	—	—
Real-time information	38	—	63	—	—	—
Emergency alert and crisis information	25	—	13	—	—	—
Other (specify below)	50	—	—	—	—	—

Note: Dash indicates not applicable

The participants were also asked about the types of data they collect from big data sources. Incident information, short-duration traffic volume data, and traffic crash data are the key data types according to the participant responses (see Figure 5). Other data types are weather data, traffic volume by vehicle type, speed by vehicle type, travel time by vehicle type, and social media feed. One of the participants mentioned, "Overall travel times along with congestion monitoring via Google maps, cameras, physical traffic detectors, and probe data populate and confirm congestion locations. We typically do not mitigate reoccurring congestion."

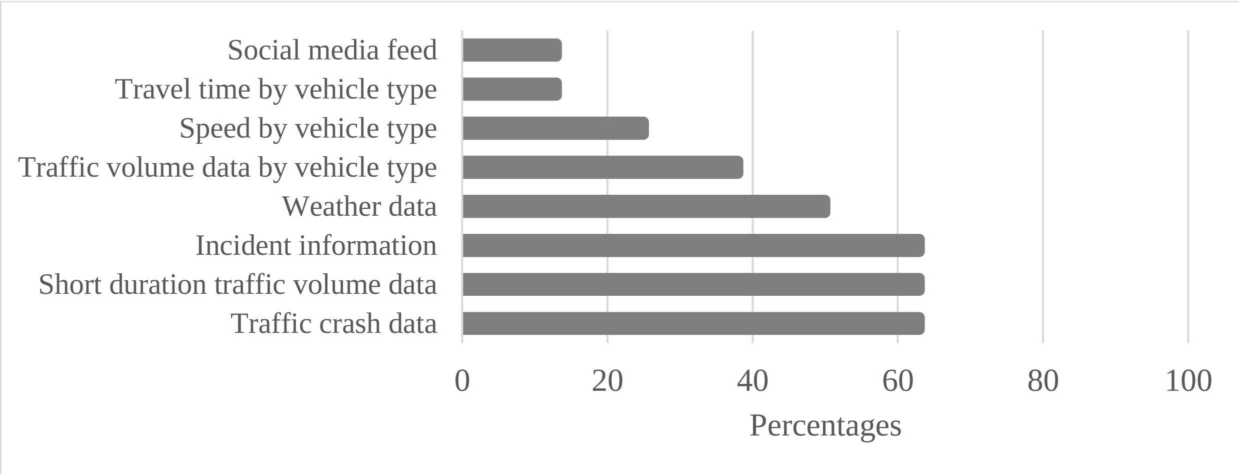


Figure 5. Data Types Used by the Participant Agencies.

The participants were asked about whether the collection of these data is continuous or occasional. The majority of the responses indicated that the data collection is continuous (see Table 6).

Table 6. Type of Data Collection by Periods.

Type of Information	Continuous	Once a week	Once a month	Once a year	Others
Traffic volume data	88	—	12	12	—
Traffic volume data by vehicle type	63	—	12	12	—
Travel time	100	—	12	—	—
Speed	88	—	12	—	—
Travel time by vehicle type	—	—	25	—	—
Other (specify below)	25	—	25	12	—

The cloud platforms used by the survey participants are Amazon AWS, Microsoft Azure, and Google Cloud (see Figure 6). Around one-fourth of the participants mentioned that they do not use cloud platforms.

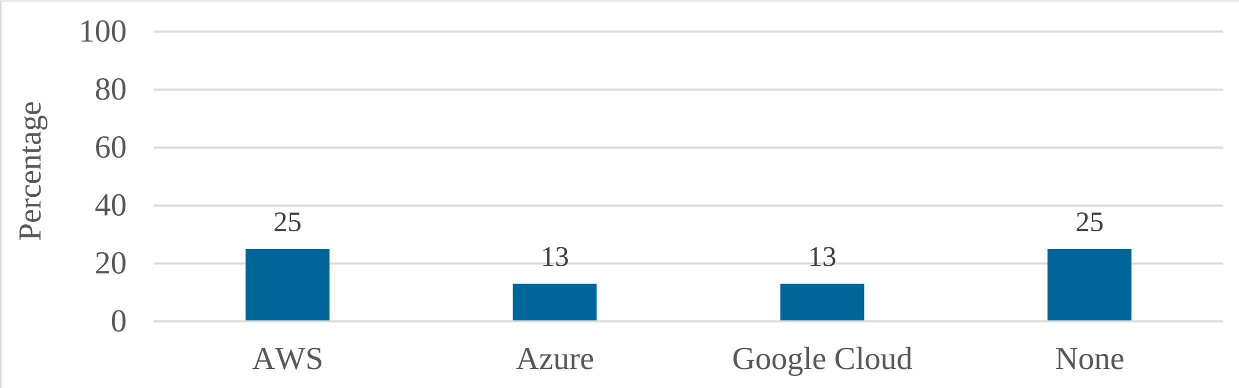


Figure 6. Cloud Platforms Used by the Participant Agencies.

The most used analytical platforms are Power BI and Tableau (see Figure 7). Around 38 percent of participants mentioned that they do not use any analytical platforms. Similarly, around 67 percent of the participants mentioned that they do not use any enterprise software. However, when asked about DOT-maintained dashboards, 63 percent of the participants mentioned that state DOTs maintain an interactive congestion dashboard.

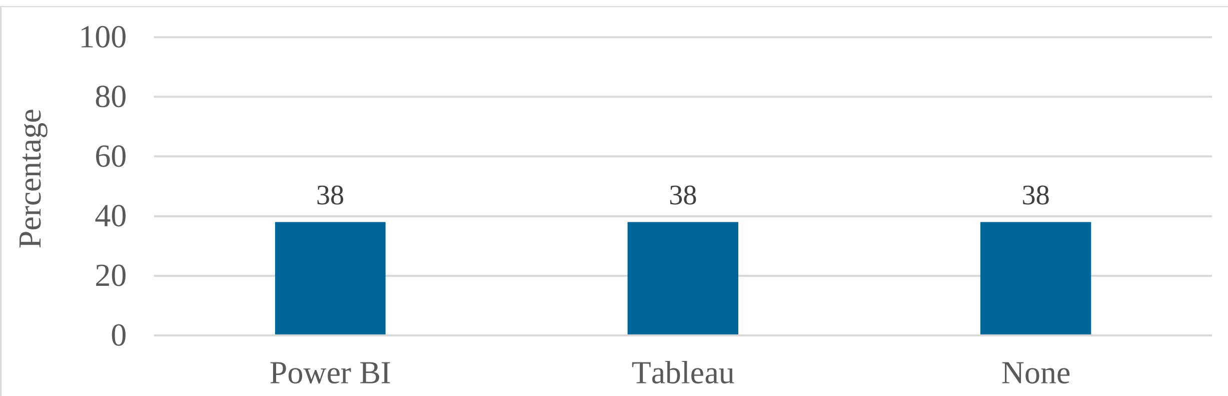


Figure 7. Analytics Platforms Used by the Participant Agencies.

The majority of the participants did not respond to the questions regarding lessons learned and future plans. The responses of two participants are below:

- Adaptive ramp metering has helped reduce congestion in parts of the Phoenix Metro area. Applying restriping judiciously has helped reduce congestion in some areas of the Phoenix Metro area. Implementing a dedicated Incident Response Unit has improved response and incident clearance times in the Phoenix Metro area.
- We are currently addressing some of the data challenges through an National Cooperative Highway Research Program (NCHRP) project, CV pool fund studies, and working through a data Governance Program that is in its early stages in the agency. Our largest challenge seems to be having the bandwidth to work on this effort.

3.3 CHAPTER SUMMARY

As discussed in this chapter, the TTI team conducted a survey to explore the use of big data by state DOTs to reduce freeway congestion. The survey was sent to 60 DOT contact persons, and 14 responses were received. The main big data sources used by the agencies were INRIX and Waze, with some also using Streetlight and Wejo data. The collected data primarily served tasks like measuring travel time, real-time information, incident management, and incident clearance time. INRIX and Waze were also the top choices for specific types of information. Key data types collected included incident information, short-duration traffic volume data, and traffic crash data. Most data collection was continuous, and Amazon AWS, Microsoft Azure, and Google Cloud were commonly used cloud platforms. The preferred analytical platforms were Power BI and Tableau. While some agencies did not provide responses on lessons learned and future plans, a few mentioned successes like adaptive ramp metering and incident response units for reducing congestion, as well as challenges related to data governance and bandwidth limitations.

CHAPTER 4: BIG DATA VALIDATION

4.1 INTRODUCTION

This chapter presents research findings of novel big data applications within traffic management, traffic forecasting and modeling, and real-time, reliability monitoring-oriented systems developed by state and federal agencies. The goal of this review was to identify noteworthy practices and tools that could be transferable at TxDOT. To collect the information presented herein, the TTI team conducted a series of activities in the following order:

- Conducted a big data validation case study using Wejo data.
- Developed a dataset for Texas freeways using Road-Highway Inventory Network Offload (RHiNO), National Performance Management Research Data Set (NPMRDS), and TxDOTs's Crash Records Information System (CRIS) data.

4.2 BIG VALIDATION DATA CASE STUDY: WEJO

CVs are rapidly becoming the new paradigm of road transport and are widely believed to positively influence transportation safety, efficiency, and sustainability. CVs represent the unification of various connectivity technologies, enabling vehicles to communicate with other vehicles, transportation infrastructures, and the cloud to achieve the goal of self-driving. Although most commercially available vehicles are still far from completely automating the driving task, most of them already can monitor the driving environment and vehicle movements through vehicular sensors. Many world-leading auto manufacturers, such as Toyota™, General Motors™, BMW™, Tesla™, and others, have ramped up the production of CVs, which will increase access and transmission of vehicular sensors' data to the cloud. Many automotive data companies have also emerged to facilitate the use of CV data, such as Wejo, Otonomo, Smartcar, Vinili, and CarAlgo. These data companies bridge the data providers (auto manufacturers) with data users by ingesting, aggregating, and normalizing the raw CV data and delivering the enriched and organized datasets to end users.

The CV data are collected from vehicles, thereby directly reflecting the dynamics of traffic mobility. CV data show great superiority in data quality, volume, consistency, and richness over traditional mobility data sources, making them promising data sources for monitoring urban mobility dynamics. For example, Wejo, as a leading CV data start-up, provides high sampling rates and multidimensional vehicle movements and driving events (e.g., hard braking, hard acceleration, and speeding) data. This data platform is currently partnered with multiple world-leading auto manufacturers and has collected data from millions of vehicles, with a sampling rate of 3 seconds per waypoint. Each waypoint describes the timestamp, location, and movement-related information (e.g., speed and heading) of a vehicle's trajectory. Wejo claims that its CV data products can access over 90 different vehicular sensors and cover 95 percent of road networks in the United States, with about 12 billion data points collected every day at its best temporal resolution of every 3 seconds. This preliminary database development also demonstrates that Wejo data have good spatiotemporal coverage in urban regions of Texas.

Note that Wejo mostly collected CV data from passenger vehicles and not from commercial motor vehicles (CMVs). However, the driving behavior of both passenger vehicles and CMVs would be affected similarly by shockwaves while traversing work zones. Therefore, the TTI team assumed that Wejo CV data can be effectively leveraged to characterize driving behavior for both passenger vehicles and CMVs. To support this assumption, Figure 8 shows how Wejo data reveal changes in driving behavior, such as hard-braking events, reductions in speed, and lane changes due to a lane closure that is causing congestion and forcing vehicles to move from the right lane to the left lane of the roadway.

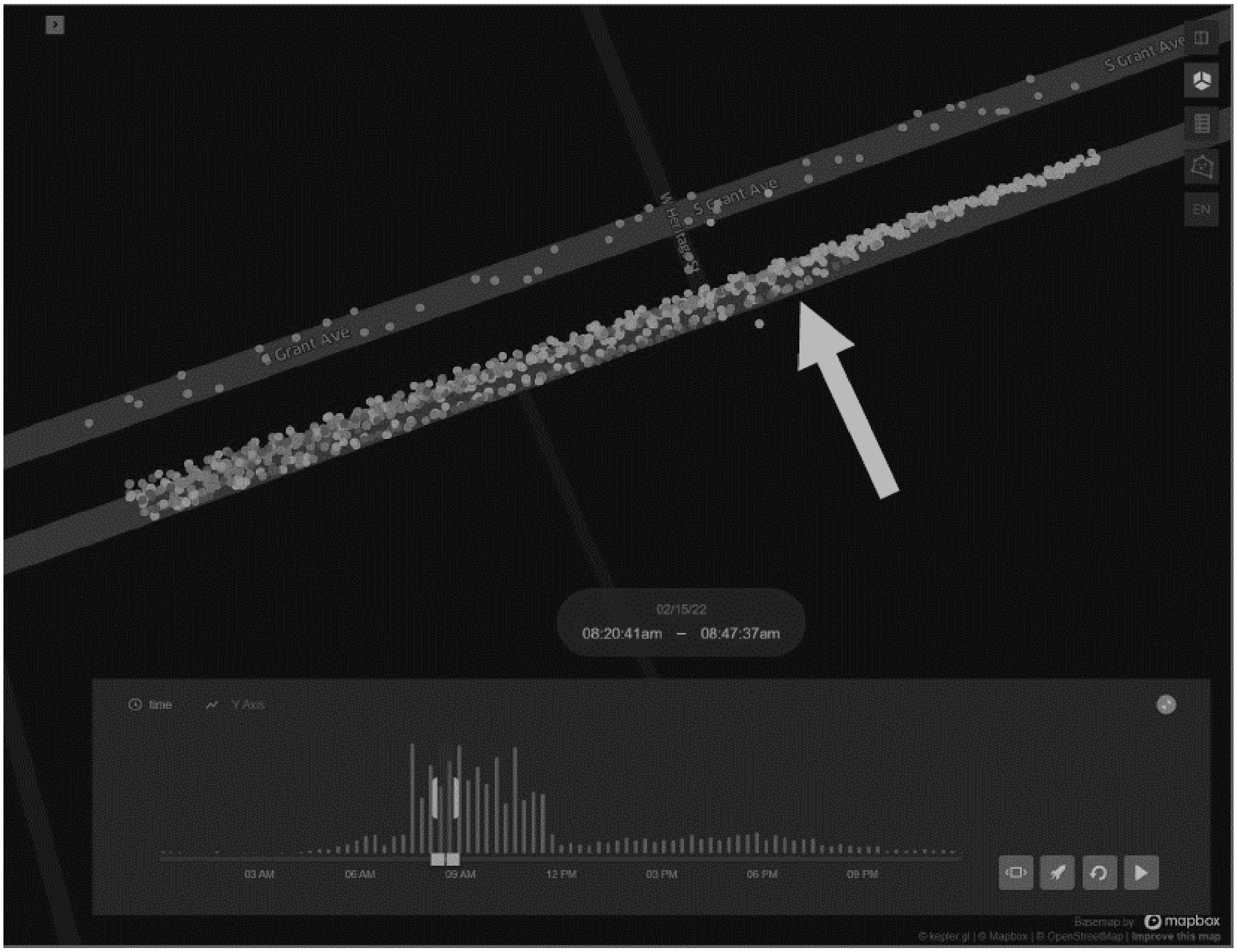


Figure 8. Changes in Driving Behavior Due to Lane Closure.

In this project, the TTI team used the CV movement data provided by Wejo to generate variables characterizing the speed variation. The CV data were reprocessed by Wejo and delivered to the Azure Cloud storage account. The data were organized in the Apache Parquet format. The TTI team used an online big data analytics platform, Azure Databricks™, to process the big CV dataset. Azure Databricks supports the latest version of Apache Spark™, which allows its users to seamlessly integrate with any open-source libraries and quickly establish a fully managed Apache Spark environment. The TTI team primarily used Apache Sedona™ to load, partition, process, and spatially analyze the big CV data and used other open-source libraries (e.g., Datashader) to visualize the large dataset. To calculate the desired parameters, the unique dataPointIds within each segment that contained data for 365 consecutive days were selected. Table 7 provides a list of variables for congestion measures using Wejo data, along with their definitions. In the table, n represents the number of unique dataPointIds within each segment, v_i

represents speed of the vehicle, $\bar{v}$ represents the average speed, v_{85} represents the 85th percentile speed for the segment (speed that is greater than 85 percent of the data points), v_{50} represents the 50th percentile (median) speed, v_{93} represents the 93rd percentile speed for the segment (speed that is greater than 93 percent of the data points), v_{07} represents the 7th percentile speed for the segment (speed that is greater than 7 percent of the data points), and a_i represents the acceleration of the vehicle. To obtain the stop_freq variables, whenever a vehicle speed went down from above 5 miles per hour (mph) to below 5 mph it was counted as a stop. Then, the number of stops was divided by the segment length and the number of unique journeys in that segment.

Table 7. Developed Congestion Measures Using Wejo Data.

Variable	Definition
Mean_spd	$\text{Average Speed} = \frac{1}{n}\sum_{i=1}^{n} v_i$
Variance_spd	$\text{Sample variance of speeds} = \frac{1}{n-1}\sum_{i=1}^{n}(v_i^2 - \bar{v}^2)$
Percentile_83_spd	85th percentile of speeds
Percentile_93_spd	93rd percentile of speeds
RMS_spd	$\text{Root mean square of speeds} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} v_i^2}$
Diff_to_mean	$\text{Difference to mean speed} = v_{85} - \bar{v}$
Coeff_upper_speed	$\text{Coefficient of upper speed} = \frac{v_{85} - v_{50}}{v_{50}}$
SI	$\text{Skewness index} = 2 \times \frac{v_{93} - v_{50}}{v_{93} - v_{07}}$
Acc_noise	Acceleration noise or root mean square of acceleration $= \sqrt{\frac{1}{n}\sum_{i=1}^{n} a_i^2}$
Stop_freq	Stop frequency: number of stops per trip per mile

The final dataset can be retrieved from the following link: https://tti-my.sharepoint.com/:f:/g/personal/s-das_tti_tamu_edu/EngdaEI8phJHg_R-N0CDPYkByspNdGCyj-mk2zjkGaHu3g?e=PIlR5Y.

4.3 DATABASE PREPARATION

The work for Texas State data with the Roadway Inventory annual data, the NPMRDS speed measure, and crash events count mainly includes five major parts: (1) preparation of homogeneous roadway segments, (2) conflation of the segment dataset with the NPMRDS dataset, (3) speed measure calculation, (4) crash and point events assignment, and (5) final combination process. The freeway is considered the type of roadway investigated in this study. The detailed steps are described in the following sections.

4.3.1 Preparation of Homogenous Roadway Safety Segments

A based layer was created to extract roadways based on the functional classification. A data conflation method was applied to prepare the base layers. In Part 1, there are two data conflation steps to prepare the based layers based on the original Roadway Inventory annual data. First, all interstate roadways were selected by setting “F_SYSTEM” = 1. Second, an additional segment identification variable was added. The original Roadway Inventory annual data used “RIA_RTE_ID” to identify the road segments. However, it contains both digits and characters of different lengths. To make the further conflation steps easier, a “RouteID” variable was created by giving a unique value for each “RIA_RTE_ID” number. The “RouteID” starts at 1 and ends at 1361. If multiple segments have the same “RIA_RTE_ID”, they will have the same “RouteID” as well (see Figure 9).

Table

2020_RHiNO_Freeway

FID	Shape	routeid	OBJECTID	REC	RIA_RTE_ID	RTE_GRID	GID	FRM_DFO	TO_DFO	C_SEC	CON	SEC	BMP	EMP	F
34293	Polyline M	1360	116102	1	IH0820-RG	153722	1599256522	7.488	7.701			0	0	0	
34295	Polyline M	1360	116105	1	IH0820-RG	153722	1599256522	25.335	25.429			0	0	0	
34313	Polyline M	1360	116132	1	IH0820-RG	153722	1599256522	18.569	18.57			0	0	0	
34314	Polyline M	1360	116133	1	IH0820-RG	153722	1599256522	33.841	33.847			0	0	0	
34322	Polyline M	1360	116147	1	IH0820-RG	153722	1599256522	1.226	1.285			0	0	0	
34342	Polyline M	1360	116178	1	IH0820-RG	153722	1599256522	12.755	12.78			0	0	0	
34343	Polyline M	1360	116179	1	IH0820-RG	153722	1599256522	30.495	30.759			0	0	0	
34372	Polyline M	1360	116225	1	IH0820-RG	153722	1599256522	7.468	7.488			0	0	0	
34374	Polyline M	1360	116228	1	IH0820-RG	153722	1599256522	23.532	23.596			0	0	0	
34417	Polyline M	1360	116295	1	IH0820-RG	153722	1599256522	20.02	20.098			0	0	0	
34418	Polyline M	1360	116296	1	IH0820-RG	153722	1599256522	32.112	32.377			0	0	0	
34458	Polyline M	1360	116359	1	IH0820-RG	153722	1599256522	15.444	15.464			0	0	0	
34459	Polyline M	1360	116360	1	IH0820-RG	153722	1599256522	32.839	33.149			0	0	0	
34486	Polyline M	1360	116404	1	IH0820-RG	153722	1599256522	13.533	13.578			0	0	0	
34487	Polyline M	1360	116405	1	IH0820-RG	153722	1599256522	26.933	27.086			0	0	0	
34517	Polyline M	1360	116453	1	IH0820-RG	153722	1599256522	9.815	10.393			0	0	0	
34519	Polyline M	1360	116456	1	IH0820-RG	153722	1599256522	24.635	25.017			0	0	0	
34543	Polyline M	1360	116496	1	IH0820-RG	153722	1599256522	1.285	1.814			0	0	0	
34562	Polyline M	1360	116524	1	IH0820-RG	153722	1599256522	16.211	16.643			0	0	0	
34563	Polyline M	1360	116525	1	IH0820-RG	153722	1599256522	27.086	27.167			0	0	0	
34589	Polyline M	1360	116570	1	IH0820-RG	153722	1599256522	13.626	13.663			0	0	0	
34619	Polyline M	1360	116616	1	IH0820-RG	153722	1599256522	5.695	5.721			0	0	0	
34621	Polyline M	1360	116619	1	IH0820-RG	153722	1599256522	22.64	22.645			0	0	0	
31495	Polyline M	1361	111647	1	IH0820-SG	153722	1574720254	10.446	10.468			0	0	0	
31740	Polyline M	1361	112042	1	IH0820-SG	153722	8476	13.772	14.185			0	0	0	
31849	Polyline M	1361	112213	1	IH0820-SG	153722	8476	14.205	14.231			0	0	0	
32212	Polyline M	1361	112788	1	IH0820-SG	153722	1574720254	10.397	10.424			0	0	0	
32718	Polyline M	1361	113592	1	IH0820-SG	153722	1574720254	9.838	10.397			0	0	0	
33111	Polyline M	1361	114220	1	IH0820-SG	153722	8476	14.185	14.205			0	0	0	
33152	Polyline M	1361	114282	1	IH0820-SG	153722	1574720254	10.424	10.446			0	0	0	
33330	Polyline M	1361	114567	1	IH0820-SG	153722	8476	14.231	14.383			0	0	0	
33476	Polyline M	1361	114800	1	IH0820-SG	153722	8476	14.383	14.536			0	0	0	
33616	Polyline M	1361	115027	1	IH0820-SG	153722	1574720254	10.468	11.123			0	0	0	
34593	Polyline M	1361	116575	1	IH0820-SG	153722	8476	14.536	14.751			0	0	0	

0 (0 out of 34624 Selected)

2020_RHiNO_Freeway

Figure 9. Illustration of RouteID.

4.3.2 Conflation of the Segment Dataset with the NPMRDS Dataset

In this step, the TTI team integrated two linear systems (NPMRDS and segment polyline files) to enable segment-based analysis with speed measures. This study used two formats of segment files: (a) roadway linework without roadway characteristics (i.e., RouteID), and (b) roadway linework with roadway characteristics (i.e., roadway segments created in the first step). In the former file, each route is a continuous polyline in the GIS database, and it includes only the basic information for the route (e.g., route name, beginning, and end mileposts). In the latter file, routes are split into various numbers of homogeneous segments with roadway assets or features (e.g., functional class, lane, shoulder, posted speed limit [PSL], presence of rumble strips, and lighting situation).

The conflation work considered the 2017 Texas NPMRDS file and the homogeneous segment file created from the Roadway Inventory annual data in the first step. The TTI team used two software packages (ArcGIS and R) to conflate these databases. The following steps were taken in this task.

4.3.2.1. Step 1: Divide the NPMRDS File by Direction

In this step, the TTI team divided the NPMRDS file into two files: positive and negative. The direction of the NPMRDS segments is determined by the TMC name: (a) "+" or "P" indicates positive, and (b) "–" or "N" indicates negative (see Figure 10 and Figure 11).

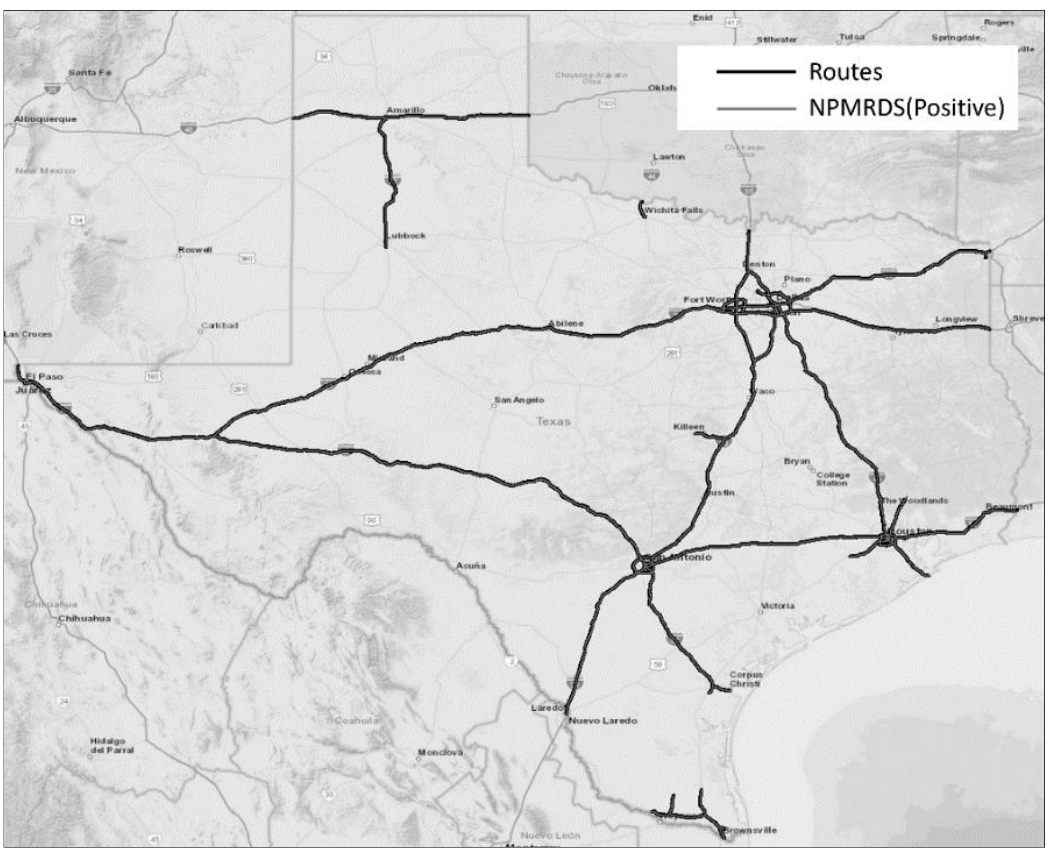


Figure 10. Step 1 (Roadway Segment and NPMRDS Linework).

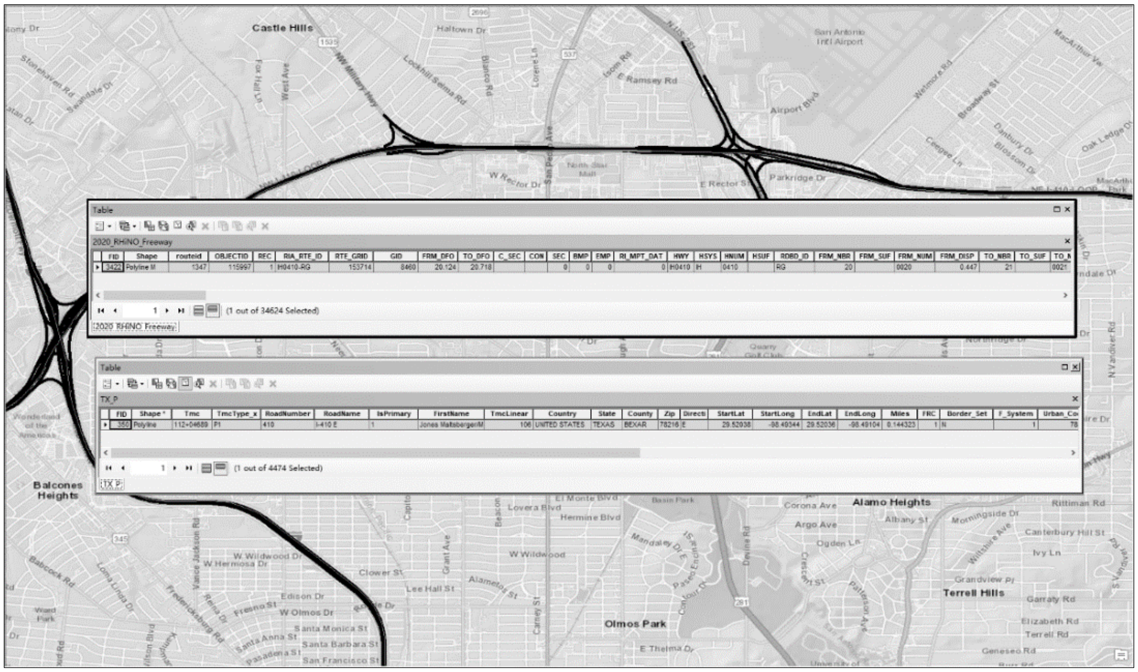


Figure 11. Example of a Route and an NPMRDS TMC Segment.

4.3.2.2. Step 2: Locate NPMRDS Segments along with Roadway Routes

In this step, the TTI team used one direction of NPMRDS files as the input feature and roadway route linework without roadway characteristics as the input route feature and then located the NPMRDS segments on the roadway routes. The event table generated from the locating process was exported as a CSV file. Each direction of NPMRDS was located separately (see Figure 12).

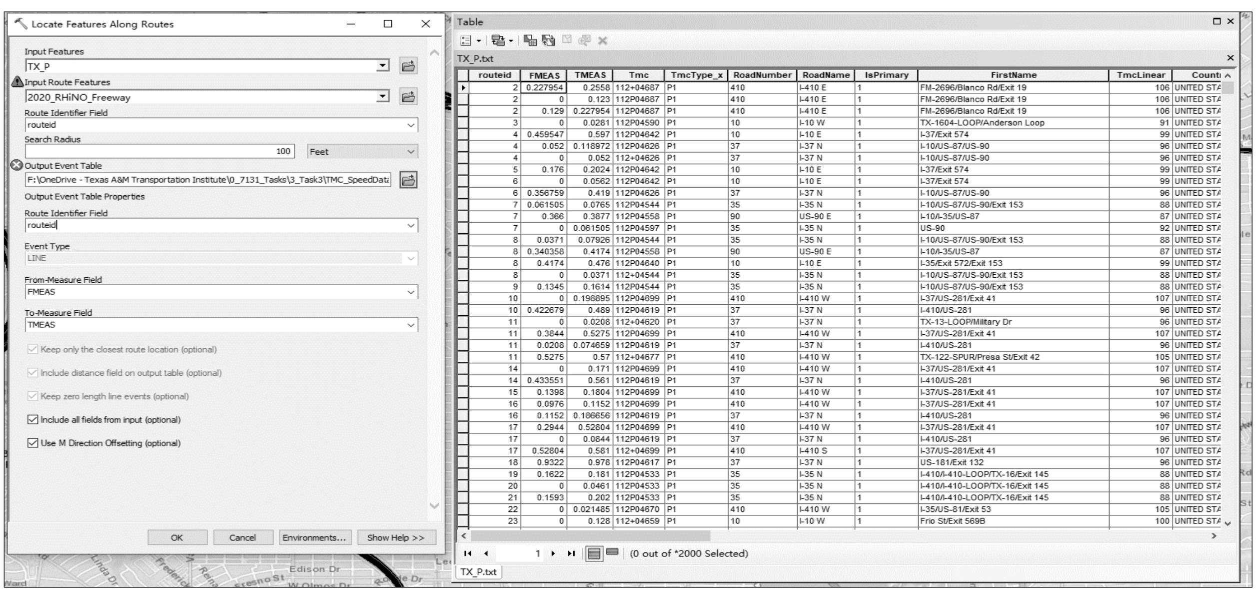


Figure 12. Step 2 (Locating NPMRDS Segments along Roadway Routes).

4.3.2.3. Step 3: Refine the Event Table

In Step 2, the two files (i.e., NPMRDS and Routes) were located based on spatial relationships. A few segments were mismatched in the process. Step 3 eliminated the mismatched events based on the roadway information and spatial matching results. In the refined event table, each NPMRDS segment had a route name and beginning and end mileposts relative to the roadway route linework.

4.3.2.4. Step 4: Create the Final Table

The TTI team refined the event table and roadway linework with roadway characteristics. In this step, the association between NPMRDS segments and roadway segments with roadway characteristics was created based on the route name and mileposts. For each roadway segment, the data from NPMRDS segments were collected. The information included the TMC name and the effective length ratio of the TMC matching with the roadway segment (see Table 8).

Table 8. Step 4 (Roadway Segments with NPMRDS Information).

RouteID	Unique ID	Beg MP	End MP	TMC	Effective Ratio
92	2	15280.1	16871.3	114+05423	0.041
92	3	16871.3	18564.8	114+05423	0.043
92	4	60632.2	61160.4	114+05420	0.078
92	4	60632.2	61160.4	114+05421	0.001
92	5	58192.9	60630.6	114+05421	0.108
92	6	67278.1	67471.1	114P05419	0.049
92	7	67471.1	68624.31	114P05419	0.291

Similar steps were taken for the conflation of the roadway routes, segments, and NPMRDS in the negative direction. The current data conflation work is limited to the quality of the original Roadway Inventory database data as well as the NPMRDS dataset. For example, if a roadway feature is missing in the Roadway Inventory data, the feature will be shown as “NA” or “0” in the final conflated dataset. A few segments had two or more entries in the attribute table, and the information contradicted each other.

4.3.3 Speed Measure Calculation

The TTI team downloaded 5 years of speed measure data (2017 to 2021) from NPMRDS by using the TMC number from Step 2. First, the TTI team developed an R code to summarize the 5-year speed measure for each TMC name. Speed measures used in this study are listed in Table 9.

Table 9. Speed Measure Variables and Definitions.

Attribute Name	Definition
SpdAve	Average speed determined for the year using all data
SpdStd	Standard deviation (STD) of speed determined for year using all data
Spd85	85th percentile speed determined for the year using all data
PSL	Posted speed limit
SpdAveDay	Average speed determined for year (hour > 5 and hour < 18) using all data
SpdStdDay	STD of speed determined for year (hour > 5 and hour < 18) using all data
SpdAveNight	Average speed determined for year (hour > 17 and hour < 24 and hour > −1 and hour < 6) using all data
SpdStdNight	STD of speed determined for year (hour > 17 and hour < 24 and hour > −1 and hour < 6) using all data
SpdAveMTWT	Average speed determined for year (Mon., Tues., Wed., and Thurs.) using all data
SpdStdMTWT	STD of speed determined for year (Mon., Tues., Wed., and Thurs.) using all data
SpdAveFSS	Average speed determined for year (Fri., Sat., and Sun.) using all data
SpdStdFSS	STD of speed determined for year (Fri., Sat., and Sun.) using all data
SpdFFAve	Average speed determined for year using speed data where 5-min speed is > PSL (or PSL+5 or +10)
SpdFF85	85th percentile speed determined for year using speed data where 5-min speed is > PSL (or PSL+5 or +10)

In Part 2, it is known that each roadway segment has both negative and positive TMC names, and for some roadway segments, there is more than one negative or positive TMC name. This situation leads to each roadway segment having more than one set of speed measure data. The TTI team developed another R code to calculate the weighted average of all the speed measure attributes for each unique roadway segment. The original speed measure file was merged with the roadway segments with the NPMRDS information file created in Part 2. Then, the speed measure data were grouped by "unique id" and the weighted average values of all speed measure attributes in Table 9 were calculated based on the effective ratio of each TMC. Note that the weighted average calculation methods for "SpdStd," "SpdStdDay," "SpdStdNight," "SpdStdMTWT," and "SpdStdFSS" are different from other speed measure attributes.

Equation 1 was applied to the following speed measurement variables: "SpdStd," "SpdStdDay," "SpdStdNight," "SpdStdMTWT," and "SpdStdFSS."

$$A = \sqrt{\sum_{i}^{n} (w_i s_i)^2} \tag{1}$$

Where,
A: Aggregated speed measurement variable.
n: Number of unique TMC names of a RHiNO segment.

w_i: Normalized effective ratio of the i^th TMC name of a RhiNO segment.
s_i : the speed measurement value of the i^th TMC name a RhiNO segment.

Equation 2 was applied to the following speed measurement variables: "SpdAve," "Spd85," "RefSpd," "SpdAveDay," "SpdAveNight," "SpdAveMTWT," "SpdAveFSS," "SpdFFAve," and "SpdFF85."

$$A = \sum_{i}^{n} w_i s_i \tag{2}$$

Where,
A: Aggregated speed measurement variable.
n: Number of unique TMC names of a RHiNO segment.
w_i: Normalized effective ratio of the i^th TMC name of a RHiNO segment.
s_i : the speed measurement value of the i^th TMC name a RHiNO segment

Moreover, three new columns were added—"num_tmc_n," "num_tmc_p," and "ratio_tmc_n_p." "Num_tmc_n" indicates the number of negative TMC for each "unique id," "num_tmc_p" indicates the number of positive TMC for each "unique id," and "ratio_tmc_n_p" indicates the ratio between the sum of negative TMCs' "Effective Ratios" and the sum of positive TMCs' "Effective Ratios."

The speed measurements were summarized based on the above explanation and added to the based layer prepared in the first step. Columns 208 to 277 in the final shapefile represent the speed measurements for each year ordered from 2017 to 2021. Columns 278 to 291 represent the speed measurements for all years combined. The detailed field definition of speed measurement for 2017 is shown in Table 10. For the remaining years, the naming criteria are the same.

Table 10. Speed Measure Variables for 2017 in the Final Shapefile.

Field Name	Definition
SpdAve17	Average speed determined for 2017
SpdStd17	STD of speed determined for 2017
Spd85_17	85th percentile speed determined for 2017
PSL_17	PSL for 2017
SpdAveD17	Average speed during daytime determined for 2017
SpdStdD17	STD of speed during daytime determined for 2017
SpdAveN17	Average speed during nighttime determined for 2017
SpdStdN17	STD of speed during nighttime determined for 2017
AveMTWT17	Average speed determined for 2017 (Mon Tues Wed Thurs)
StdMTWT17	STD of speed determined for 2017 (Mon Tues Wed Thurs)
AveFSS17	Average speed determined for 2017 (Fri Sat Sun)
StdFSS17	STD of speed determined for 2017 (Fri Sat Sun)
SpdFFAve17	Average speed determined for 2017 using speed data where 5-min speed is > PSL (or PSL+5 or +10)
SpdFF85_17	85th percentile speed determined for 2017 using speed data where 5-min speed is > PSL (or PSL+5 or +10)

4.3.4 Crash and Point Events Assignment

The TTI team downloaded Texas crash data from 2017 to 2021. The crash events shape files contain a large amount of information, including crash date, crash severity, and the latitude and longitude of the crash events. The TTI team first applied the near feature tool in ArcMap (see Figure 13) to assign crash events to roadway segments that were created in Part 1. Each crash was assigned an attribute named "Near_FID." "Near_FID" stands for the Feature Identifier (FID) number (roadway segment corresponding row number starting from 0 in the crash characteristics file) of the roadway segment on which the crash event happened. Here, the TTI team chose 30 ft as the threshold to determine if a crash event happened on a roadway segment, which means that the crash event is assigned to this specific roadway segment only if a crash event is within 30 ft of the roadway segment. If a crash event cannot be assigned to any roadway segment, its "Near_FID" attribute is equal to −1. The TTI team filtered out the crash events whose "Near_FID" equaled −1 because they cannot relate to any roadway segments. Note that in the Roadway Inventory annual data, all the different types of roadbeds were classified in "RDBD_ID" variables. The crash events should not be assigned to the centerline/single roadbed (i.e., "RDBD_ID" = "KG"). Therefore, before assigning crash events, all the centerline/single roadbed segments were excluded by filtering out "RDBD_ID" = "KG."

After crash events were assigned to roadway segments for both 2017 and 2018, the TTI team summarized the total number of crashes that happened on each roadway segment based on crash severity. The crash severity levels range from 0 to 5. The detailed severity explanations of levels are shown in Table 11. The crash types are categorized as single vehicle (SV) or multiple vehicle (MV). Therefore, for each "NearFID" number, the TTI team summarized 30 new columns named “S7_0”, “S17_1” … “M21_5”, and “M21_5” (see Table 12). For example, the column “S17_1” indicates the number of severity level 1 crashes that happened during 2017 with the crash type of single vehicles.

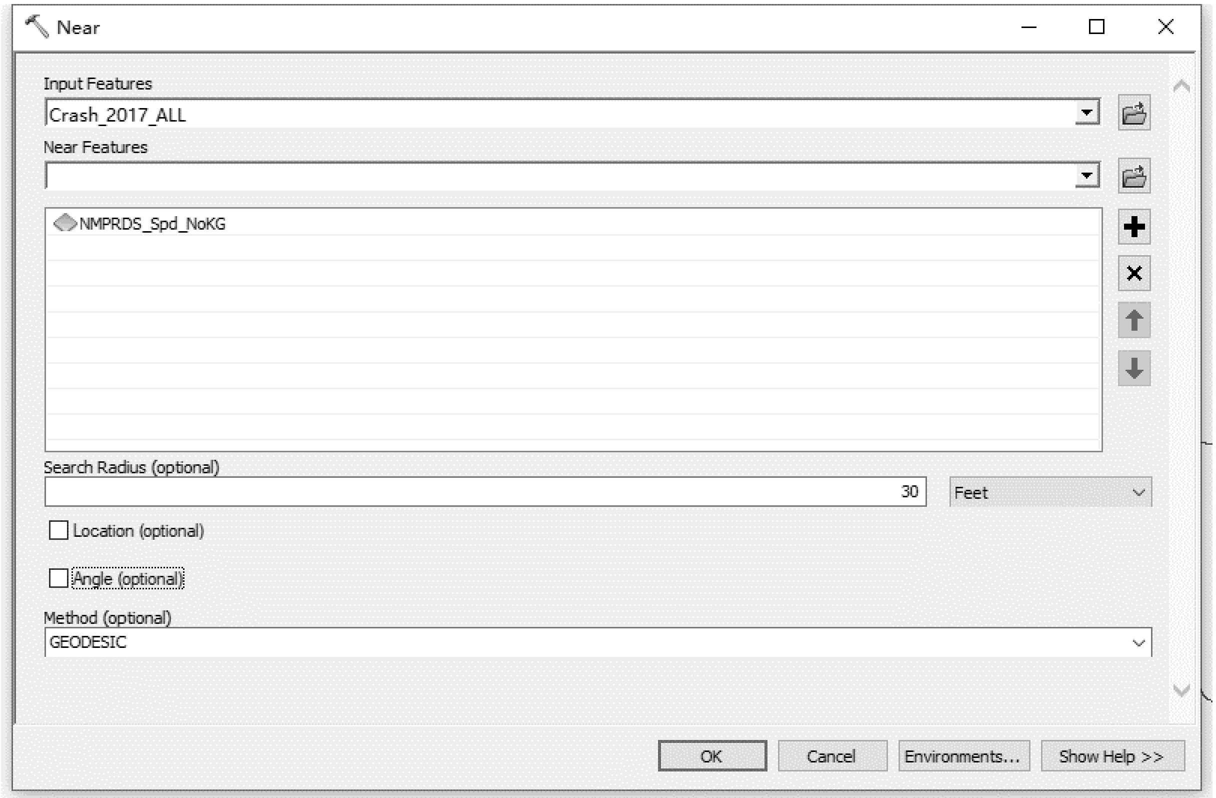


Figure 13. Near (Analysis) Tool.

Table 11. Scales of Crash Severity Levels.

CRASH_SEV_ID	Description	Scale
0	Unknown	O
1	Incapacitating Injury	A
2	Non-Incapacitating	B
3	Possible Injury	C
4	Fatal	K
5	Not Injured	O

Table 12. Crash Summary for Each "Near_FID" (Partially Displayed Example).

NEAR_FID	S17_0	S17_1	M21_3	M21_4	M21_5
1	0	0	0	0	0
4	0	0	0	0	3
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	2
11	0	0	0	0	0
15	0	0	0	0	0
17	0	0	0	0	1
20	0	0	0	0	1
21	0	0	0	2	4
22	0	0	0	0	1
25	0	0	1	1	4

Note: only the first two columns (S17_0 and S17_1) and last three columns (i.e., M21_3, M21_3, and M21_4) are displayed.

To make future analysis easier, additional columns were added based on different statistical methods. All the crashes were categorized as "F" and "I," where "F" is equivalent to severity level 4 in Table 11. The summarization of severity 1, 2, and 3 represents "I." In total, 54 additional columns were added to the data (explained in Table 13).

Table 13. Additional Summarization Columns Explanation.

Code	Description
S17F	S17_4
S17FI	S17_1 +S17_2+S17_3+S17_4
S17All	S17F + S17FI
M17F	M17_4
M17FI	M17_1 +M17_2+M17_3+M17_4
M17All	M17F + M17FI
F17	S17F+ M17F
FI17	S17FI+ M17FI
All17	S17All+ M17All
All21	S21All+ M21All
SV_TF	S17F+ S18F+ S19F+ S20F+ S21F
SV_TFI	S17FI+ S18FI+ S19FI+ S20FI+ S21FI
SV_Tall	S17All+ S18 All + S19 All + S20 All + S21 All
MV_TF	M17F+ M18F+ M19F+ M20F+ M21F
MV_TFI	M17FI+ M18FI+ M19FI+ M20FI+ M21FI
MV_Tall	M17All+ M18 All + M19 All + M20 All + M21 All
F	SV_TF+ MV_TF
FI	SV_TFI+ MV_TFI
All	SV_Tall+ MV_Tall

4.3.5 Final Combination Process

In Part 1, the TTI team created the file for roadway segments with roadway characteristics, and each unique roadway segment was assigned a "unique id." In Part 2, the TTI team conflated the segment dataset with the NPMRDS dataset in four steps. In Part 3, the speed measure was calculated for each "unique_id," then the speed measure file was joined to the roadway characteristics file by "unique_id." In Part 4, the TTI team assigned crash events to each roadway segment. The number of crash events was summarized by crash severity level for each "Near_FID." In this case, "Near_FID" indicates the corresponding roadway segment's row number (starting from 0) in the roadway characteristics file from Part 1. Thus, in order to connect "unique_id" with "Near_FID," the TTI team added a new column, "Near_FID," to the roadway characteristics file. Because in R the row number starts from 1, "Near_FID" equals "row number – 1" in calculations. Then, the crash events file was joined to the roadway characteristics file by "Near_FID.". Finally in Part 5, the TTI team combined all the data processed in the previous parts.

The final dataset can be retrieved from the following link: https://tti-my.sharepoint.com/:f:/g/personal/s-das_tti_tamu_edu/EngdaEI8phJHg_R-N0CDPYkByspNdGCyj-mk2zjkGaHu3g?e=PIlR5Y.

4.4 CHAPTER SUMMARY

This chapter presents the findings of the big data validation and development of a comprehensive dataset for model development. A big data validation case study using Wejo data and the development of a dataset for Texas freeways using roadway inventory, NPMRDS, and CRIS data were undertaken. For the big data validation case study, the TTI team used the CV movement data provided by Wejo to generate variables characterizing the speed variation. Five major steps were taken to prepare the dataset for Texas freeways: the preparation of homogeneous roadway segments, the conflation of the segment dataset with the NPMRDS dataset, the speed measure calculation, the crash and point events assignment, and the final combination process.

CHAPTER 5: NOVEL APPLICATION IDENTIFICATION

5.1 INTRODUCTION

The TTI team searched available state DOT records from 2010–2022 within the Transportation Research Board TRID research database to document state congestion mitigation measures and related efforts to identify general trends in relation to the following:

- Congestion performance measures and related data sources.
- Application of AI strategies (e.g., ML, natural language process) and big data sources within predictive and dynamic traffic demand modeling.
- Use of predictive analytics across operational regimes within real-time, travel time reliability monitoring of nonrecurring congestion.

This review included 108 research reports, guides, plans, special studies, and deployment/development case studies.

5.2 CONGESTION MEASURES

Each state measures certain performance aspects to determine the effectiveness of its traffic congestion reduction efforts. Table 14 shows all congestion measures applied in the review documentation.

Table 14. Congestion Measures.

Delay	Reliability	System Congestion
• Travel Time Index • Travel Time Delay (total or per capita) • Maximum Throughput Travel Time Index (MT3I) • Annual Cost of Vehicle Delay • Annual Hours of Vehicle Delay • Annual Hours of Person Delay • Vehicle Hours of Delay (VHD) • Delay per User • Percent Lane Miles Delayed (<85% posted speed) • Vehicle Delay per Capita	• Planning Time Index (PTI 95) • Reliability Index • TTAve • Median Travel Time (50th percentile) • Buffer Index (BI) • Planning Travel Time • Semivariance Measure • Level of Travel Time Reliability (LOTTR) • Work Zone Impact Ratio • Prediction Interval • Historical Travel Time Ratio (HTTR) • Cumulative Volume Distribution Function • Cumulative Travel Time Distribution Function	• Duration of Congestion • Commute Congestion Cost • Percent of Weekdays When Average Travel Speeds Are Below 36mph • Percent of Freeway Miles Below 45 mph (stop-and-go conditions) in AM or PM Peak • Percent Lane Miles Congested (<70% Posted Speed) • AM and PM Percent of Miles of Directional Congestion • Annual Peak Spreading Factor (Stability of Proportion of 24-Hour Traffic Volume that Occurs During Peak Hours) • 24-hour Volume-to-Capacity Ratio • Average Peak Travel Time • MT3I • Volume/Service Flow Ratio (VSF) • Capacity for Vehicles per Hour • Historical Volume Ratio (HVR) • K-Cluster Travel Time-Based Traffic Profile • K-Cluster Volume-Based Traffic Profile

Figure 14 shows the count of 22 select congestion measures in use out of the 38 identified measures referenced in the novel applications search. Delay is the most common measure of congestion, followed by the travel time index and the buffer time index (BTI).

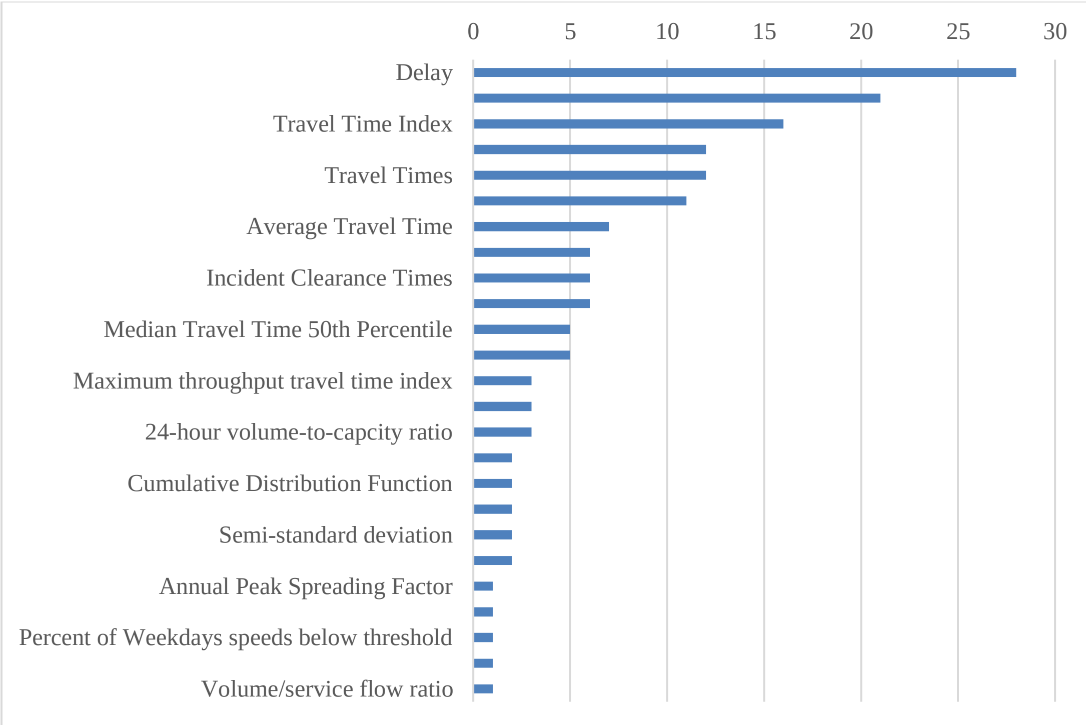


Figure 14. Count of Congestion Measures Identified in Novel Applications Search.

5.3 DATA TYPES

Figure 15 displays the count of data types in use by states for congestion analyses and predictive congestion analytics. The most common factors are volume and speed data, followed by crash and weather.

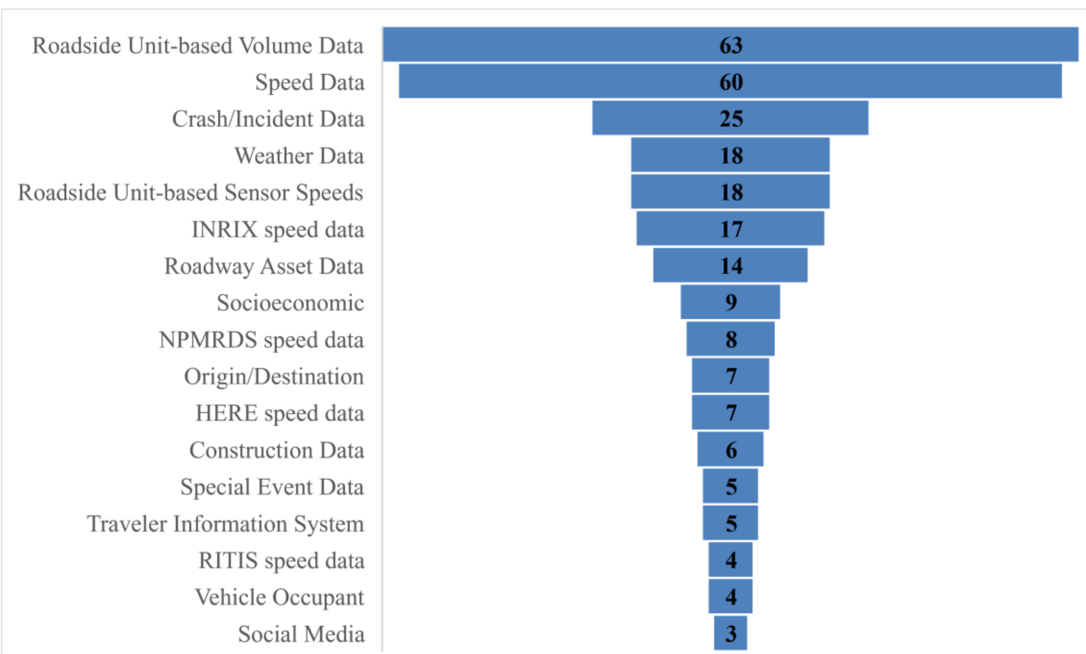


Figure 15. Count of Datasets Identified in Novel Applications Search.

5.4 METHODS AND APPLICATION

Figure 16 identifies the types and counts of analytic methods and applications used on a recurring basis across the various research, case studies, and congestion performance measure programs. K-Means cluster analyses, regression analyses, and random forests (RFs) were the top three types of novel applications applied to congestion measure studies and efforts tied to forecasting congestion related to recurring and nonrecurring sources of congestion. Factors accounted for within these approaches typically included weather, demographics, peak spread, crashes, work zones, and special events. Much of the focus was on achieving a solid balance between linkages of spatial networks with the potential for temporal and semantic information to be included in the predictive capabilities.

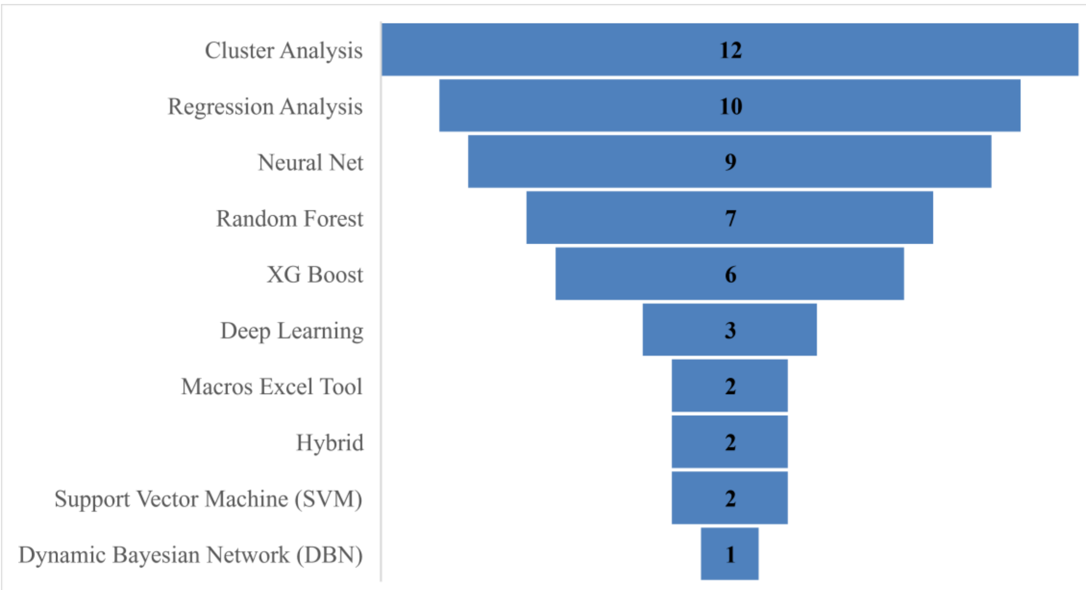


Figure 16. Count of Analysis Methods Identified in Novel Applications Search.

Figure 17 shows the distribution of use of novel applications within state DOT activities related to planning, operations, or both. Planning takes up the majority of novel applications.

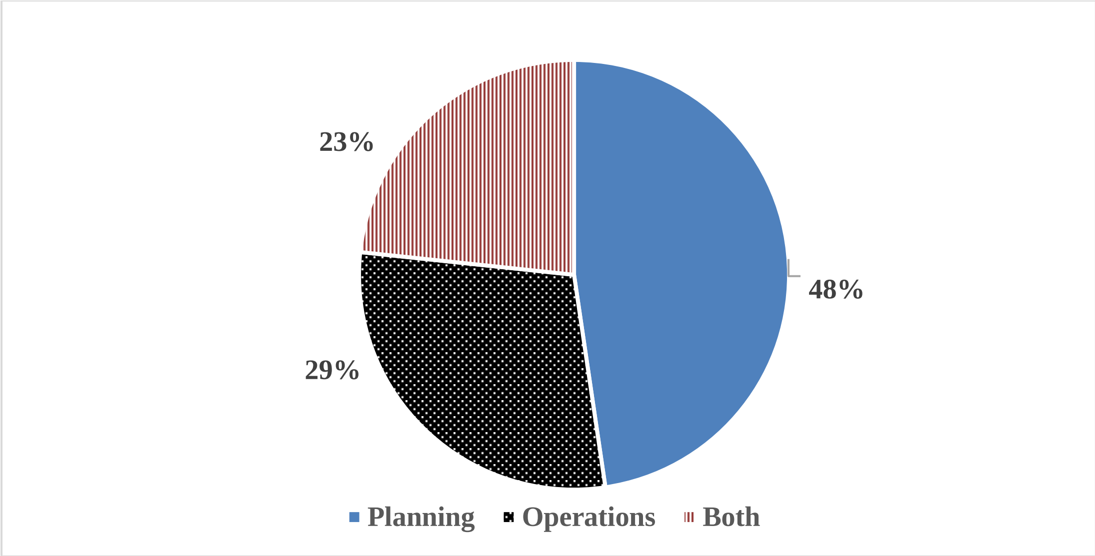


Figure 17. Novel Application of Congestion Predictive Analytics to Planning, Operations, or Both.

Figure 18 indicates the count of activities across states, internationally, and academic findings of a novel applications search involving congestion performance measures, traffic forecasting analytics, and the use of big data in congestion analyses. Leaders in the deployment of novel applications include California, Florida, Minnesota, and international sites predominantly located in China.

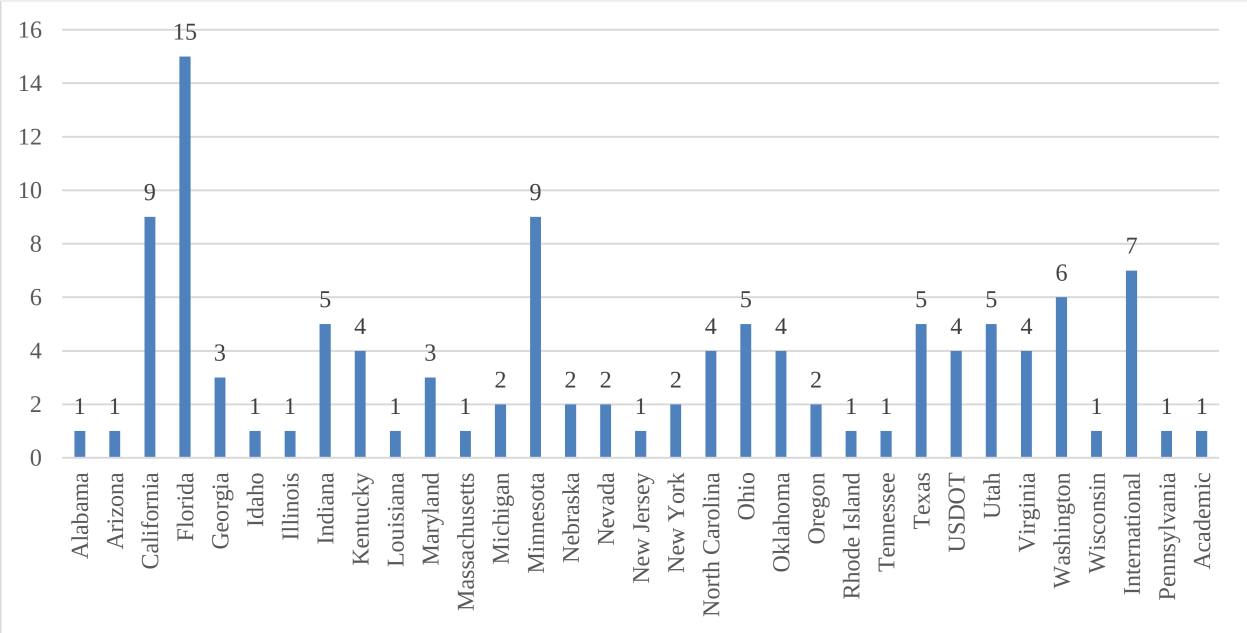


Figure 18. Count of Activities across States, International, and Academic Findings in Novel Applications Search.

5.5 STATE EVALUATION PRACTICES AND TOOLS

The search of TRID database records for state DOT publication sources between 2010 and 2022 resulted in finding important insights identified through searches of keywords and titles that included but were not limited to the following:

- Congestion Performance Indicators.
- Congestion Performance Metrics.
- Traffic Congestion Forecast.
- Big Data.
- Traffic Congestion Forecast and Performance Measures.
- Reliability Measures.
- Traffic Congestion Measures.
- Travel Time Measures.

Eleven states reported formulas and methods for analyzing traffic congestion and applying big data sources and ML/AI-derived analytics to current or future efforts to update their programs. The TTI team focused on states that published sources of research, program activities, congestion reporting analyses, and associated methods and applications. Table 15 lists these states along with the methods and applications used. Overall, the table indicates that travel time and speed-related variables are the most used congestion measures. Some of the studies also utilized reliability-related indices, such as travel time reliability, and PTI as congestion measures. Some studies also integrated weather and roadway-related factors in their congestion analysis. Considering the number of studies, Florida is the leader in the development of novel applications with the highest number of congestion analyses, followed by California and Minnesota. Table 15 also shows that a wide variety of methodologies were applied to conduct the congestion analysis, ranging from traditional statistical models, such as regression and correlation analysis, to more advanced data mining, ML, and AI-based methods, such as XGBoost, CatBoost (CB), and neural networks (NNs).

Table 15. Novel Big Data Applications.

State/ Region/ Agency	Study	Congestion Measures	Method/Application
Alabama	(Isukapati and List, 2016)	(1) Duration of congested hours of speeds at or below 45mph, (2) 70th percentile speed differential, (3) Congestion index (weighted congestion hours normalized by IH length)	Utilized TMC-based probe-speed data (vendor unlisted).
Arizona	(McLellan, 2013)	Peak hour traffic congestion reduction	Utilized regression analysis within a simulation model to predict or forecast congestion levels stemming from new freeway segments.
California	(Bento, 2021)	Travel time	Applied regression analysis.
	(Duvvuri and Pulugurtha, 2021a)	(1) TTAve, (2) PTI, (3) Travel time index, (4) BTI	Applied Pearson correlation coefficient analysis.
	(Kim and Moon, 2022)	(1) Congestion-free travel time in minutes, (2) Congestion delay, (3) Peak congestion delay	Developed a model of trip scheduling under peak congestion to construct California commuters' travel time profiles.
	(Liu and Shetty, 2021)	None (strictly focused on testing large data sets on four different crash prediction methods)	Conducted exploratory analysis, correlation analysis, and feature extraction for the period of the day and location of crashes. Cleaned crash data using the synthetic minority oversampling technique to balance the dataset by duplicating the minority data from the minority population. Tested the predictive capabilities of multiple linear regression, Decision Tree, Random Forest, and XG Boost models on both congestion and crash outcomes. Additionally, developed a framework for the integration of real-time environmental conditions to enhance a smart transportation system's capabilities, enabling integration at the level of a Decision Support System (DSS)..
	(Molan et al., 2020)	Freeway efficiency (ratio of VMT to vehicle hours traveled [VHT])	Calculated the annual average of freeway efficiency using performance measurement system (PeMS) data from 2012 to 2017.
	(Pozdnoukhov, 2018)	Travel times	Application of input-output hidden Markov model (IOHMM) and long short-term memory (LSTM) models drawing on locational data to create activity-based travel demand model.
	(Seeherman and Anderson, 2017)	Flow (volume/5 minutes)	Applied regression analysis.

State/ Region/ Agency	Study	Congestion Measures	Method/Application
	(Varaiya, 2008)	Travel time delay	Developed an algorithm to systematically test and optimize ramp metering strategies in proximity to freeway bottlenecks.
	(Zhang, 2008)	(1) TTAve, (2) Hours of delay	Calibrated the DynaSmart analysis, modeling, and simulation (AMS) model using travel times, speeds, and delay data obtained from the PeMS platform.
Florida	(Brewer et al., 2003a)	(1) Travel speed, (2) Weather	Developed the Final Work Plan to guide the overall sequence of activities and management approach for the model deployment, which identifies all of the organizations involved in the model deployment as well as contractual and other working arrangements in them and addresses systems engineering and software acquisition practices to be followed. It also included a project schedule of all model deployment tasks.
	(Brewer et al., 2003b)	(1) Travel speed, (2) Weather	Developed the Florida model, which can 1) expand and integrate existing data collection and monitoring system; 2) collect and share data; 3) use the data operationally to improve transportation system security, safety, reliability, and performance; and 4) distribute the data to the traveling public.
	(Elefteriadou et al., 2010)	(1) Travel time reliability, (2) Hourly demand, (3) Expected frequency of congestion for each hour	Aimed to (a) refine the previously developed framework and travel time estimation models to evaluate the impacts of ITS applications on travel time reliability, (b) continue to assess the use of field data across Florida to refine the current model, and (c) apply the refined model to the entire strategic intermodal system (SIS) freeway system.
	(Elefteriadou et al., 2013)	Travel time reliability	Developed a CORSIM model estimating more accurate travel time in arterials than previous models.
	(Elefteriadou et al., 2012)	(1) Travel time reliability, (2) Travel time index, 3) PTI	Proposed a framework for considering travel time reliability in a multimodal context. Arterial travel time estimation models were developed using the simulation program CORSIM, considering a total of 1200 scenarios. A spreadsheet similar to that previously developed for freeway sections was developed to obtain travel time reliability measures along arterials based on the results of the developed travel time estimation models.

State/ Region/ Agency	Study	Congestion Measures	Method/Application
	(Elefteriadou and Xu, 2007)	Travel time reliability	Reported travel time reliability for freeways through Florida; developed a preliminary framework for reporting travel time reliability on arterials.
	(Gan et al., 2016)	(1) Excess fatalities, (2) Number of excess injuries, (3) Volume-to-capacity ratio, (4) AADT per lane, (5) Truck volume per lane, (6) Truck percent and delay	Updated the existing performance measures for highway projects and developed visual mapping tools for Congestion Management Process system.
	(Hadi et al., 2016)	(1) Travel time reliability, (2) Tolling price, (3) Travel delay, (4) Queue length	Developed a multi-resolution modeling framework for use in support of agency analyses and modeling of congestion impacts and advanced strategies. Estimated origin-destination demand matrices diversion models due to work zones. Dynamic traffic assignment (DTA) modeling. Mesoscopic and microscopic models.
	(Lawphongpanich and Yin, 2012)	(1) Road pricing, (2) Vehicle miles traveled (VMT) fee	Found nonlinear road pricing by different impacts.
	(Lee and Jin, 2020)	(1) Travel time reliability, (2) market accessibility, (3) intermodal connectivity, (4) ADT, (5) VMT, (6) VHT	Produced a standard benefit-cost analysis and metrics for the wider economic benefits.
	(Moses, 2015)	(1) Peak hour traffic, (2) 99th percentile hourly volume	Analyzed 24-hour peaking relationship to various traffic performance measures. Proposed models to estimate future changes in traffic volumes that are used by transportation planners to determine if the peak period is expected to spread.
	(Stevanovic and Mitrovic, 2019a)	(1) HVR, (2) Travel time index, (3) HTTR, (4) Cumulative volume distribution function, (5) Cumulative travel time distribution function, (6) Volume-based traffic profile (K-Cluster), (7) Travel time-based traffic profile, (8) Volume prediction	Developed a modular framework that operationalizes performance measures using AI and predictive analytics, machine vision, and more. The framework includes ready examples and applies techniques like K-Means clustering to real-time and archived data sources to present Traffic Management Center TMC-relevant performance measures for operational and planning use. It also employs machine learning algorithms such as support vector regression (SVR) and pattern matching to predict travel times and volumes..
	(Washburn and Ko, 2007)	Level of service (LOS)	Found the determinants of LOS perceived by truck mode users and measured their relative

State/ Region/ Agency	Study	Congestion Measures	Method/Application
			importance based on which truck LOS estimation.
Georgia	(Amekudzi-Kennedy et al., 2020)	(1) Travel time reliability, (2) Total delay, (3) Delay per mile, (4) Travel time index, (5) PTI	Identified effective practices for TSMO at the strategic, programmatic, and tactical levels via the TSMO Capability Maturity Model. Developed a tool for calculating transportation system performance metrics. Developed a tool to analyze and report on transportation system performance.
	(Hunter et al., 2012)	(1) Real-time freeway information, (2) Travel time	Integrated real-time data streams with an arterial simulation to support an arterial performance monitoring system.
	(Southworth and Gillett, 2011)	(1) Travel time, (2) Monetary travel cost, (3) Regional accessibility, (4) Average truck speed, (5) Corridor PTI, (6) Corridor BTI, (7) Estimated daily costs of traffic delay in the corridor, (8) Estimated cost of travel time variability in the corridor, (9) A corridor per mile delay cost index, (10) Typical trucks' speed on major truck route connectors	Provided quantitative evidence of how well the system is performing and whether travel conditions have been improving or getting worse over time. Offered useful benchmarks against which the success of the transportation planning process can be assessed and possibly redirected when a particular trajectory needs adjustment.
Idaho	(*Idaho Transportation System: 2008 Performance Report*, 2009)	Cracking index	Provided the reader with an accurate and useful review of the historical and current condition of Idaho's roads, bridges, and railroad crossings, with a goal to eventually provide information on several other facilities, such as pedestrian and bicycle systems, public transit, and congestion.
Illinois	(Sriraj et al., 2017)	Integration of transportation congestion and reliability	Took a broad look at transportation integration by providing an understanding of the different dimensions of integration as defined in the literature, followed by a scan of various best practices of integration/mobility case studies to provide a basis for understanding the significant issues associated with achieving improved mobility through integration.
Indiana	(Achillides and Bullock, 2004)	(1) Travel speed, (2) Traffic volume, (3) Vehicle length	Examined several quality control issues; recommendations for improving construction and configuration procedures were proposed. Documented several simple performance metrics that transportation agencies can use to assess the

State/ Region/ Agency	Study	Congestion Measures	Method/Application
			quality of traffic data and sustain that quality over time.
	(Martchouk et al., 2010a)	(1) Travel time, (2) Travel time reliability	Proposed three models: an autoregressive model to predict individual vehicle travel times on a freeway segment; a duration model to provide insights into how one can predict the probability of a car's duration of time on a roadway segment changing over time; and a seemingly unrelated regression equation model to predict travel time and travel time variability.
	(Ong et al., 2010)	Pavement quality	Investigated the inherent variability of the automated data collection processes and proposed guidelines for an automated data collection quality management program in Indiana.
	(Paleti et al., 2014)	(1) Demand volume, (2) Travel time saving, (3) TTAve, (4) Ramp metering	Implemented the high occupancy vehicle (HOV) lane strategy, which led to improved traffic flow conditions on the HOV lanes; however, it also exacerbated congestion on the general-purpose lanes. The HOV lane strategy was economically unfeasible due to the low HOV volume on these lanes. The reversible lane and ramp metering strategies were found to be economically feasible with positive net present values (NPVs), with the NPV for the reversible lane strategy being the highest.
	(Peeta et al., 2011)	Stability, consistency, and convergence of traffic assignment algorithms	Developed an enhanced transportation planning framework by augmenting the sequential four-step planning process with post-processing techniques.
Kentucky	(Chen, 2010)	(1) Travel time index, (2) PTI, (3) Percentage of travel under congestion (PTC), (4) BI	Calculated travel time index, PTI, PTC, and BI for freeways within the traffic response and incident management assisting the river cities (TRIMARC) coverage. Explored the potential correlation between the reliability measures and incident characteristics.
	(Chen et al., 2015)	(1) Travel time index, (2) PTI, (3) BI, (4) Annual hours of delay, (5) PTC	Evaluated private sector speed data and their use in generating travel time-based performance measures. Created a mechanism to integrate these speed data with networks maintained by Kentucky TC and Kentucky. Met with metropolitan planning organizations to facilitate congestion management and travel model improvement.

State/ Region/ Agency	Study	Congestion Measures	Method/Application
			Generated performance measures, including travel time index, PTI, BI, annual hours of delay, and PTC.
	(Chen and Zhang, 2017)	(1) Travel time index, (2) PTI, (3) BI, (4) Annual hours of delay, (5) PTC	Evaluated datasets at selected locations in previous studies to better understand their characteristics. Further, a procedure was developed to calculate system performance measures established by the Federal Highway Administration's (FHWA's) final rule on system performance measures to assist KYTC in its preparation process.
	(Chen and Zhang, 2015)	(1) VSF, (2) Peak hour congestion	Determined that VSF calculated by the Highway Performance Monitoring System software had limitations that cannot reflect variations in travel time (or speed). The future of VSF as the sole measure of congestion performance should be evaluated.
Louisiana	(Schneider et al., 2019)	A census of all crashes that happened within 5 miles of the work zone (before, within, and after) during active project dates	Provided a review of current work zone crash reporting practices in the United States in general and specifically in Louisiana.
Maryland	(Haghani et al., 2014)	Travel time reliability	Found that some travel time reliability measures are more sensitive to the data source than others, and performance measures for HOV and general purpose lanes must be calculated separately.
	(Kim et al., 2017)	Incident duration	Developed an innovative transferability assessment method that allows the construction of a new system to take advantage of existing Incident Detection and Prediction Models (IDPMs) embedded knowledge in order to circumvent the demanding expertise and efforts needed for data quality control and calibration of IDPMs' prediction rules from many incident records.
Michigan	(Crawford et al., 2011)	(1) Travel demand, (2) Benefit-cost ratio	Created *A Michigan Toolbox for Mitigating Traffic Congestion* to be a useful desk reference for practitioners and an educational tool for elected officials acting through public policy boards to better understand the development, planning, and implementation of congestion mitigation strategies.
	(Kassens-Noor et al., 2022)	(1) Travel time, (2) Travel reliability, (3) Speed volume relationship	Investigated the performance, safety, and public perceptions of the US-23 Flex route.

State/ Region/ Agency	Study	Congestion Measures	Method/Application
Minnesota	(Kwon and Park, 2018)	Travel time reliability	Developed a computerized Travel Time Reliability Measurement System (TTRMS), which can automate the time-consuming process of gathering and managing data from multiple sources and calculating various types of reliability measures under user-specified conditions for given corridors.
	(Kwon et al., 2022)	(1) Travel time reliability, (2) Traffic flow performance, (3) Vulnerability index	Estimated and analyzed the travel time reliability and traffic flow performance trends of the freeway corridors in the Twin Cities metro area of Minnesota.
	(Kwon, 2020a)	Detector health system classified into four classes: healthy, tolerable, impaired, and nonfunctional	Implemented a detector health system as a client-server model in which a relational database became the center of the server. A client software program called detHealth_app.exe was developed that provided retrieval and visualization of various health parameters along with detector or station volumes.
	(Kwon, 2020b)	(1) Vehicle length, (2) Speed, (3) Traffic volume and occupancy	Integrated these classification data into the existing traffic flow analysis (TFA) volume data, which could save cost and time for TFA in the future using existing classification data. The TTI team also integrated the real-time monitoring and control (RTMC) speed data for the locations where it was available.
	(Liao, 2018)	(1) Travel time reliability, (2) Truck delay	Recommended a framework for truck data collection and analysis to better understand the relationships between truck traffic and congestion during rush hours.
	(*Metropolitan Freeway System 2013 Congestion Report*, 2014)	AM and PM percent of miles of directional congestion	Identified locations that are over capacity, project planning, resource allocation (e.g., RTMC equipment and incident management planning), construction zone planning, and department performance measures reporting.
	(*Minnesota Tolling Study Report Modern Tolling Practices and Policy Considerations*, 2018)	(1) Tolling rate, (2) Traffic volume	Analyzed three distinct areas of tolling. Any implementation of tolling requires a balanced approach that aligns the goals and objectives of the project with the needs of the users and the transportation network.
	(Morris and Parikh, 2022)	(1) Travel demand, (2) Traffic volume, (3) Traffic occupancy, (4) Speed	Researched the level of VMT. Looked at the relationship between VMT and congestion at the corridor level to assess the sensitivity of

State/ Region/ Agency	Study	Congestion Measures	Method/Application
			congestion on specific roadways to changes in travel demand.
	(*Twin Cities Metro Freight Initiative: Performance Management Framework*, 2011)	(1) Safety, (2) Mobility, (3) Environmental quality, (4) LOS, (5) Congestion, (6) Travel time reliability, (7) Economic development	Built the Twin Cities Metro Freight Initiative Performance Management Framework to balance competing goals and objectives, (e.g., safety, mobility, environmental quality) and set priorities among alternative actions to resolve freight issues.
Nebraska	(Sharma and Ahsani, 2019)	(1) Travel time delay, (2) Travel time reliability	Identified hotspots, the unusually high-risk zones in a spatiotemporal space containing traffic congestion that occurs on almost all game days. Identified the factors affecting the sizes of hotspots and other parameters. Applied the Dynamic Bayesian Network's approach to forecast the start times and locations of hotspot clusters.
	(Sharma et al., 2016, p. 1)	(1) Travel time delay, (2) Travel time reliability	Developed real-time crash risk prediction models for the studied road section. A sensitivity analysis was conducted on different models and different temporal and spatial windows to estimate/predict crash risk.
Nevada	(Jensen et al., 2019)	Cost savings measure	Developed an AV Feasibility Study Framework by the TTI team to allow NDOT to apply the following decision-making framework not only to this study but also to any future AV roadway studies: (1) Identify potential AV developer partners; (2) Review AV Developer Product Roadmap; (3) Identify mutually beneficial use cases; (4) Determine physical and intelligent transportation system infrastructure needs; (5) Identify suitable Nevada corridor; and (6) Estimate benefits of use cases.
	(Tian et al., 2020)	(1) The attainability of ideal progression, (2) The attainability of user satisfaction	Developed a quality of signal timing performance measure methodology for arterial operation, which may have great potential for enhancing agencies' capabilities of cost-effectively monitoring the quality of arterial signal timing, proactively addressing signal timing issues, and reporting the progress and outcomes in a timely, concise, and intuitive manner.
New Jersey	(Balgowan, 1987)	(1) Forecast accuracy, (2) Forecast reliability, (3) Forecast effectiveness, (4) Expansibility,	Evaluated the accuracy, reliability, effectiveness, expansibility, and additional potential benefits of the moisture, forest, and ice early warning system and indicated that the system is effective

State/ Region/ Agency	Study	Congestion Measures	Method/Application
		(5) Additional potential benefits	as an early warning system. It appears that utilization of the system should result in a reduction of snow and ice-related accidents.
New York	(Fu, 1993)	System reliability	Presented an important-sampling method for the first-order problem of reliability analysis of structural systems, which has a failure domain defined by linear or linearized functions.
	(List et al., 2008)	(1) Traffic delay, (2) Traffic incident	Aimed for the following: 1) The development of New York City input data for the New York City's application of the New York State DOT's delay prediction model (congestion needs analysis model [CNAM]), and 2) Investigation of the published literature identifying models/methods that could improve the CNAM approach for estimating nonrecurring delay.
North Carolina	(Cunningham et al., 2016)	(1) Travel time, (2) Route VHD, (3) Life cycle cost	Produced a sound foundation for evaluating ramp metering outcomes on which the state can build a ramp metering program that is both sustainable and efficient.
	(Pulugurtha et al., 2017)	Travel time reliability	Aimed to 1) review various definitions of travel time reliability proposed by researchers and practitioners in the past, 2) estimate and assess the differences in the performance measures to recommend the most appropriate and viable measures, 3) define and identify travel time reliability thresholds based on additional costs incurred to motorists, and 4) monetize reliability based on the recommended definition to assess the impact of transportation alternatives for use by the North Carolina DOT.
	(Williams et al., 2013)	(1) Travel time, (2) Traffic volume, (3) Traffic mobility, (4) Travel time reliability	Delivered a robust and validated algorithm for estimating route travel times from segment travel times, a decision framework, and accompanying models for estimating (a) volume, (b) VMT, (c) system delay in the absence of direct volume observation, and (d) a preliminary framework for project lifecycle mobility value estimation.
	(Williams et al., 2016)	(1) Freeway service, (2) Benefit-cost	Provided a solution that enables criteria-based selection and prioritization of future incident management assistance patrol (IMAP) system expansion.
Ohio	(Chinnam et al., 2010)	(1) Traffic volume/density/pattern/states (2) Traffic state transitions, (3) Traffic interaction	Proposed three major milestone goals: Milestone #1: Data collection for MI-OH road network structure and historical incident data form MDOT, ODOT, and other agencies using ArcGIS software; Milestone #2: Developing road

State/ Region/ Agency	Study	Congestion Measures	Method/Application
		between links, (4) Traffic incident clearance pattern	network models representative of major freight transportation routes; and Milestone #3: Developing a static re-routing optimization model and implementation for a limited set of scenarios.
	(Coifman and Ponnu, 2018)	(1) Travel speed, (2) Speed index	Pursued several objectives: 1) assessing the performance of the current system, 2) ensuring good performance from a traffic responsive metering system, 3) investigating the operation of the coordinated ramp metering system, and 4) working with ODOT to improve the performance of the ramp metering system.
	(Zwahlen and Oner, 2006)	(1) Interarrival time, (2) Service time, (3) Travel delay	Aimed to improve driver guidance and delineation cues for drivers in work zones to minimize delays and improve safety in the work zone and to thoroughly measure all traffic before and through the work zone.
Oklahoma	(Najumudeen and Bin, 2020)	(1) Speed turbulent occurrences, (2) Traffic flow	Covered several traffic accident-based analyses, methodology of analyses, model training, and model validation and demonstrated the best possible approach for acquiring and preparing necessary data from NPMRDS for use with supervised learning and RNN model training to execute near real-time accident detection.
	(Refai et al., 2017)	Travel time	Presented research detailing the use of the first version of the NPMRDS, which comprised highway vehicle travel times used for computing performance measurements in Oklahoma.
	(Saidi, 2020)	Travel time	Used a dataset from the Oklahoma DOT to compare the accuracy of statistical and machine learning approaches to predicting travel time.
Oregon	(Monsere et al., 2008)	(1) Travel time delay, (2) VMT, (3) Vehicle hours traveled, (4) Ramp queues	Conducted before and after comparisons, comparing pre-timed plans with dynamic smart work zone and road network management (SWARM) ramp metering. Utilized simple queuing theory calculations based on manual timestamps for vehicle departures, which were validated using cameras..
Pennsylvania	(Yao and Qian, 2021)	(1) Travel time index, (2) PTI (95th percentile), (3) Congestion duration	Applied tweet2traffic social computing model to perform neural language modeling, sentiment analysis, and correlation analysis of work and sleep patterns derived from a bounded Pittsburgh region population on Twitter and travel demand on the morning after. Using a clustered learning structure making use of ordered spatiotemporal congestion patterns on freeways, the team

State/ Region/ Agency	Study	Congestion Measures	Method/Application
			forecast morning commute congestion using tweeting profiles extracted by 5 a.m. on the day of, taking tweet congestion peak data from the night prior up to midnight for analysis purposes.
Rhode Island	(“Rhode Island Statewide Model Update,” 2016)	Congestion measures not listed	Integrated state travel demand model with performance measure database through TransCAD.
Tennessee	(Baroud et al., 2021)	Congestion conditions (ratio of Free-Flow Speed in INRIX minus current speed)	Used K-Means Algorithm to group segments into distinct clusters that are generalizable together for associated machine learning models for each cluster. RF with 250 decision trees; NNs using a sequential architecture and fully connected layers (3 hidden layers) and ReLU activation function for all hidden layers. RF and NN outperform naive models, logistic regression, and zero-inflated Poisson regression models.
Texas	(Chaudhary et al., 2018)	(1) Travel time index, (2) Average number of congested peak hours (total congestion counts stop/go traffic durations when speeds are below 50 mph threshold; severe congestion duration counts are made for links or segments below 30 mph), (3) PTI	Used a DTA and microscopic models to analyze benefits of ramp metering in the Dallas area.
	(Lomax et al., 2013a)	(1) Travel time index, (2) Peak period delay, (3) Hours of congested road, (4) BTI	Developed spreadsheet using project-specific calculations of speed and volume to develop congestion measures.
	(Pandey and Juri, 2018a)	(1) BI, (2) PTI, (3) STD, (4) Coefficient of variation	Cleaned and processed NPMRDS data for the San Antonio region using R and Javascript K-means clustering to identify typical days. mean absolute percentage error (MAPE) and mean absolute error (MAE) were used to evaluate the quality of clustering and machine learning applications.
	(Schrank et al., 2019)	(1) Travel delay, (2) Annual PERSON Delay, (3) Annual delay per auto commuter, (4) Travel time index, (5) Commuter stress index, (6) PTI, (7) Time of congestion, (8) Wasted fuel	Conflated highway volume segmentation with INRIX speed segmentation, data cleaning, and congestion measure calculation.
	(Shelton et al., 2020)	(1) Travel time reliability, (2) Average vehicle delay	Developed Mesoscopic model in Houston, El Paso, Austin, El Paso/Juarez, and Dallas-Fort

State/ Region/ Agency	Study	Congestion Measures	Method/Application
		(seconds), (3) Total vehicle delay (hours)	Worth areas using DTA to estimate benefits of freight investments.
Utah	(Liu and Chen, 2017)	(1) Frequency of Congestion, (2) Travel time distribution at 15th, 50th, and 85th percentiles, (3) Vehicle hours traveled, (4) Incident-induced delay	Estimated incident-induced delay through pattern matching of historic incidents, speeds, and volumes, and K-nearest neighbor (KNN) method.
	(Swanson and Culp, 2022)	(1) BTI, (2) Volume/capacity	Filtered nonrecurring congestion to forecast BTI on recurring congestion only for freeways. Regression analysis was used to connect BTI with volume to capacity ratio (V/C).
	(Zhang et al., 2020)	Average speed	Developed hybrid XGBoost, RF, and Artificial Neural Network (ANN) for traffic predictive analytics.
	(Zlatkovic and Zhou, 2017)	Travel time reliability	Used methods and tools to address limitations in construction analysis for pavement rehabilitation strategies QuickZone and VISUM, piecing them together while developing an open-source Google Maps/Google Earth interface for scenario-based traffic simulation analysis of work zones, with a DTA hosting input data from traffic sensors online. It also developed a decision support tool for advanced travel demand management in daily practice.
Virginia	(Dutta and Fontaine, 2020)	Crash frequency	Developed crash prediction models using a negative binomial form and a Zero-Inflated Negative Binomial (ZINB) form through a generalized linear model and by using GLMM models within a "glmmTMB" package that used a template model builder in R statistical software.
	(Fontaine and Miller, 2012)	(1) Mean travel time, (2) Total travel time, (3) Delay, (4) BTI, (5) PTI	Applied microscopic traffic simulation models (VISSIM, PARAMICS) to predict ATM treatment benefits after preliminary sketch analyses indicated ATM strategy is likely to be successful.
	(Miller, 2012)	(1) Annual peak spreading factor (stability of proportion of 24-hour traffic volume that occurs during peak hours), (2) 24-hour volume-to-capacity ratio	Analyzed variance on an entire data set to determine which variables had greatest potential to explain variability in K-factors; an annual K-Factor was developed, and models were developed to forecast a K-factor with simple correlation analysis to confirm model results.
	(X. Zhang et al., 2021a)	(1) LOTTR (50th, 80th, and 90th percentiles), (2) Variance and STD,	Used linear quantile mixed models (LQMM) and gaussian random fields (GRF) models to estimate 50th, 80th, and 90th percentile travel

State/ Region/ Agency	Study	Congestion Measures	Method/Application
		(3) BTI, (4) PTI (95th percentile), (5) Travel time index, (6) Percent on-time travel, (7) Misery index	times to quantify the effects of travel time reliability impact factors and predict reliability measures. All data mapped to TMS and TMC match segments on the network using the Schrank conflation approach through a "Spatial Join" function in ArcGIS.
Washington	(Hallenbeck et al., 2015)	(1) Average hourly freight delay, (2) Travel time reliability, (3) average hourly vehicle delay, (4) Frequency of congestion	Compared speeds from WSDOT sensors to NPMRDS to evaluate accuracy and identify issues.
	(Hendricks et al., 2016)	Maximum throughput	Used highway segment analysis program that contains a mobility analysis screening tool in macros Excel spreadsheet format that computes segment performance to determine whether a travel demand management strategy should be prioritized because it delivers congestion reduction.
	(Wang et al., 2010)	(1) Average incident duration (in minutes), (2) Incident-induced delay	Applied deterministic queuing theory to predict impacts upstream from incidents downstream.
	(Wang et al., 2016)	(1) Throughput productivity, (2) Maximum throughput, (3) Duration of congestion, (4) Commute congestion cost, (5) Hours of travel delay, (6) Annual hours of vehicle delay, (7) Annual cost of vehicle delay, (8) Median travel time (50th percentile), (9) TTAve, (10) MT3I, 11) Percent of weekdays when average travel speeds are below 36mph	Used DRIVE-Net web-based data fusion-driven archive that functions as a real-time DSS with digital roadway mapping element that can perform travel time reliability estimation, predictive modeling, and visualization functions.
	(*The 2010 Congestion Report*, 2010)	(1) Lane miles of state highway system congested (70% of posted speed), (2) Percent of state highway system congested, (3) Total VHD (times when speeds drop below 85% of posted speed), (4) Average peak travel time, (5) 95% Reliable travel time, (6) MT3I, (7) Percent days speeds at/below 35mph, (8) Vehicle	Analyzed Statewide planning data.

State/ Region/ Agency	Study	Congestion Measures	Method/Application
		throughput, (9) Lost throughput productivity, (10) Average incident clearance time	
Wisconsin	(Srivastava et al., 2018a)	(1) Travel time, (2) Reliability rating (% of trips at < 1.33 freeways), (3) PTI, (4) Travel time index (80th percentile), (5) Semistandard deviation, (6) Percentage trips with space mean speed less than 50, 45 and/or 30mph, (7) STD, (8) Misery index	Developed predictive models using regression modeling to estimate work zone reliability measures based on the baseline reliability
USDOT	(Fan and Chen, n.d.)	TTAve	Applied XG Boost to predict travel time and compare against ground truth travel times.
	(Qiu and Fan, 2021)	Travel times	Tested XG Boost and RF methods on traffic forecasting.
	(*Establishing Monitoring Programs for Travel Time Reliability*, 2014)	(1) Travel time index, (2) PTI, 80th and 95th percentile, (3) SV	Integrated variables from nonrecurring event data into the segment and route-level travel time data to determine effects on speeds and travel times. Used the Monte Carlo model with incidence matrices to estimate probability distribution functions for various time periods of the day over 1 year to determine (a) the extent of travel time rate variances and impact of congestion when no nonrecurring event exists, (b) the impact of nonrecurring events, and (c) the consistency that exists within observations for the same operating condition.
	(Sobolewski et al., 2014)	(1) Travel time, (2) VMT, (3) Delayed Vehicle Hours	Reconfigured various data types into a macros-enabled Excel spreadsheet to fit the 5-minute bin recommendation. The analysis tool referenced the TTRMS database with all the associated records. It is a macro-enabled spreadsheet that produces basic travel time reliability measures such as cumulative density function curves, reliability indices, and more.
International	(Cao et al., 2022)	(1) Average speed, (2) Average occupancy, (3) Total volume	Compared multi-head self-attention spatiotemporal graph convolutional network (MSASGCN) to other models using MAE, root mean squared error (RMSE), and MAPE to estimate potential for forecasting/predicting traffic flow and congestion. PyTorch deep learning framework applied in MSASGCN model development.

State/ Region/ Agency	Study	Congestion Measures	Method/Application
	(Petalas et al., 2017)	Delay	Fused multiple forecast models, including RF, autoregressive integrated moving average (ARIMA), and a social data miner using natural language processing (NLP)-based architecture using linear regression experiments within a singular traffic forecasting application.
	(Yang et al., 2022)	Traffic speed	Applied spatiotemporal deepwalk gated recurrent neural network (ST-DWGRU) model that 1) incorporates non-Euclidean topological relationship of the road network through clustering or convolutional neural network (CNN) to explore spatial correlation and temporal correlation; 2) incorporates semantic information such as factoring in roadway incidents; and 3) does it all simultaneously for real-time use in operations and planning.
	(Zhang et al., 2022)	Travel time	Applied adaptive graph learning algorithm (AdapGL) to NNs to acquire spatial dependencies between PeMS sensors through a novel parameterized graph learning module for improved travel time forecasting.
	(Zhou et al., 2019)	Average speed	Applied spatial-temporal deep tensor NN to mitigate the influence of manually stacking traffic detectors on the prediction results and to open up traffic forecasting to probe data inputs.
	(Kang et al., 2020)	TTAve	Applied a hybrid EDMCN-XG Boost model to predict travel times on a roadway network in Guiyang, China.
	(Liu and Tan, 2021)	(1) Traffic speed, (2) Traffic flow	Surveyed GNN applications, data needs, data sources, and measures suitable for forecasting volume, speed, and crash analytics.

5.6 NOVEL APPLICATION CASE STUDY: COMPUTER VISION

The TTI team, in collaboration with InfraSix, developed a comprehensive plan for extracting highly accurate traffic flow data based on ML and AI. Initially, InfraSix conducted a proof-of-concept study to develop and test a fully functional ML algorithm based on the YOLO V5 architecture to perform vehicle detection, tracking, and counting from closed-circuit television (CCTV) traffic cameras available from open data sources. The main purpose of this proof-of-concept was to demonstrate the feasibility of utilizing CCTV camera footage to train and test ML algorithms and then to accurately detect multiple vehicle classifications, including cars, motorbikes, trucks, buses, etc. Each of the detected vehicles were then tracked with a unique classification ID to get accurate counts of each. Additionally, an Excel file in CSV format was exported once the algorithm had been run on CCTV video footage, which contains the timestamp, vehicle ID, and confidence of the detections from the video. The main purpose of this

CSV file is to make it possible to utilize the ML algorithm results as input parameters for calculating traffic counts, speed, headroom, and related traffic identifiers. Using these metrics, numerous traffic-related variables can be introduced to explain and predict traffic patterns. This information can then be used by the DOT for decision-making regarding both short-term and long-term solutions or provisions. Note that the proof-of-concept study can be extended to develop traffic algorithms that can generate and optimize highly reliable traffic data. Using this database, a complete series of both expected and unexpected variables can be run against the traffic data to provide an incredibly rich view of traffic patterns that can be used to conduct the following tasks for TxDOT:

- Selecting and evaluating best data sources and preparing them to be "AI Ready."
- Gathering real-time CCTV camera analytics from multiple cameras located statewide.
- Extracting useful analytics and metrics from the ML and computer vision based model outputs from sources such as traffic cams and data from other third-party sensors and sources.
- Developing an interactive, user-friendly GIS-based dashboard to visualize and work with collected and analyzed data.
- Post-processing analyzed results to predict choke points in traffic based on present video/imagery input and build a database to support predictive analytics in the long term to enable efficient highway planning practices.

Figure 19 presents screenshots from the working system. As can be seen from the figure, the system can detect, count, and classify vehicles with good accuracy.

Figure 19. Vehicle Counts Using Computer Vision.

The first step in identifying and counting vehicles was a proof-of-concept based on ML to recognize vehicles, identify the vehicle class, and provide a micro perspective of traffic flow. The next step of this project was to provide a clear and quantifiable understanding of cause and effect and look at the impact of variables on traffic patterns on a larger or macro scale by potentially using thousands of cameras across the state. These cause/effect scenarios will be highly beneficial to TxDOT for planning, prioritizing, and decision-making to improve the safety, mobility, and operation of Texas highways. Note that the project can be extended to determine and analyze the following variables:

- Vehicle speed.
- Average speed.
- Vehicle headway.

In addition, the extracted data from the newly developed algorithm can be integrated with the following variables related to roadway and weather conditions, driver demographics, crash events, driver distraction, and roadway geometry:

- Day of week (weekdays, weekends, etc.).
- Time of day (peak hour, off-peak hour, etc.).
- After a vehicle crash (fender bender to multiple units assisting).
- After an immobilized vehicle stops (flat tire, out of gas, etc.).
- Stopped vehicles inside or outside or in the traffic lanes.
- Range of weather conditions (sunny, cloudy, rain, snow, fog, etc.).
- Range of lighting conditions (daytime, dawn, dusk, nighttime, etc.).
- Road condition, closed lanes, cones, lane switches, presence of construction equipment.
- Number of lanes, merging traffic.
- Use of cell phones, texting, and navigation.

A complete and accurate understanding of traffic counts, speed, and headway is critical for an in-depth understanding of the causal effect of crashes. As a proof-of-concept, the TTI team developed a working system for collecting traffic data based on the ML algorithm. The developed algorithm will be further enhanced to improve the accuracy and to make it capable of proving information in real time. There is potential in building a platform for conducting the actual cause/effect research. The platform development can contain writing scripts (code), integration, and testing, and then the workflow can be established so that the TxDOT traffic analysis team has all of the tools and data needed to run the analytics. The platform can be used to identify the largest variations in traffic flow, determine which cameras expose the biggest issues, isolate the variables, and determine a single or multivariable cause and effect.

5.7 CHAPTER SUMMARY

Chapter 5 focused on identifying trends in congestion performance measures and related data sources, the application of AI strategies and big data sources within predictive and dynamic traffic demand modeling, and the use of predictive analytics across operational regimes within real-time, travel time reliability monitoring of nonrecurring congestion. The literature search showed that the most common congestion measure was delay. It was also found that volume and speed data are the most common data sources used, and K-Means cluster analyses, regression analyses, and CB were the top three types of novel applications applied to congestion measure studies. Planning was shown to be most novel application within state DOT activities. In addition, leaders in the deployment of novel applications include California, Florida, Minnesota, and international sites that are predominantly located in China. To extract accurate traffic flow data, the TTI team, in collaboration with InfraSix, developed a computer-vision-based novel approach.

CHAPTER 6: KEY PERFORMANCE METRICS

6.1 INTRODUCTION

The development of congestion performance measures, especially for freeways, has become a major area of interest for transportation and planning agencies due to the demands of the public and state legislation. Federal transportation authorization legislation is emphasizing freeway performance monitoring, especially with regard to system operation and management. However, the development of appropriate congestion measures that can truly benefit practitioners, agencies, and road users faces many challenges. Therefore, these measures are not being integrated into the transportation decision-making process and have not yet become standard practices. Many transportation professionals have been using the *Highway Capacity Manual* (HCM) for defining the quality of traffic flow largely tied to the LOS concept. However, the LOS may not necessarily measure the nature and extent of the congestion. Therefore, it is essential to develop more detailed and meaningful congestion performance measures for freeways than HCM-based levels of service to describe the effects of operational strategies. Keeping these research needs in mind, this study suggests several congestion performance measures and reports on associated analysis procedures.

6.2 TYPE OF PERFORMANCE MEASURE USED BY AGENCIES

Transportation agencies use both outcome and output measures to define congestion performance measures. Outcome measures are defined based on the physical quantity of the items, the scale or scope of activities, and the efficacy of converting resources into usable products, whereas output measures are defined based on the quality of service the transportation agencies provide to transportation users. The operating and planning agencies generally apply derivatives of speed, travel time, and delay to develop outcome performance measures. For instance, the travel time index is a common performance measure adopted by the agencies. Although the use of the LOS as a freeway performance measure is decreasing, many operating and planning agencies still use this matrix. However, performance measures related to reliability have become more popular in recent years. Although these metrics are usually formulated for short segments or at key locations, many agencies are leveraging and extending these concepts for multiple freeway routes of extended lengths. Output performance measures are used mainly by operating agencies for the operation of field equipment, such as sensors and cameras, and for activities related to incident management.

6.3 BASIC PRINCIPLES OF FREEWAY PERFORMANCE MEASUREMENT

In order to develop effective performance measures appropriate for a particular freeway, some basic principles need to be followed by transportation and planning agencies. The NCHRP Project 3-68 has reported several basic principles for the successful development of freeway performance measures, as listed in Table 16 (Margiotta et al., 2007).

Table 16. Basic Principles for Freeway Performance Monitoring.

Principle	Description
Principle 1	Mobility performance measures must be based on the measurement or estimation of travel time.
Principle 2	Measure where you can—model everything else.
Principle 3	Multiple metrics should be used to report freeway performance, especially for mobility.
Principle 4	Traditional HCM-based performance measures for mobility (i.e., V/C and LOS) should not be ignored but should serve as supplementary, not primary measures of performance in most cases.
Principle 5	Both vehicle- and person-based performance measures of throughput are useful and should be developed, depending on the application.
Principle 6	Both quality-of-service (outcome) and activity-based (output) performance measures are required for freeway performance monitoring.
Principle 7	Activity-based (output) measures should be chosen so that improvements in them can be linked to improvements in quality-of-service measures.
Principle 8	Customer satisfaction measures should be included with quality-of-service measures for monitoring freeway performance.
Principle 9	The measurement of travel time reliability is a key aspect of freeway performance measurement, and reliability measures should be developed and applied.
Principle 10	Three dimensions of freeway mobility/congestion should be tracked with mobility performance measures: source of congestion, temporal aspects, and spatial detail.
Principle 11	Communication of freeway performance measurement should be done with graphics that resonate with a variety of technical and nontechnical audiences.
Principle 12	Continuity should be maintained in performance measures across applications and time horizons; the same performance measures should be used for trend monitoring, project design, forecasting, and evaluations.

6.4 RECOMMENDED FREEWAY PERFORMANCE MEASURES

The TTI team conducted a comprehensive literature review to determine the best possible freeway congestion performance measures for TxDOT to develop appropriate strategies for optimum corridor management. It was found that most of the measures are derivatives of speed and travel time. For instance, TxDOT has developed a simple but robust metric called the Texas congestion index (TCI), which indicates the amount of additional time a road user needs for a particular trip (Texas A&M Transportation Institute, 2020). TCI is defined as the ratio of the actual travel time for a trip to the time required to complete the same trip in free-flow conditions. The free-flow travel time can be calculated using the 85th percentile speed at night when there is no obstruction in flow due to congestion. The PTI is another effective congestion measure reported by TxDOT (Texas A&M Transportation Institute, 2020). The PTI represents the reliability of travel time and indicates the total travel time that needs to be planned for a trip. The PTI is defined as the ratio of the 95th percentile travel time to the free-flow travel time. For example, a PTI value of 2 indicates that a trip that requires 20 minutes in light traffic would need 40 minutes in current conditions. From a system manager's perspective, PTI indicates how much worse a system can perform relative to the free flow. This measure can be beneficial for individual commuters as well as for truck drivers when planning urgent trips.

Another interesting performance matrix is the peak spreading K-factor reported by the Virginia DOT (VDOT) (Miller, 2022). The K-factor is defined as the proportion of the 24-hour traffic volume that occurs during the peak hour. With an increase in traffic, the K-factors may decrease. Motorists may change their travel departure times to slightly before or after the peak period in response to increasing traffic congestion, and this behavioral response is known as peak spreading. The peak spreading K-factor can be effectively leveraged by practitioners to estimate travel demand and resulting transportation performance.

WSDOT has developed several congestion performance matrices, including average peak travel time, MT3I, and duration of congestion. The average peak travel time of a corridor is defined as the mean of travel time during the peak 5-minute intervals for all weekdays of a whole year. The MT3I is the ratio of the average peak travel time to the maximum throughput speed travel time. The duration of congestion is calculated by summing up all the travel time during which the speed of that corridor dropped below 45 mph.

In recent years, freeway congestion performance measures based on the concept of reliability have been growing in importance. Travel time reliability indicates the level of consistency in travel over time and is calculated by observing the trend in travel time over a significant amount of time. This measure is beneficial for planning trips and selecting routes effectively. Travel time reliability can be affected by either recurrent or nonrecurrent variability or congestion. Although the main cause of recurrent congestion is insufficient capacity, nonrecurrent congestion mainly occurs due to unexpected traffic events, such as accidents, inclement weather, and construction. Many simple yet meaningful performance measures to define travel time reliability representing both recurrent and nonrecurrent variability can be found in the literature. For example, the Indiana DOT reported two measures, namely travel window and percent variation, as measures of travel time reliability (Martchouk et al., 2010b). The travel window provides a range of travel time that might be required to complete a trip and is calculated by adding and subtracting the STD from the TTAve. Similarly, the percent variation is defined by the ratio of TTAve to the STD. These performance measures are beneficial for professionals in communicating the extent of reliability; however, they might not be useful to individual travelers because of the challenges they might face in applying the concept to their own travel time. Another interesting form of reliability is the BI, which indicates the amount of additional time that a traveler needs to consider on top of their usual travel time to complete a trip on time. To determine BI, first the difference between the 95th percentile travel time and the mean travel time is calculated, and then the resulting difference is divided by the mean travel time. A single value of BI is usually reported for a corridor to represent the overall corridor reliability. In addition, the variation of BI during the times of day can provide important insights for improving corridor operations. A list of performance measures is reported in Table 17, and the recommended measures are highlighted in bold. Many agencies have developed more sophisticated performance measures to monitor freeway congestion, as listed in Appendix A. Some of these measures require more detailed data resolution and continuous surveillance and planning data; however, they might not necessarily provide additional value compared to the measures reported in Table 17.

Table 17. Recommended Congestion Performance Measures and Associated Formulas.

Performance Metric	Equation	Source
TCI	$Texas\ Congestion\ Index = \frac{Actual\ Travel\ Time\ (minutes)}{Free{-}Flow\ Travel\ Time\ (minutes)}$	TxDOT (Texas A&M Transportation Institute, 2020)
PTI	$Planning\ Time\ Index = \frac{95th\ Percentile\ Travel\ Time\ (minutes)}{Free{-}Flow\ Travel\ Time\ (minutes)}$	TxDOT (Texas A&M Transportation Institute, 2020)
Peak Spreading K-Factor	$K - factor = \frac{Kth\ highest\ volume}{AADT} \times 100\%$ K-factor refers to the stability of the proportion of 24-hour traffic volume that occurs during peak hours	Virginia DOT (Miller, 2022)
Average Peak Travel Time	$Average\ peak\ travel\ time = \frac{Trip\ length}{Average\ speed_{5-min}}$	Washington State DOT (Wang et al., 2013)
MT3I	$Maximum\ throughput\ travel\ time\ index (MT^3I) = \frac{Average\ travel\ time\ _{Peak\ 5-min}}{Travel\ time\ _{Max.throughput\ speed}}$	Washington State DOT (Wang et al., 2013)
Duration of Congestion	$Duration\ of\ congestion = \sum (Time\ for\ all\ intervals\ with\ speeds < 45\ mph)$	Washington State DOT (Wang et al., 2013)
Statistical Indices of Reliability	$Travel\ Window = Average\ Travel\ Time \pm Standard\ Deviation$ $Percent\ Variation\ of\ Travel\ Time\ (PVTT) = \frac{Standard\ Deviation}{Mean} \times 100\%$ *Different speed measures:* SpdAve, SpdStds, and Spd85s.	Indiana DOT (Martchouk et al., 2010b)
BI	$Buffer\ Index\ (BI) = \frac{TT_{95\%} - T_{Mean}}{TT_{Mean}}$	TxDOT (Pandey and Juri, 2018b)

6.5 CHAPTER SUMMARY

Chapter 6 reviewed key performance metrics used by different agencies. It covered types of performance measures used by different agencies, including output and outcome measures and the basic principles of freeway performance measurement. Finally, the recommended freeway performance measures were discussed, including the TCI, the PTI, the peak spreading K-factor, and travel time reliability.

CHAPTER 7: FORECASTING USING ARTIFICIAL INTELLIGENCE

7.1 DATA DESCRIPTION

The present study aimed to forecast suitable congestion measures within Texas. The dataset utilized for the analysis comprises a total of 28,684 rows of freeway segments. Out of these segments, 23,318 segments, or 81.28 percent of total data, have valid congestion-related information recorded for the respective road segments. Conversely, there is a lack of congestion-related information for the remaining 18.72 percent of the road segments within the dataset.

7.1.1 Congestion Performance Measures

The current study focused on four congestion performance measures in the analysis of freeway segment speeds within Texas: SpdAve, SpdStd, Spd85, and percent variation of travel time (PVTT). The speed information utilized for this analysis was collected over a period of 5 years (2017–2021). Descriptive statistics for these congestion measures are presented in Table 18, which differentiates between rural and urban freeway segments and various PSLs of 65 mph, 70 mph, 75 mph, and an ALL category, which represents all PSLs. The third column of the table indicates the number of samples under each land use (i.e., rural and urban) and PSL. For each land use and PSL combination, the table presents descriptive statistics for the four congestion measures. The statistics include the minimum and maximum values, the mean, and the STD. For example, in the rural category with a PSL of 65, the sample size is 49. The SpdAve values for this sample range from 57.576 to 66.404, with a mean of 64.122 and a STD of 3.112. The SpdStd values range from 3.698 to 7.185, with a mean of 5.844 and a STD of 0.907, and the Spd85 values range from 61.152 to 70.972, with a mean of 68.87 and a STD of 3.067. In this chapter, predictive models will be developed to estimate road segment speeds based on land use, and more granular models will be developed that consider different PSLs.

Table 18. Sample Descriptive Statistics for Dependent Variables.

Land Use	PSL (mph)	Sample Size	Dependent Variables	Min	Max	Mean	STD
Rural	65	49	SpdAve	57.576	66.404	64.122	3.112
			SpdStd	3.698	7.185	5.844	0.907
			Spd85	61.152	70.972	68.87	3.067
			PVTT	1.329	85.440	23.437	25.012
	70	4502	SpdAve	60.76	69.727	67.411	1.222
			SpdStd	1.769	8.819	3.976	1.108
			Spd85	65.7	74.21	70.39	1.247
			PVTT	0.503	99.978	24.651	24.368
	75	5470	SpdAve	61.539	71.904	69.028	1.077
			SpdStd	1.834	8.708	4.586	0.954
			Spd85	67.678	77	73.2	1.762
			PVTT	0.701	99.796	24.666	24.510
	ALL	10037	SpdAve	38.055	71.904	68.249	1.616
			SpdStd	1.769	8.819	4.32	1.076
			Spd85	44.49	77	71.89	2.212
			PVTT	0.503	99.978	24.569	24.386
Urban	65	1300	SpdAve	39.764	65.363	55.713	5.311
			SpdStd	2.257	20.447	8.09	3.354
			Spd85	53	70.5	62.4	2.657
			PVTT	2.329	93.299	31.937	22.882
	70	8700	SpdAve	37.937	69.128	62.641	3.818
			SpdStd	1.807	20.62	6.575	2.73
			Spd85	57.757	76	67.453	2.096
			PVTT	0.567	99.308	31.269	22.995
	75	3164	SpdAve	48.262	70.934	65.361	3.487
			SpdStd	2.45	18.224	5.922	2.149
			Spd85	61	77.851	70.439	2.18
			PVTT	1.555	99.955	31.086	24.832
	ALL	13281	SpdAve	16.319	70.934	62.488	4.949
			SpdStd	1.807	20.62	6.578	2.743
			Spd85	29.5	77.851	67.575	3.289
			PVTT	0.567	99.955	31.168	23.458

Note: STD = standard deviation.

7.1.2 Independent Variables

Table 19 provides descriptive statistics for a set of independent variables of freeway segments that are used to predict congestion measures. The data conflation steps were detailed in the previous deliverables. This study utilized multi-source data (data from NPMRDS, INRIX, and Wejo) to collect congestion-related information. After doing a variable importance analysis, the following key variables were considered for the AI modeling:

- Number of lanes (num_lanes), which ranges from 1 to 4 lanes with a mean of 2.292 lanes and a STD of 0.71 lanes across all segments.
- Jobs within 45 minutes auto travel time, time decay (network travel time) (d5ar), which ranges from 0 to 900 jobs, with a mean of 350.401 and a STD of 93.389.
- AADT (adt_adj), which ranges from 8 to 117,230, with a mean of 25,218.638 and a STD of 15,828.853 across all segments.
- Median width (med_wid), which ranges from 0 to 455 ft, with a mean of 52.566 ft and a STD of 35.441 ft across all segments.
- K-factor (k_fac), which ranges from 4.9 to 26.7, with a mean of 9.581 and a STD of 1.717 across all segments.
- Length of the section (len_sec), which ranges from 0.004 to 10.408 miles, with a mean of 0.843 miles and a STD of 1.31 miles across all segments.
- Daily vehicle miles traveled (dvmt), which ranges from 0.288 to 459,376.1, with a mean of 19,537.897 and a STD of 35,184.667 across all segments.
- Total number of crashes over a 5-year period (five_year_crashes), which ranges from 0 to 249, with a mean of 5.065 and a STD of 10.257 across all segments.

The table also separates the statistics for each variable based on the PSL of the road segments. Of the segments, 4,502 segments have a PSL of 70 mph, and 5,470 segments have a PSL of 75 mph. The mean and STD of the variables differ between different PSLs.

Table 19. Variable Statistics of Rural Roads.

PSL (mph)	Sample Size	Independent Variable	Min	Max	Mean	STD
65	49	med_wid	4	52	31.143	11.993
		num_lanes	2	4	2.245	0.656
		d5ar	250	300	278.694	22.75
		adt_adj	15255	63802	51054.98	19933.411
		k_fac	7.3	14	9.79	1.466
		len_sec	0.017	3.693	0.96	1.141
		dvmt	494.824	144849.3	39955.265	46480.154
		five_year_crashes	0	89	11.571	17.272
70	4502	med_wid	0	455	48.449	36.15
		num_lanes	1	4	2.23	0.547
		d5ar	0	900	349.665	88.314
		adt_adj	8	117230	32706.692	16815.601
		k_fac	5.1	25	9.627	1.519
		len_sec	0.005	10.408	0.864	1.361
		dvmt	1.276	459376.1	25661.631	44852.951
		five_year_crashes	0	249	6.525	12.768
75	5470	med_wid	0	455	56.173	34.592
		num_lanes	1	4	2.343	0.818
		d5ar	0	785	351.795	97.588
		adt_adj	10	79854	18846.172	11426.181
		k_fac	4.9	26.7	9.536	1.862
		len_sec	0.004	9.38	0.826	1.269
		dvmt	0.288	238098.7	14342.768	23115.816
		five_year_crashes	0	108	3.807	7.229
ALL	10037	med_wid	0	455	52.566	35.441
		num_lanes	1	4	2.292	0.71
		d5ar	0	900	350.401	93.389
		adt_adj	8	117230	25218.638	15828.853
		k_fac	4.9	26.7	9.581	1.717
		len_sec	0.004	10.408	0.843	1.31
		dvmt	0.288	459376.1	19537.897	35184.667
		five_year_crashes	0	249	5.065	10.257

Table 20 provides the independent variable descriptive statistics of all urban roads. Of the 13,281 urban road segments, 87,000 roads have a PSL equal to 70 mph. For example, for urban roads with a PSL of 65, the median width (med_wid) ranges from 0 to 350, with a mean value of 16.685 and a STD of 20.975. The number of lanes (num_lanes) ranges from 1 to 6, with a mean value of 2.967 and a STD of 1.107. Similarly, for each variable, the range (minimum and maximum), mean, and STD values are displayed in the table.

Table 20. Variable Statistics of Urban Roads.

PSL (mph)	Sample Size	Variable	Min	Max	Mean	STD
65	1300	med_wid	0	350	16.685	20.975
		num_lanes	1	6	2.967	1.107
		d5ar	0	750	295.08	120.899
		adt_adj	85	240182	104736.288	59490.065
		k_fac	6.6	25.5	10.239	1.258
		len_sec	0.007	1.908	0.223	0.262
		dvmt	15.98	303772.7	21276.402	30313.316
		five_year_crashes	0	674	37.376	70.839
70	8700	med_wid	0	350	23.836	27.188
		num_lanes	1	7	2.958	1.055
		d5ar	0	800	315.425	115.208
		adt_adj	85	272758	91240.487	56909.504
		k_fac	6.1	32.3	10.108	1.089
		len_sec	0.003	3.703	0.306	0.413
		dvmt	2.168	558809.8	25169.835	40257.738
		five_year_crashes	0	1567	29.422	62.39
75	3164	med_wid	0	350	31.662	29.083
		num_lanes	1	7	2.803	1.073
		d5ar	0	800	322.594	118.948
		adt_adj	33	272758	73607.344	63419.184
		k_fac	6.1	23.7	9.95	1.362
		len_sec	0.006	3.733	0.35	0.482
		dvmt	5.39	558809.8	22176.565	37984.209
		five_year_crashes	0	960	27.216	69.228
ALL	13281	med_wid	0	350	24.917	27.535
		num_lanes	1	7	2.919	1.065
		d5ar	0	800	314.798	116.9
		adt_adj	33	272758	87998.081	59419.215
		k_fac	6.1	32.3	10.085	1.18
		len_sec	0.003	3.733	0.308	0.419
		dvmt	2.168	558809.8	23925.894	38739.149
		five_year_crashes	0	1567	29.583	64.802

7.2 METHODOLOGY

7.2.1 Artificial Intelligence Algorithms

7.2.1.1. Random Forest (RF)

RF, a supervised ML algorithm, utilizes a bootstrap sampling method to generate a random number of samples from the original training data to produce new training sample sets by building a decision tree that is developed on the new training data (Hefner et al., 2014). RF iw comprised of a decision tree and bagging framework in which each of the trees is independent of the other (Liaw and Wiener, 2002). While developing the decision tree, RF does not carry out any pruning. Although as a single tree the prediction accuracy is not high, the combination generates high probability measures of the whole prediction (Xu and Luo, 2021).

The approach is based on two basic steps: (1) the formation of the forest and (2) the process of decision-making. For the prediction, the operation process of RF is to divide the training samples into *n* samples at random, construct *n* Classification and Regression Tree decision trees, and determine the classification results according to the simple voting method. The prediction results are the mean values of *n* leaf nodes; that is, the simple average values of the prediction results of multiple decision trees are taken as the prediction results. The operation process of the RF algorithm is shown in Figure 20.

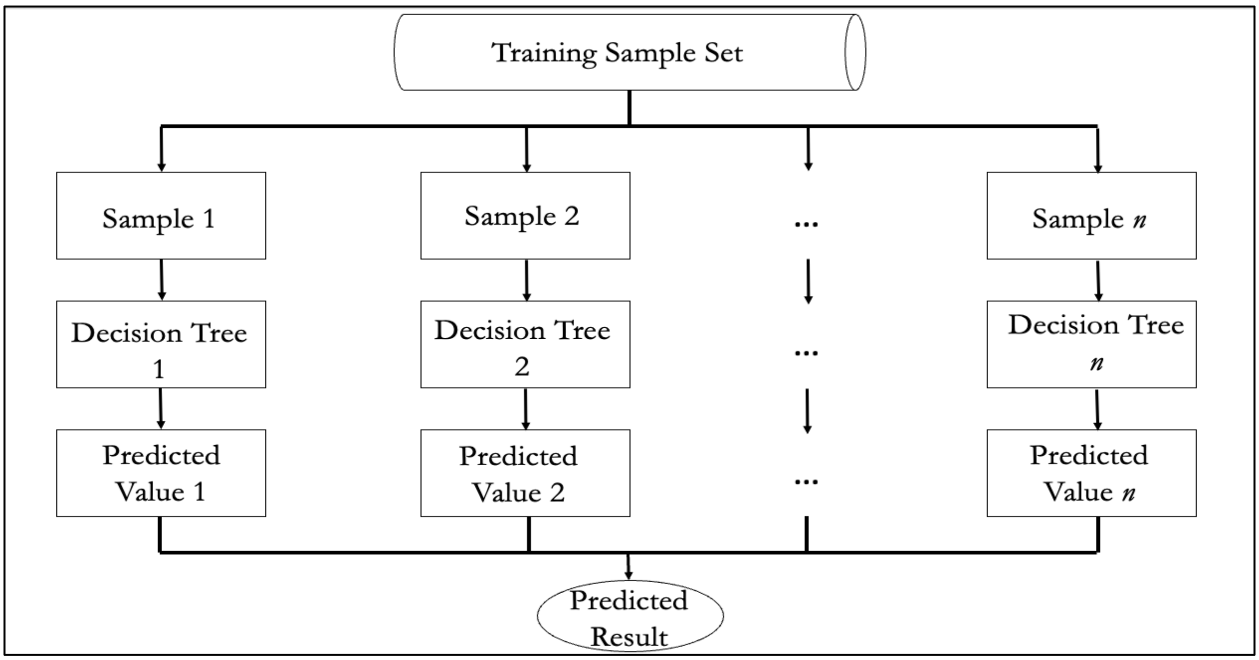


Figure 20. RF Algorithm (Xu and Luo, 2021).

The margin function in CB is given by Equation 3 (Breiman, 2001).

$$\mathrm{mg}(X, Y) = \mathrm{av}_k \, I(h_k(X) = Y) - \max_{j \neq Y} \mathrm{av}_k \, I(h_k(X) = j) \tag{3}$$

Where,

$h_1(x)$, $h_2(x)$, . . ., $h_k(x)$ are ensembles of classifiers,

X, and Y are random vectors, and

I is the indicator function.

The margin function evaluates the difference in the average number of votes between the right class that exceeds the average vote for all other classes. A larger margin indicates a higher level of confidence in the classification (Breiman, 2001). The error generalization is given by Equation 4.

$$PE^{*} = P_{X,Y}\,(mg(X, Y) < 0) \tag{4}$$

Where, X, Y indicates that the probability is over the X, Y space.

This study used the Python ML library scikit-learn function "sklearn.ensemble.RandomForestRegressor" to utilize this algorithm.

7.2.1.2. Gradient Boosting (GB)

The GB method was proposed by Friedman (Friedman, 2001; Friedman and Meulman, 2003). GB is commonly known as multiple additive trees and is a unique improvement in data mining and an enhancement on top of decision trees, which uses stochastic GB (Friedman, 2001). Boosting is a technique used to boost the accuracy of a learning algorithm by combining multiple models that have low error rates. In addition, combining multiple models into an ensemble makes them perform better.

Figure 21 shows a flow chart of the GB ML method. The ensemble classifiers consist of a set of weak classifiers. The weights of the incorrectly predicted points are increased in the next classifier. The final decision is based on the weighted average of the individual predictions (T. Zhang et al., 2021). A GB model can be viewed as a series expansion approximating the true functional relationship. The algorithm for the GB model is described in Equation 5 (De'ath, 2007; Hastie et al., 2009).

$$f(x) = \sum_{n} f_n(x) = \sum_{n} \beta_n g(x, \gamma_n) \tag{5}$$

Where x is a set of predictors and f(x) is the estimate of the response variable. $g(x, \gamma_n)$ are single decision trees with the parameter γ_n signaling the split variables. Coefficients β_n (n = 1,2,...,n) determine how each single tree is joined together. Values of β_n depend on the minimization of a specified loss function, $L(y_i, f(x_i))$. Prediction performance is measured by a loss function such as deviance. A numerical optimization method named functional gradient descent was proposed by (Friedman, 2001). The algorithm to initialize $f_0(x)$ is given below:

1. For n = 1,2,3,…m (number of trees)
   a. For i = 1 to m (number of observations), calculate the residuals
   $$\tilde{y}_{in} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}$$
   b. Fit a decision tree to $\tilde{y}_{in}$ to estimate γ_n
   c. Estimate β_n by minimizing $L(y_i, f_{n-1}(x_i) + \beta_n g(x, \gamma_n))$
   d. Update $f_n(x) = f_{n-1}(x) + \beta_n g(x, \gamma_n)$
2. Calculate

$$f(x) = \sum_{n} f_n(x)$$

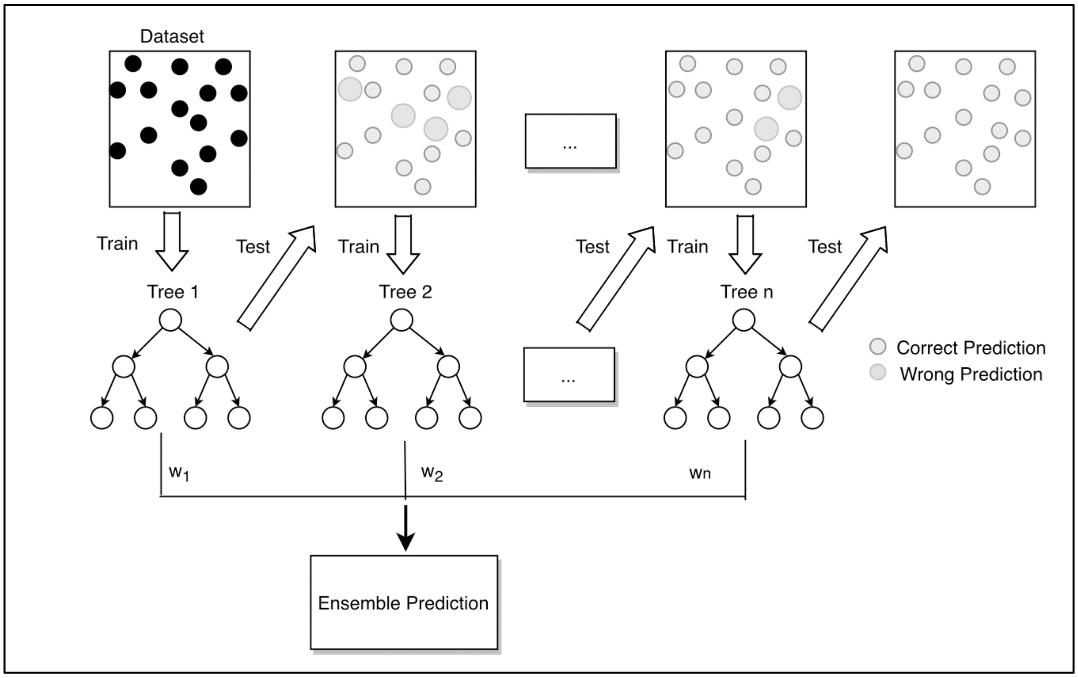


Figure 21. Flowchart for Gradient Boosting (T. Zhang et al., 2021).

The tree building process adds trees iteratively till all results are best fit, which leads to overfitting. In that scenario, trained models perform well only on trained data and have low prediction accuracy with other datasets. To avoid overfitting, the model is also tested by fitting a test dataset. Iterative training will stop when the performance of the model reaches a point at which the model predicts well for both the training and test dataset. Regularization parameters help overcome overfitting and improve model performance. These parameters have two components: (1) learning rate and (2) tree complexity. The learning rate decides how quickly the model is updated or improved after each stage. The learning rate ranges from 0.0001 to 1.0. A learning rate with lower value benefits in minimizing the loss function, but it needs more data on trained trees and time to run the model (De'ath, (2007). Values closer to 1.0 will need less training data efforts but result in overfitting and poor performance. Tree complexity describes the number of nodes per single simple tree. The simplest tree is a tree with two nodes, with only one split (Hastie et al., 2009). Both the learning rate and tree complexity rate are balanced to avoid the overfitting issue.

The GB regression model is implemented in a Python ML library named scikit-learn by using the function "sklearn.ensemble.GradientBoostingRegressor."

7.2.1.3. K-Nearest Neighbor (KNN)

The KNN algorithm was originally a classification algorithm proposed by Cover and Hart (1967). In recent years, it has been widely used as a nonparametric regression method. The basic algorithm process of KNN can be summarized as follows: the target value is predicted using state vectors that are in turn constructed using historical and current data. Using Euclidean distance between the present state vector and each previous state vector, *k* historical moments with the smallest distances are selected as the KNN. The prediction result of the target time can be obtained by calculating the average value of k neighbors at the next time point. Figure 22 illustrates the detailed process of KNN.

KNN is commonly known as sample-based learning. This algorithm is a useful data mining technique that allows past data samples to be used with known output values to estimate an unknown output value of a new data sample. Instead of making generalized assumptions, the algorithm compares new problem sets with those sets seen in the test and stored in memory. The most important advantage of the neighbor algorithm is that it can adapt its model to large amounts of data. KNN estimates a value or class for a new sample while calculating distances or similarities with previous training examples. The value is found by calculating the distances from each point in the KNN master data set to a point in the test data, of which the core value is unknown. Thus, neighbors are calculated by selecting the k number of observations with the closest distance. This method uses Euclidean distance, which is formulated in Equation 6 for points i and j when calculating distances (Çavuşoğlu and Kaçar, 2019).

$$\mathrm{d}(\mathrm{i},\mathrm{j}) = \sqrt{\sum_{k=1}^{p}(x_{ik} - x_{jk})^2} \tag{6}$$

Where i and j are data points on the graph. The KNN approach, using K = 3, is illustrated in Figure 22, which is a simple situation with six blue observations and six orange observations. On the left side of Figure 22, a test observation at which a predicted class label is desired is shown as a black cross. The three closest points to the test observation are identified, and it is predicted that the test observation belongs to the most commonly occurring class, in this case blue. On the left side of Figure 22, the KNN decision boundary for this example is shown in black. The blue grid indicates the region in which a test observation will be assigned to the blue class, and the orange grid indicates the region in which it will be assigned to the orange class.

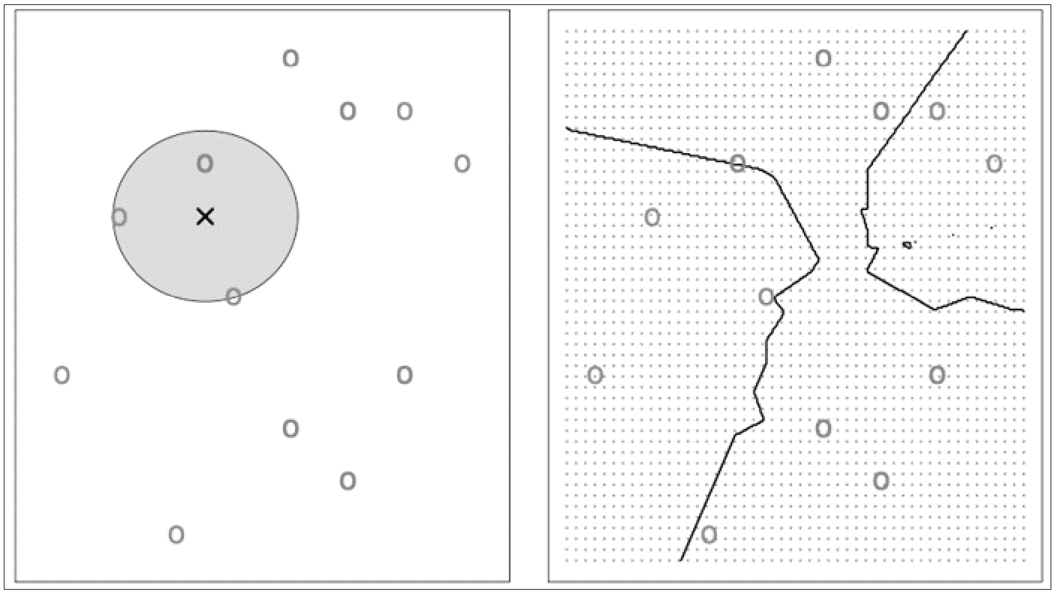


Figure 22. KNN Approach Using K = 3 (James et al., 2021).

In Figure 23, the black curve indicates the KNN decision boundary on the data, using K = 10. The Bayes decision boundary is shown as a purple dashed line. The KNN and Bayes decision boundaries are very similar. The KNN regression model is implemented in a Python ML library scikit-learn with the function “sklearn.neighbors.KNeighborsRegressor.”

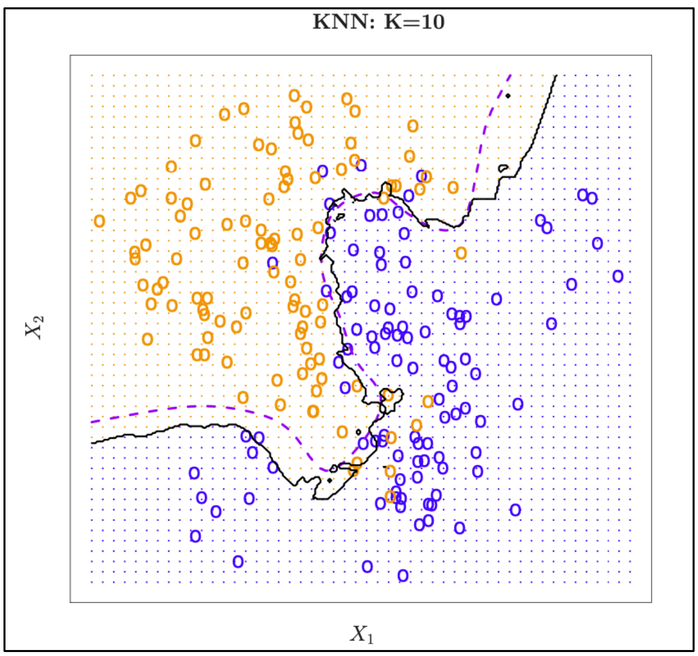


Figure 23. KNN Approach Using K = 10 (James et al., 2021).

7.2.1.4. Support Vector Regression (SVR)

SVR is a supervised ML model derived from a support vector machine (SVM) (Smola and Schölkopf, 2004; Vapnik, 2000). SVR’s approach is quite similar to SVM, although it has a few minor updates (Yang et al., 2017). SVR maps the original data nonlinearly and analyzes the linear regression problem in advanced dimensional feature space (Figure 24). SVR first arranges the initial input data nonlinearly and solves the problem in linear regression on a higher dimensional feature space. Thus, SVR creates a function to describe a nonlinear relationship.

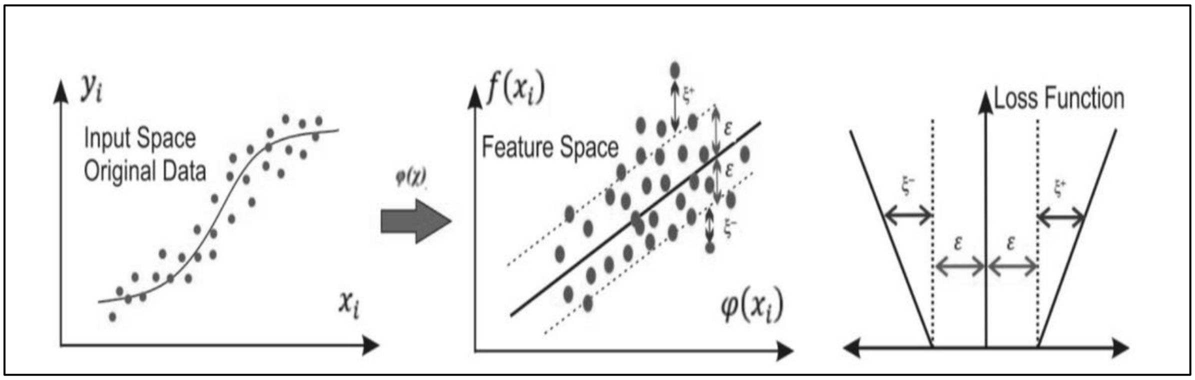


Figure 24. SVR Approach (Özdoğan-Sarıkoç et al., 2023).

The regression function is defined by the following equation (Özdoğan-Sarıkoç et al., 2023).

$$f(x) = \sum_{i=1}^{n} (\partial_i^- - \partial_i^+) \, K \, (x_{i,} x_j) + b \tag{7}$$

Where K(xi, xj) is the kernel function. In this study, the kernel functions tested were radial basis function, polynomial, and linear kernels, which are shown in Equation 8.

$$\begin{aligned} &\text{Linear Kernel}: \ \mathrm{K}\,(\mathrm{x_i}, \mathrm{x_j}) = \mathrm{x.x_i} \\ &\text{Polynomial Kernel}: \ \mathrm{K}\,(\mathrm{x_i}, \mathrm{x_j}) = (\gamma(\mathrm{x.x_i}) + \mathrm{b})^{\mathrm{d}} \\ &\text{Radial Basis Function Kernel} = \mathrm{K}\,(\mathrm{x_i}, \mathrm{x_j}) = \exp(-\frac{\|\mathrm{x} - \mathrm{x_i}\|}{2\sigma^2}) \end{aligned} \tag{8}$$

Where γ is the structural parameter in the radial basis function and polynomial kernel, ν represent the residuals, and d is the degree of the polynomial term.

7.2.1.4. Artificial Neural Network (ANN)

ANNs are modeled to mimic the human brain and can be defined as a computational model that performs simulation analysis on structural and functional parts of biological NNs. The creation of an ANN model is dependent on the type of data used. ANNs are artificially intelligent predictive models capable of efficient prediction of parameters. They are also a computing model created to mimic the human brain and nervous systems of the human body. ANNs are regarded as superior models to classical models because they can effectively simulate intelligence. ANNs are data-driven and contain several numbers of operational elements connected to each other called neurons (Mohamed, 2013; Pirouzmand and Kazem Dehdashti, 2015).

ANNs are made up of different types of layers—the input layer, the hidden layer, and the output layer, as shown in Figure 25. ANN models are models that can learn, save, and train appropriate input and output parameters that can be missed from experimental training and testing procedures of the ANN model (Elsafi, 2014; Mattar et al., 2015). ANN learning operations occur when the weights related to the correlations are appropriately regulated; due to this regulation, the real output and the input will be well correlated. ANNs are significant artificial intelligent models that are also called nonparametric regression models. ANNs have been applied over the years to find solutions to complex life problems by developing a prediction and classification model for uncertain situations (Azizi and Ahmadloo, 2016; Mia and Dhar, 2016; Nasr et al., 2012). An ANN is a suitable option over classical models due to its prediction accuracy (Chiteka and Enweremadu, 2016).

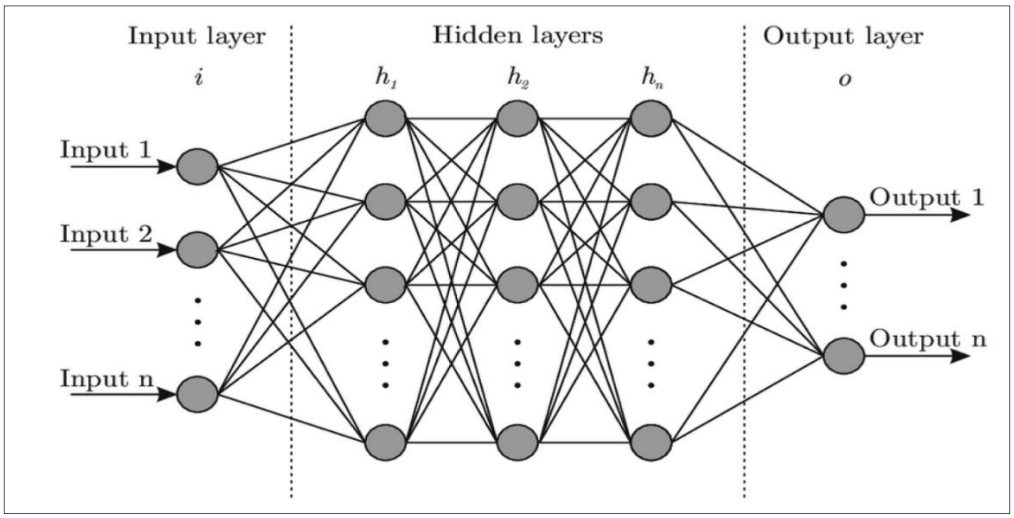


Figure 25. ANNs (Olayode et al., 2021).

ANNs are made up of three layers called the input, hidden, and output layers. The node of the bias in the input layer of the ANN model is written in Equation 9.

$$X_H = \sum_{j\,1}^{j\,n} W_{h_j i_j} + W_{h_b b_i} \tag{9}$$

Where, X_H represents the inputs in the ANN model. W_{h_j} is the weight in between the input layer and the hidden layer; W_{h_b} is the weight between the hidden layer and the bias node; b is the input layer, and it is called a bias node; and i is the input node of the ANN model.

Mathematically, X_H is used after applying the activation function. Several researchers have used various activation functions to predict and understand the relationships between input and output variables (Ghumman et al., 2011; Illias et al., 2016).

7.2.1.5. CatBoost (CB)

CB is an ML method derived from a gradient boosting decision tree (GBDT), and it was proposed by engineers of Yandex in 2017 (Prokhorenkova et al., 2018). Figure 26 illustrates the architecture of a CB model. GB is a powerful ML technique, capable of solving problems with heterogeneous features, noisy data, and complex dependencies.

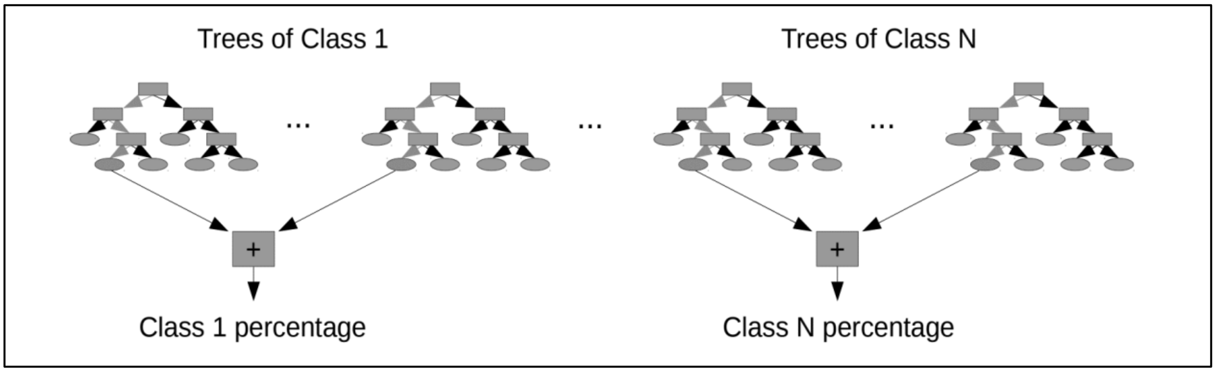


Figure 26. CatBoost (Alcolea et al., 2020).

Compared to other GBDT algorithms, CB has the following advantages: First, categorical data are handled well by CB. Traditional GBDT algorithms can replace categorical features with corresponding average label values. In decision trees, node splitting is done based on average label value. This method is called Greedy Target-based Statistics, which is defined by Equation 10 (Prokhorenkova et al., 2018).

$$\frac{\sum_{j=1}^{p}\left[x_{j,k} = x_{i,k}\right] Y_i}{\sum_{j=1}^{n}\left[x_{j,k} = x_{i,k}\right]} \tag{10}$$

When considering a dataset of observations D = {X_i, Y_i} i = 1, …, n, if a permutation is σ = (σ_1,..., σ_n), $x_{\sigma_{p,k}}$ is substituted with:

$$\frac{\sum_{j=1}^{p=1}\left[x_{\sigma_{j,k}} = x_{\sigma_{j,k}}\right] Y_{\sigma_j} + aP}{\sum_{j=1}^{p=1}\left[x_{\sigma_{j,k}} = x_{\sigma_{j,k}}\right] + a} \tag{11}$$

Where p is a prior value, and a is the weight of the prior value. This method contributes to reducing the noise obtained from the low frequency category.

Second, CB combines multiple categorical features. CB uses a greedy way to combine all categorical features and their combinations in the current tree with all categorical features in the dataset. Third, gradient bias is overcome by CB. GBDT generates a weak learner in each iteration, and each learner is trained based on the gradient of the previous learner, and the accumulation of classified results of all learners provides the output (Friedman, 2002). The final learned model might be overfit because there will be biased pointwise gradient estimation. CB uses a new method to change the gradient estimation method in the classic algorithm, which is named ordered boosting. This method can overcome prediction shifts caused by gradient bias and further enhance the generalization ability of the model (Prokhorenkova et al., 2018).

To obtain an unbiased gradient estimation, CB trains a separate model M_i for each sample X_i. The model M_i is trained with a training set that does not contain sample X_i. M_i is used to obtain a gradient estimation of the sample. Furthermore, this gradient will be used to train the base learner for the final model.

7.2.2 Evaluation Criteria

7.2.2.1. Mean Absolute Error

Absolute error represents the amount of error in a prediction and is defined by the difference between the predicted value and the true value. The mean absolute error (MAE) is the average of all absolute errors and is defined by Equation 12. The scale or unit of the MAE is the same as the original data and therefore can be named as a scale-dependent accuracy measure (Willmott and Matsuura, 2005). The MAE has only nonnegative values since absolute values are considered and therefore can avoid mutual cancellation of the positive and negative errors.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{12}$$

Where n is the number of errors or sample size, and O_i and P_i are the true value and predicted value of the i^{th} observation, respectively.

7.2.2.2. Mean Squared Error

Mean squared error (MSE) is calculated by averaging the square of all errors and defined by Equation 13. The lower the MSE value, the better the performance of the prediction models. Higher error values are penalized more than the lower ones due to the nature of the square function, and therefore for outliers, MSE will become much larger than MAE (Botchkarev, 2018). Also, the unit of MSE is different than the original data.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2 \tag{13}$$

Where n is the number of errors or sample size, and O_i and P_i are the true value and predicted value of the i^{th} observation, respectively.

7.2.2.3. Root Mean Squared Error

Root mean squared error is defined as the square root of the average of the square of all of the errors, as described in Equation 14. RMSE is always nonnegative, and a lower value indicates a good fit for the data. In general, a lower RMSE is better than a higher one. Although RMSE is a good performance evaluation matrix, it cannot be used to compare between variables since it is a scale-dependent matrix (Christie and Neill, 2022).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (14)$$

Where n is the number of errors or sample size, and O_i and P_i are the true value and predicted value of the i^{th} observation, respectively.

7.3 RESULTS AND ANALYSIS

7.3.1 Model Selection Results

In this study, a variety of regression models were utilized to predict road congestion measures. Because PVTT is the only travel time-related performance measures among the four congestion measures, this section shows the model performance results for only speed-related measures. The final tool development will consider all four congestion performance measures for data visualization. The RandomForestRegressor, GradientBoostingRegressor, KNeighborsRegressor, SVR, and MLPRegressor packages from the Scikit-learn library and the CatBoostRegressor package from the CB library were used to conduct the regression. Table 21 presents the final optimized parameters for each model. The RF model has an optimized parameter of n_estimators = 500, max_depth = 8, and random_state = 0. The GB model has an optimized parameter of n_estimators = 1000, max_depth = 8, learning_rate = 0.01, and random_state = 0. The KNN model has an optimized parameter of n_neighbors=50 and weights ='distance'. The SVR model has an optimized parameter of C = 1 and epsilon = 0.1. The Multilayer Perceptron (ANN) model has an optimized parameter of activation = 'tanh', solver = 'lbfgs', max_iter = 1000. The CatBoostRegressor model has an optimized parameter of iterations=5000, learning_rate = 0.1, and depth = 8. In the model selection process, only the different types of land use were considered.

Table 21. Parameters Used for Each Model.

Models	Optimized Parameters
RF	n_estimators = 500, max_depth = 8, random_state = 0
GB	n_estimators = 1000, max_depth = 8, learning_rate = 0.01, random_state = 0
KNN	n_neighbors = 50,weights = 'distance'
SVR	C=1, epsilon = 0.1
ANN	Activation = 'tanh', solver = 'lbfgs', max_iter = 1000
CB	Iterations = 5000, learning_rate = 0.1, depth = 8

Table 22 presents the results of model selection for urban roads in terms of MAE, MSE, and RMSE. The dependent variables used are SpdAve, SpdStd, and Spd85. The table compares the performance of various models, including RandomForestRegressor, GradientBoostingRegressor, KNeighborsRegressor, Support Vector Regressor, Multilayer Perceptron Regressor, and CatBoostRegressor, with two sets of independent variables: “with crash” or “without crash.” It can be observed that CB performed the best among all models, followed by GB and the rest of the models. Additionally, it can be seen that the modeling accuracy is higher when the number of crashes is removed from the modeling process. However, for the purpose of recurrent and nonrecurrent congestion performance measure predictions, both approaches (with and without consideration of crash data) will be applied to the tool development phase.

Table 22. Model Selection Results of Urban Freeways.

Dependent Variables	Consideration	Model	MAE	MSE	RMSE
SpdAve	Considering crash as independent variables	RF	2.055	9.615	3.101
		GB	1.521	6.138	2.478
		KNN	2.778	17.248	4.153
		SVR	3.197	23.346	4.832
		ANN	3.343	21.553	4.643
		CB	**1.451**	**5.870**	**2.423**
	Not considering crash as independent variables	RF	2.043	9.271	3.045
		GB	1.488	5.832	2.415
		KNN	2.715	16.951	4.117
		SVR	3.196	23.341	4.831
		ANN	3.839	26.082	5.107
		CB	1.400	5.896	2.428
SpdStd	Considering crash as independent variables	RF	1.361	3.483	1.866
		GB	1.092	2.482	1.576
		KNN	1.567	4.864	2.206
		SVR	1.751	6.190	2.488
		ANN	1.868	6.240	2.498
		CB	**1.041**	**2.396**	**1.548**
	Not considering crash as independent variables	RF	1.359	3.475	1.864
		GB	1.083	2.474	1.573
		KNN	1.546	4.835	2.199
		SVR	1.751	6.191	2.488
		ANN	1.968	6.716	2.592
		CB	1.011	2.362	1.537
Spd85	Considering crash as independent variables	RF	1.382	4.370	2.091
		GB	1.032	2.872	1.695
		KNN	1.865	7.633	2.763
		SVR	2.180	10.083	3.175
		ANN	2.152	9.021	3.004
		CB	**0.979**	**2.673**	**1.635**
	Not considering crash as independent variables	RF	1.378	4.321	2.079
		GB	1.016	2.902	1.704
		KNN	1.828	7.537	2.745
		SVR	2.179	10.083	3.175
		ANN	2.430	10.791	3.285
		CB	**0.935**	**2.674**	**1.635**

Note: Best modeling results under each scenario are in bold fonts.

For rural freeways, the test results of each model are presented in Table 23. The results present a similar trend to urban freeways: CB performed the best, and the modeling accuracy is higher when the number of crashes is removed from the modeling process.

Table 23. Model Selection Results of Rural Freeways.

Dependent Variables	Consideration	Model	MAE	MSE	RMSE
SpdAve	Considering crash as independent variables	RF	0.472	0.694	0.833
		GB	0.356	0.603	0.776
		KNN	0.582	1.477	1.215
		SVR	0.830	2.199	1.483
		ANN	0.923	2.390	1.546
		CB	**0.355**	**0.560**	**0.748**
	Not considering crash as independent variables	RF	0.471	0.700	0.837
		GB	0.354	0.607	0.779
		KNN	0.577	1.493	1.222
		SVR	0.830	2.198	1.483
		ANN	0.917	2.354	1.534
		CB	**0.335**	**0.577**	**0.760**
SpdStd	Considering crash as independent variables	RF	0.548	0.585	0.765
		GB	0.386	0.335	0.579
		KNN	0.561	0.684	0.827
		SVR	0.780	1.069	1.034
		ANN	0.776	1.094	1.046
		CB	**0.367**	**0.318**	**0.564**
	Not considering crash as independent variables	RF	0.547	0.584	0.764
		GB	0.376	0.328	0.573
		KNN	0.556	0.685	0.828
		SVR	0.779	1.069	1.034
		ANN	0.792	1.107	1.052
		CB	**0.348**	**0.299**	**0.547**
Spd85	Considering crash as independent variables	RF	0.619	1.083	1.040
		GB	0.389	0.552	0.743
		KNN	0.712	1.665	1.290
		SVR	1.170	2.867	1.693
		ANN	1.224	3.007	1.734
		CB	0.374	0.524	0.724
	Not considering crash as independent variables	RF	0.615	1.057	1.028
		GB	0.379	0.536	0.732
		KNN	0.704	1.672	1.293
		SVR	1.170	2.867	1.693
		ANN	1.269	3.165	1.779
		CB	**0.355**	**0.543**	**0.737**

Note: Best modeling results under each scenario are in bold fonts.

7.3.1.1. CatBoost Modeling Results

Table 24 presents the results of a CB model applied to predict speed-related congestion measures (SpdAve, SpdStd, Spd85) for two different land use types (rural and urban) at ALL PSL levels and three different PSL levels (65, 70, and 75).

The R-squared values are a measure of how well the model explains the variance in the target variable and range from 0 to 1, with higher values indicating a better fit. In this table, the R-squared values for all models are above 0.662, with some models having values above 0.9, indicating that these models have a good fit to the data.

The MAE, MSE, and RMSE are measures of the model's prediction accuracy, with lower values indicating better performance. Generally, the MAE, MSE, and RMSE results for the models in the table indicate good prediction accuracy, with values that are relatively low compared to the range of the target variable.

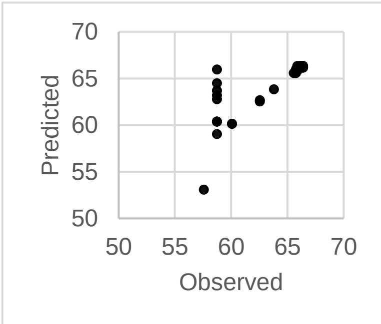
It is also worth noting that the prediction accuracy seems to improve when building the models based on different PSL levels, which suggests that the PSL level may be a useful cluster or filtering option for speed-related congestion measures. Additionally, the results suggest that the performance of the model is better for rural freeways than for urban freeways, which might be due to differences in the characteristics of the data between the two land use types.

Table 24. CatBoost Modeling Results.

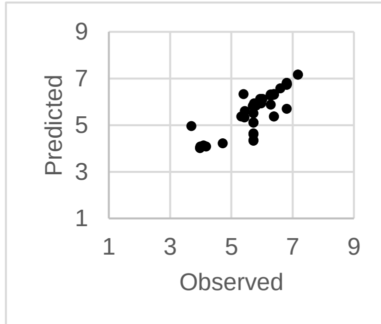
Land Use	PSL (mph)	Dependent Variables	MAE	MSE	RMSE	R^2
Rural	65	SpdAve	0.234	0.103	0.322	0.983
		SpdStd	0.406	0.385	0.621	0.897
		Spd85	0.199	0.102	0.319	0.981
	70	SpdAve	0.237	0.162	0.403	0.890
		SpdStd	0.269	0.192	0.438	0.852
		Spd85	0.237	0.150	0.388	0.903
	75	SpdAve	0.330	0.307	0.554	0.716
		SpdStd	0.336	0.291	0.539	0.697
		Spd85	0.386	0.446	0.668	0.857
	ALL	SpdAve	0.325	0.427	0.653	0.844
		SpdStd	0.330	0.246	0.496	0.784
		Spd85	0.354	0.424	0.651	0.915
Urban	65	SpdAve	1.619	6.481	2.546	0.767
		SpdStd	1.352	3.870	1.967	0.654
		Spd85	0.851	1.428	1.195	0.785
	70	SpdAve	1.137	3.436	1.854	0.760
		SpdStd	0.989	2.189	1.480	0.693
		Spd85	0.705	1.239	1.113	0.717
	75	SpdAve	0.836	2.015	1.420	0.825
		SpdStd	0.739	1.333	1.155	0.702
		Spd85	0.655	1.079	1.039	0.762
	ALL	SpdAve	1.449	7.323	2.706	0.733
		SpdStd	1.028	2.452	1.566	0.689
		Spd85	0.974	3.418	1.849	0.713

7.3.2 Model Performance

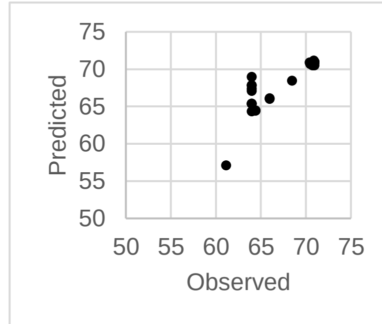
Figure 27 and Figure 28 show the scatter plots of observed and predicted congestion performance measures for different subsets based on PSLs. The linear trends of the majority of the models indicate that AI models are well suited for determining the predicted values. One of the major contributions of this project is determining congestion measures for the freeway segments with no historical speed or congestion-related information.



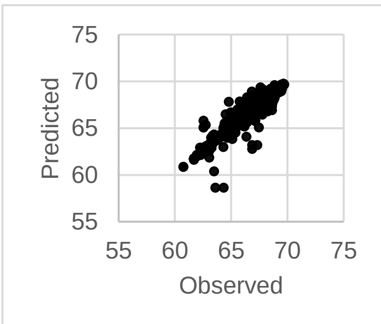
(a) **SpdAve (PSL = 65 mph)**



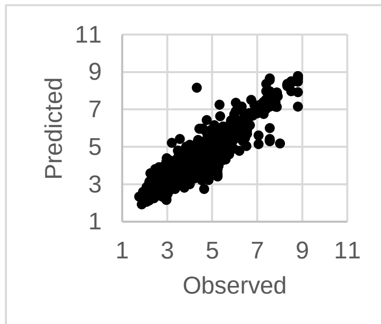
(b) **SpdStd (PSL = 65 mph)**



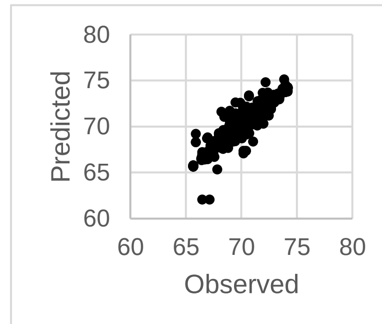
(c) **Spd85 (PSL = 65 mph)**



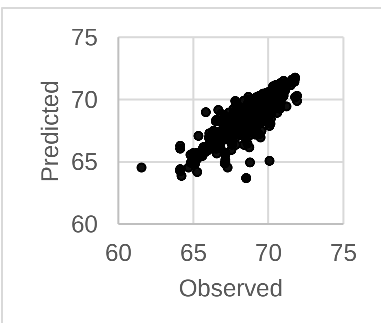
(d) **SpdAve (PSL = 70 mph)**



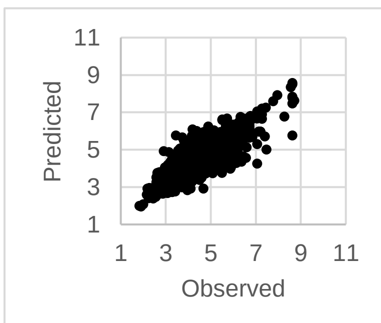
(e) **SpdStd (PSL = 70 mph)**



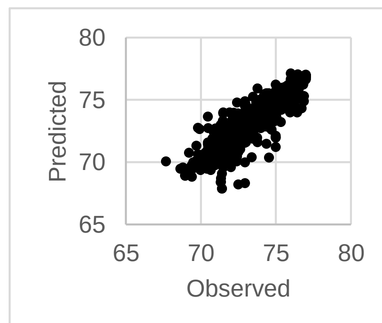
(f) **Spd85 (PSL = 70 mph)**



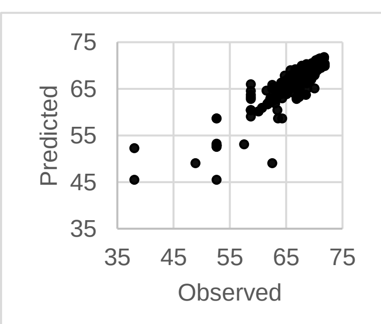
(g) **SpdAve (PSL = 75 mph & Up)**



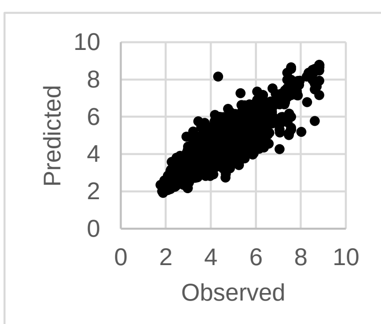
(h) **SpdStd (PSL = 75 mph & Up)**



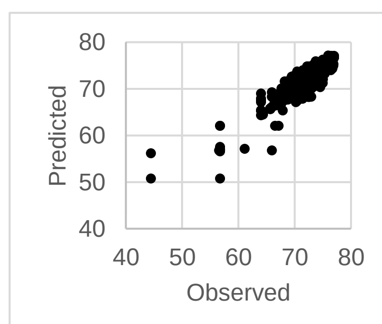
(i) **Spd85 (PSL = 75 mph & Up)**



(j) **SpdAve (PSL = All)**

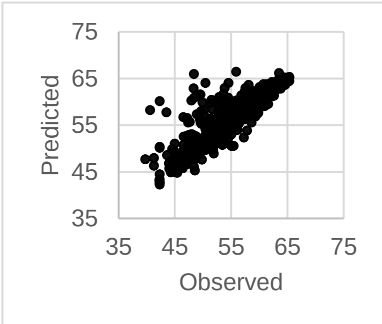


(k) **SpdStd (PSL = All)**

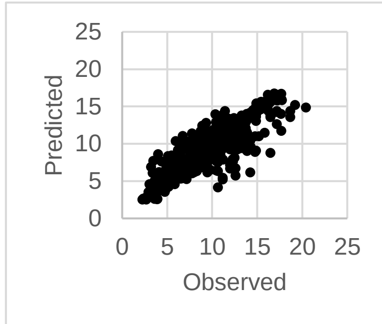


(l) **Spd85 (PSL = All)**

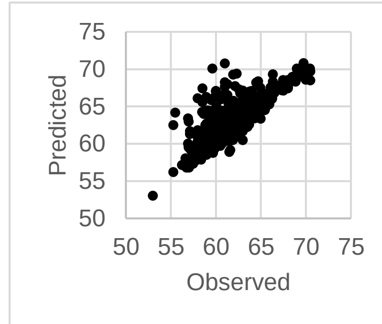
Figure 27. Rural Freeway Model Performances.



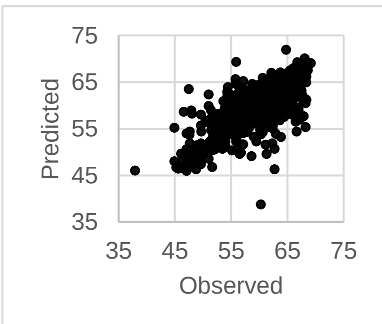
(a) **SpdAve (PSL = 65 mph)**



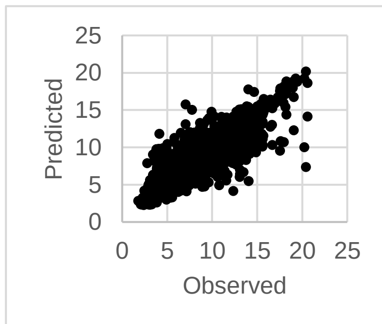
(b) **SpdStd (PSL = 65 mph)**



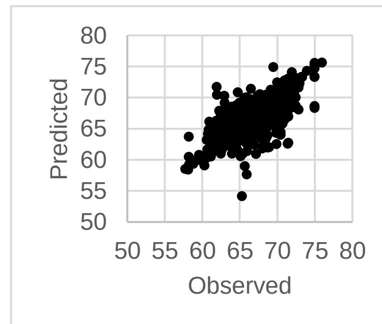
(c) **Spd85 (PSL = 65 mph)**



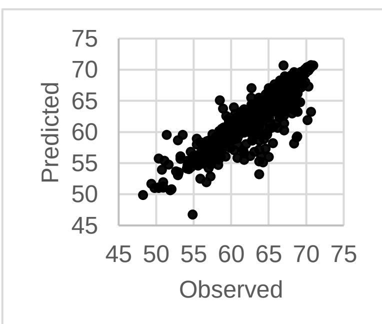
(d) **SpdAve (PSL = 70 mph)**



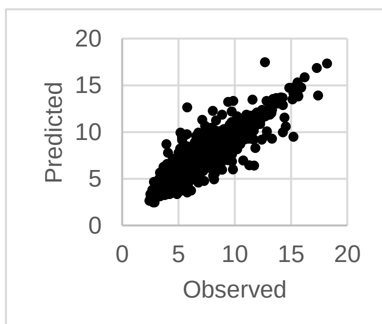
(e) **SpdStd (PSL = 70 mph)**



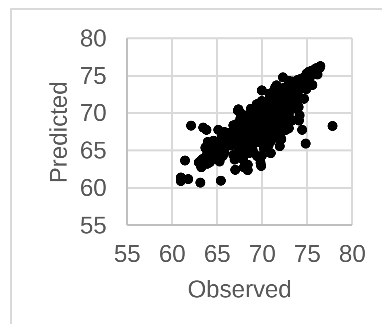
(f) **Spd85 (PSL = 70 mph)**



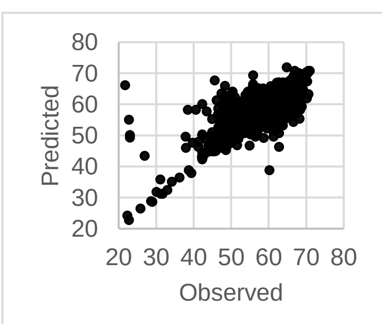
(g) **SpdAve (PSL = 75 mph & Up)**



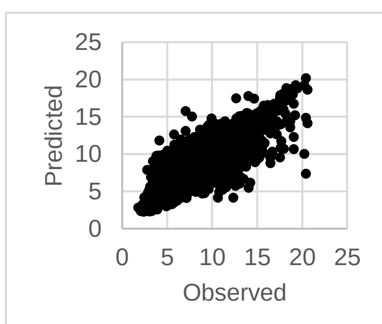
(h) **SpdStd (PSL = 75 mph & Up)**



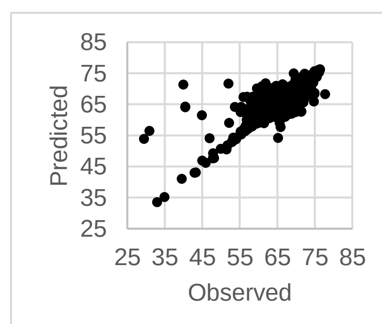
(i) **Spd85 (PSL = 75 mph & Up)**



(j) **SpdAve (PSL = All)**



(k) **SpdStd (PSL = All)**



(l) **Spd85 (PSL = All)**

Figure 28. Urban Freeway Model Performances.

7.3.3 Explainable AI

7.3.3.1. Shapley Additive exPlanations

The Shapley Additive exPlanations (SHAP) approach explains the outcomes from the XGBoost model. ML techniques are sometimes known as “black boxes” since it is difficult to understand how each individual variable affects the results of the model. The SHAP method uses a game-theoretic approach to estimate the individual contribution of each feature to the model prediction. This method can be used to explain tree-based ML models. Equation 15 can be used to determine the Shapley value of a feature i for any certain prediction:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}}^{T} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \tag{15}$$

Where,
ϕ_i : Shapley value of feature i;
S : a possible feature subset;
$|S|$: feature counts in subset S;
F : the set of all features;
$|F|$: feature counts in set F;
$f_{S \cup \{i\}}(x_{S \cup \{i\}})$: model prediction based on the features in subset S and feature i; and
$f_S(x_S)$: model prediction based on subset S features.

Here, $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ is the distinction between a model prediction generated only using features from subset S and a model prediction made using features from subset S and feature i. This variation may illustrate how the inclusion of feature i alters the model’s prediction results. The weighted average of all differences across all potential subsets is used to calculate the Shapley value of a certain characteristic for any given forecast.

To interpret the significance of each feature in the CB model, SHAP was used. The SHAP summary pattern of the SpdAve model is shown in Figure 29 for rural freeways. Each explanatory variable has been ranked according to its influence on the model’s predictions. A higher level of feature importance implies that the parameter has a significant impact on the measures of SpdAve. The top factors for rural freeways are traffic volume (adt_adj), jobs within 45 minutes auto travel time, k-factor, and median width.

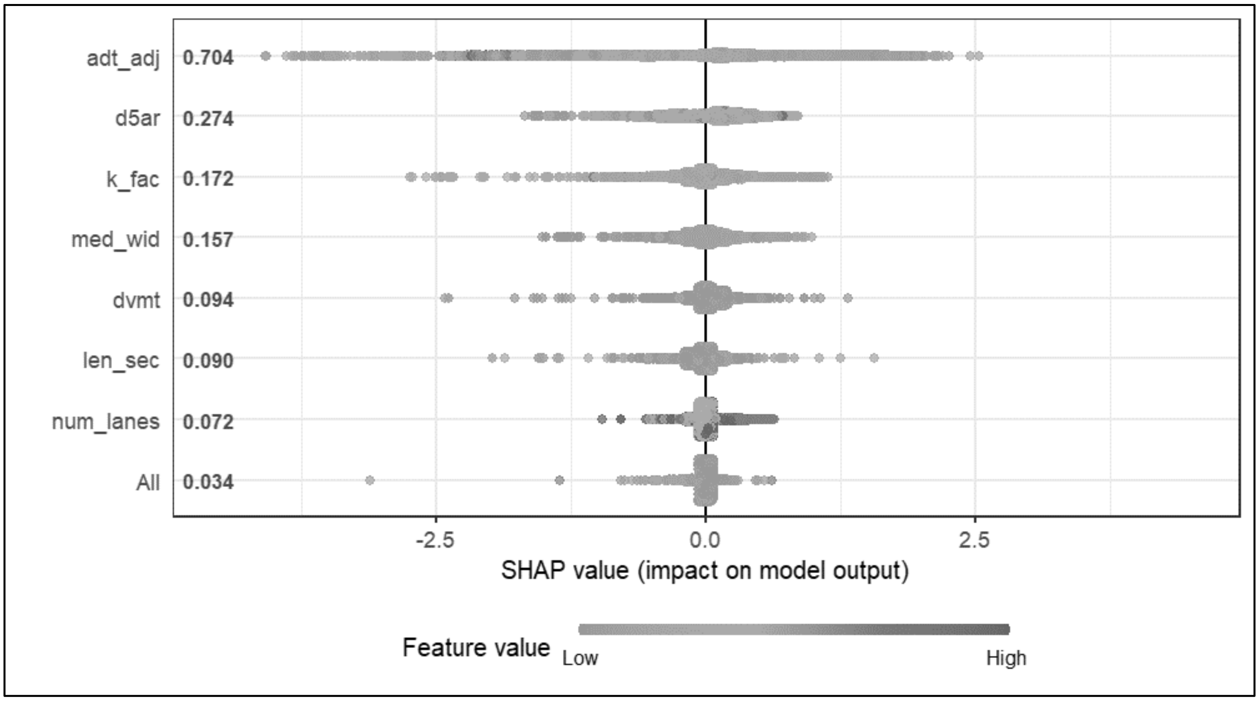


Figure 29. SHAP Plot for Rural Freeways.

The explainability of the key factors are shown in Figure 30 as partial dependence plots (PDPs). These plots indicate the effect of values in each independent variable on the congestion measures, such as SpdAve. Variables such as dvmt, segment length (len_sec), and crashes (all) have minimum impact on the congestion measures. However, traffic volume (adt_adj), jobs within 45 minutes auto travel time, k-factor, and median width show an association with congestion measures. The positive and negative association of these factors depend on different clusters of SpdAves.

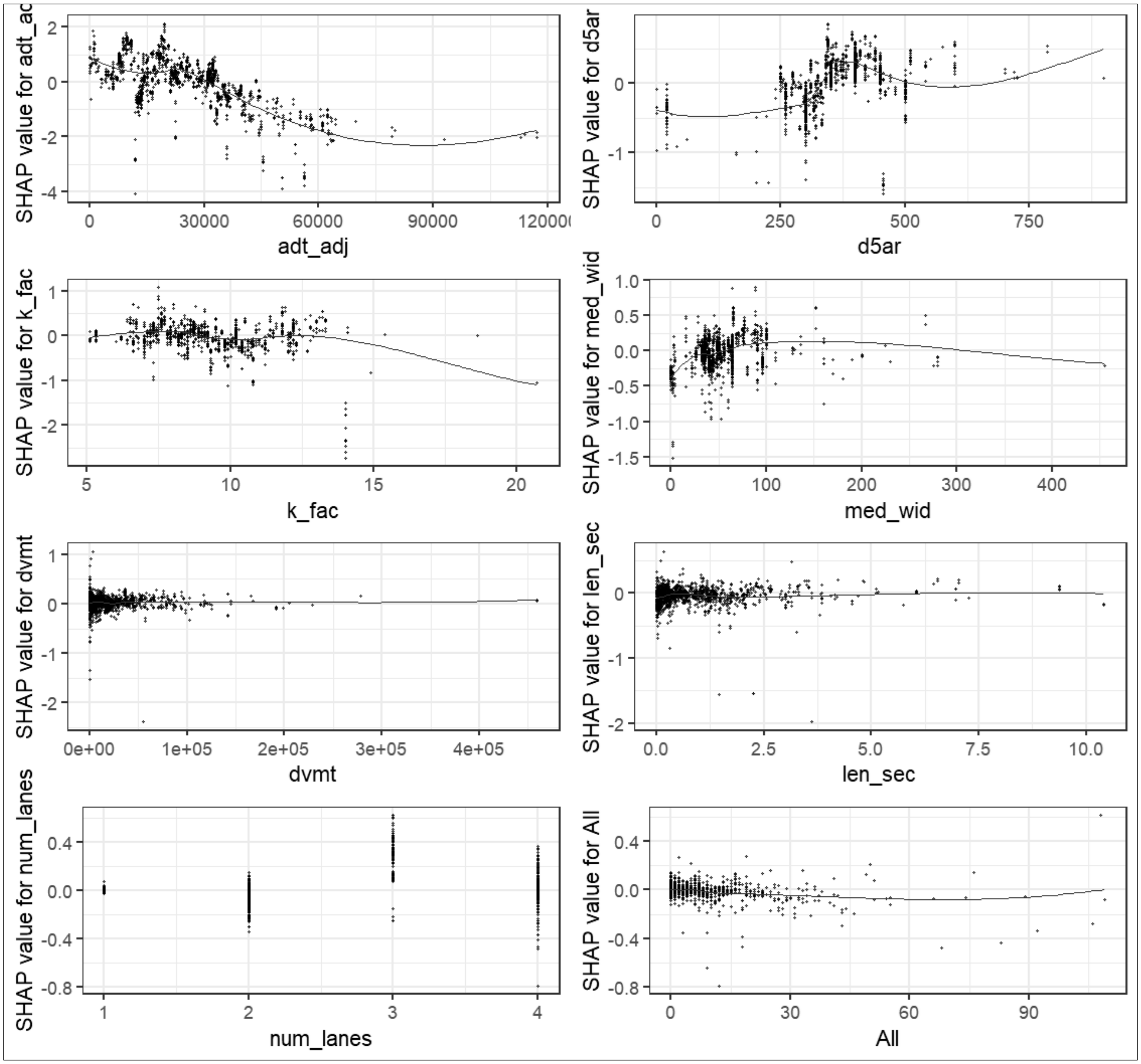


Figure 30. Partial Dependence Plots for Rural Freeways.

The SHAP summary pattern of the SpdAve model is shown in Figure 31 for urban freeways. Each explanatory variable has been ranked according to its influence on the model's predictions. A higher level of feature importance implies that the parameter has a significant impact on the measures of SpdAve. The top factors for urban freeways are traffic volume (adt_adj), median width, jobs within 45 minutes auto travel time, k-factor, and number of lanes.

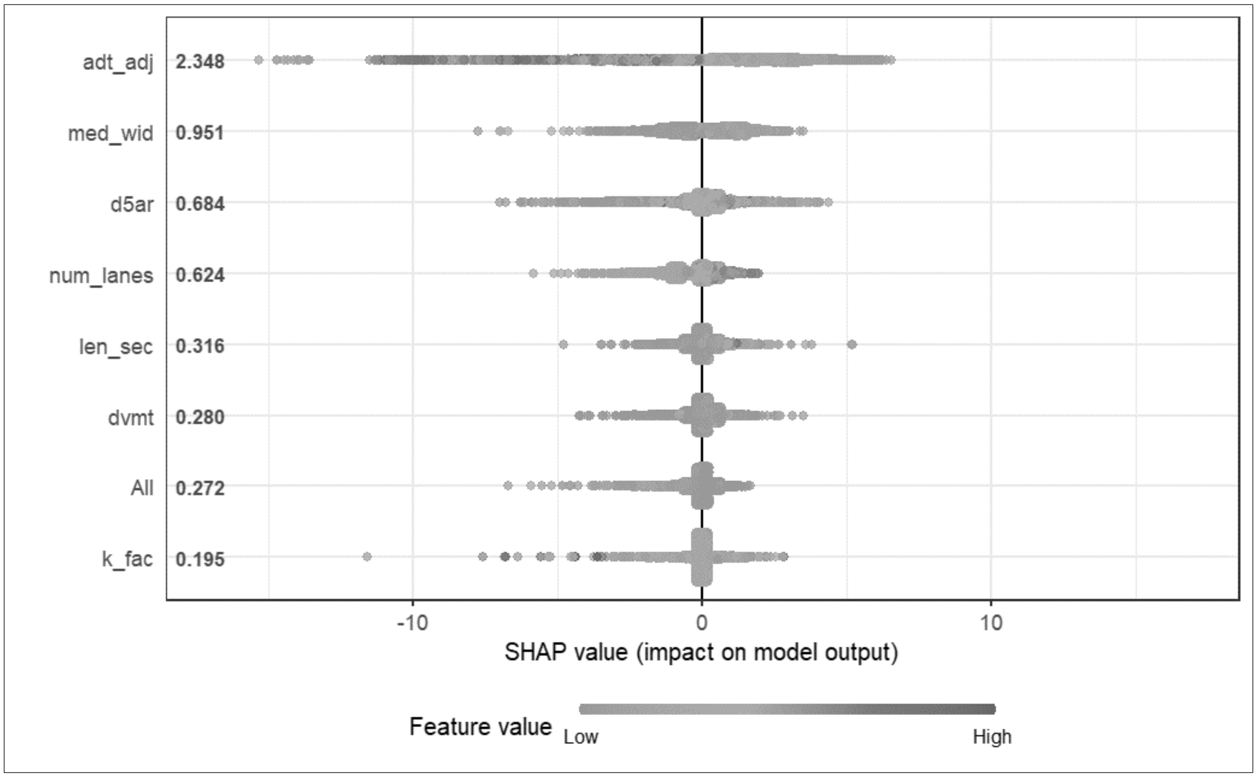


Figure 31. SHAP Plot for Urban Freeways.

The explainability of the key factors are shown in Figure 32 as PDPs. These plots indicate the effect of values in each independent variable on the congestion measures, such as SpdAve. Variables such as dvmt and segment length (len_sec) have minimum impact on the congestion measures. The number of crashes is associated with a lower SpdAve. However, traffic volume (adt_adj), jobs within 45 minutes auto travel time, median width, and lane numbers show an association with urban roadway SpdAve measures. The positive and negative association of these factors depend on different clusters of SpdAves.

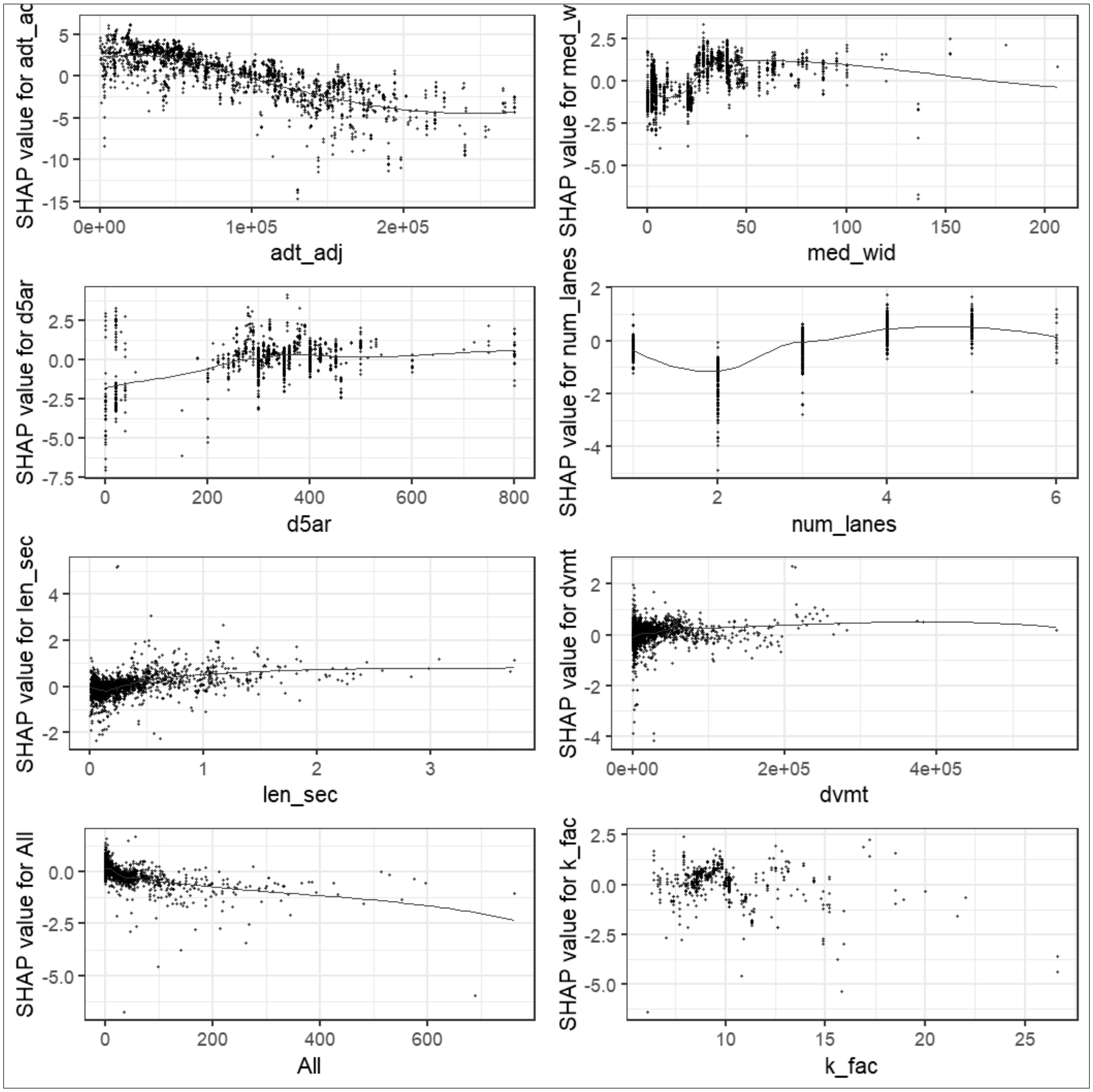


Figure 32. Partial Dependence Plots for Urban Freeways.

7.4 CHAPTER SUMMARY

Chapter 7 covered congestion measure forecasting using AI algorithms with a brief introduction to each algorithm. Model selection and model performances were discussed. It was shown that the CB model demonstrated better performance compared to other AI models. Finally, explainable AI was applied to provide interpretation of the final models.

CHAPTER 8: GUIDELINES AND SPECIFICATIONS

8.1 INTRODUCTION

The TTI team developed a GIS-based prototype decision support tool that can estimate and visually illustrate the status of freeway congestion. The decision support tool visually illustrates the congestion profile of the roadway segments (the interactive table can be used to select the most congested roadways based on different congestion measures). The GIS web-tool then color-codes the congestion level of each highway segment. The tool also has data exporting (e.g., CSV file) features. This tool can be useful for state and local transportation safety decision-makers in identifying the segments with the highest and lowest risk from the roadway networks under their supervision.

8.2 GUIDELINE ON TOOL USAGE

Figure 33 shows the interface of the opening page of the interactive tool. The tool was developed using open-source software R and the Shiny platform. The interactive web-based GIS decision support tool can be accessed at: https://txdot.shinyapps.io/0_7131/. The interface requires a login and password to open the tool in the web browser. This tab provides the core features of the tool. On the top of the page, the second tab (0-7131 Tool) contains the GIS-based interactive tool. The data and associated codes are located in a OneDrive Folder.

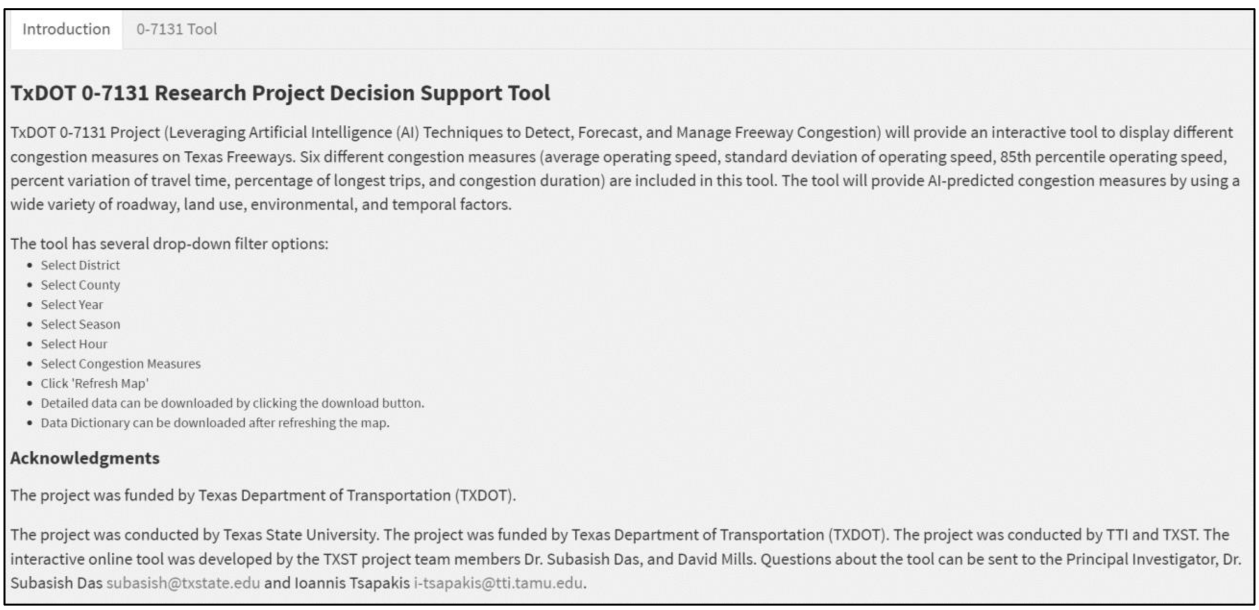


Figure 33. Interface of the 0-7131 Tool Opening Page.

Figure 34 shows the interface of the GIS-based interactive tool. The tool has three major sections: (1) the top left section has the interactive GIS-based map, (2) the top right section has several drop-down panels for different spatial and temporal clusters, and (3) the bottom section contains the interactive table with key geometric, demographic, and final congestion measures.

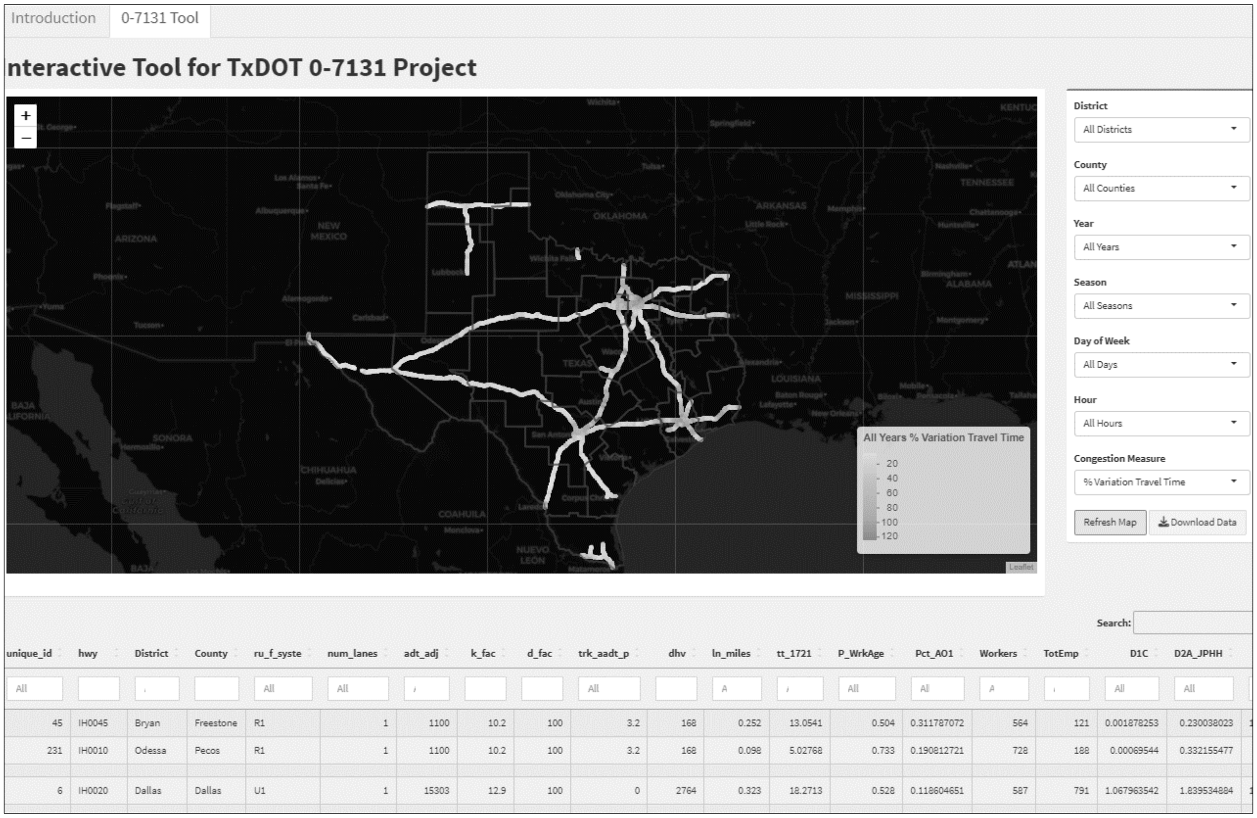


Figure 34. Interface of the Web-Based GIS Interactive Tool.

The tool contains a dashboard with various drop-down lists of steps to provide congestion measures at the segment level. Users have the flexibility of selecting several options. The tool has the following drop-down and selection options (see Figure 34):

- Districts (All Districts, Individual District).
- County (All Counties, Individual County display after selection of the district).
- Year (All Years [2018–2021] or 2018, 2019, 2020, and 2021 individually).
- Season (All Seasons or Fall, Winter, Spring, and Summer individually).
- Day of the week (All Days, or Weekends or Weekdays).
- Hours (All Hours or Daytime, Nighttime, Morning Peak, Evening Peak, and Other individually).
- Congestion Measures (SpdAve, SpdStd, Spd85, TTAve, Standard Deviation of Travel Time [TTStd], and PVTT).

The AI-based estimated congestion measures on the segment will be displayed based on the selection of different drop-down panels and the selection of the congestion measure. The estimates are graphically displayed on a color-coded map (yellow indicates a low number and red indicates a high number), and the selected segments are displayed in the bottom section of the tool in an interactive table. The selected data are also available for download. The users need to follow a few simple steps:

- Select options from drop-down panels.
- Click "Refresh Map" (will take some time to load the map and associated data).
- Zoom in/out (segments have a hovering option that shows important features of the segment).
- Click "Download Data" to download the data in CSV format.

8.3 GENERAL GUIDELINES

The users can use the tool to develop segments of interest based on the following congestion measures:

- SpdAve.
- SpdStd.
- Spd85.
- TTAve.
- TTStd.
- PVTT.

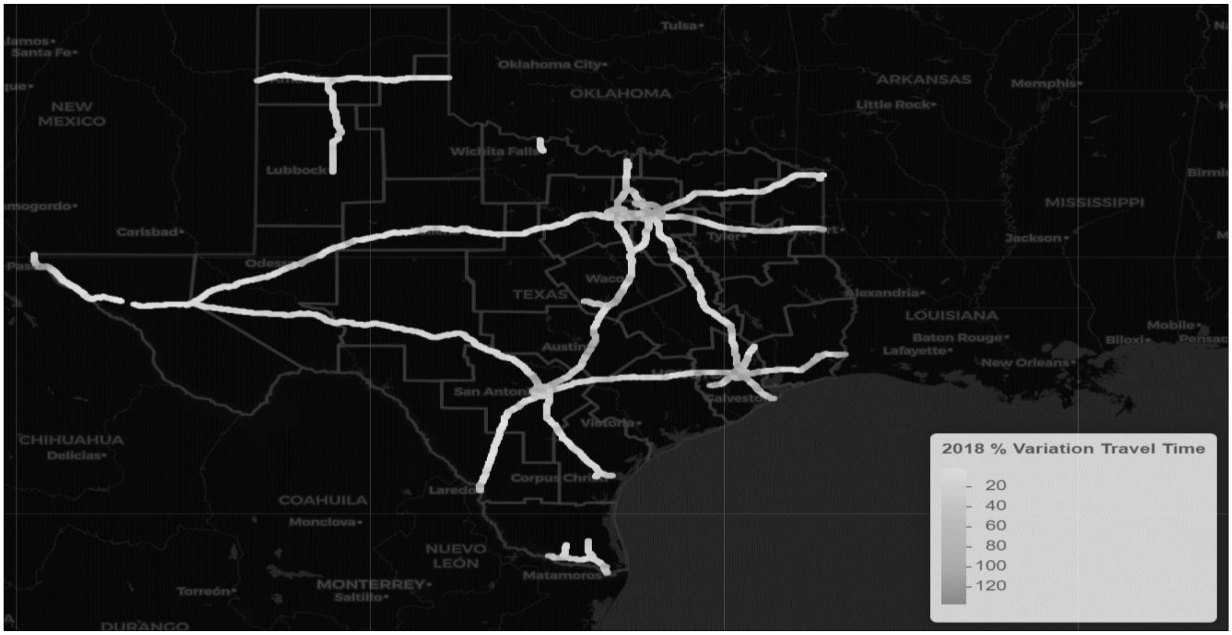
The development of AI-based predictive models forms the basis of this research, which is aimed at analyzing the year-by-year variation of congestion on Texas freeways. These models utilize historical data to forecast and understand the changing congestion patterns over time. By studying the trends and fluctuations, valuable insights into the dynamics of congestion can be obtained.

The subsequent section of this research investigates the identification of segments of interest based on the PVTT metric. PVTT is calculated as the percentage derived from the STD and mean of travel time measures, and it provides a comprehensive measure of the variation in travel times. The top segments of interest are determined by ranking the PVTT values according to two key factors: land use (urban and rural freeways) and PSLs. By examining these factors individually, a more nuanced understanding of congestion patterns and their relationship to specific contexts can be achieved.

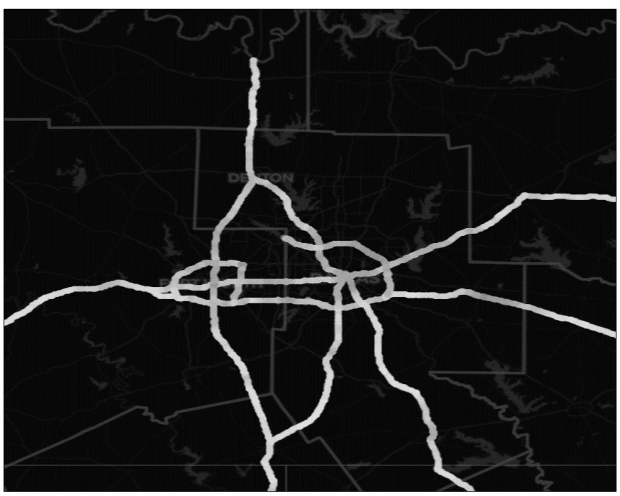
The segmentation based on land use considers both urban and rural freeways and recognizes the distinct characteristics and challenges associated with each type of roadway. Additionally, the segmentation based on PSLs provides valuable insights into how different speed limits impact congestion. By focusing on specific speed limit categories such as 65 mph, 70 mph, 75 mph, and 80 mph, researchers can analyze the relationship between speed limits and congestion levels.

8.3.1 Segments of Interest Based on 2018 Data

Figure 35 presents a comprehensive heatmap utilizing data from 2018 that depicts the segments with estimated PVTT measures. The heatmap provides a visual representation of the congestion levels experienced across various segments. To provide further granularity, Figure 35 is subdivided into Figure 35(a), 35(b), and 35(c), which focus on different regions, namely Texas as a whole, Dallas, and Houston, respectively. In each of the subfigures, the segments are color-coded based on their PVTT measures. The color gradient ranges from lighter shades to darker shades of red, with darker red colors indicating higher PVTT measures. This color scheme allows for a quick and intuitive understanding of the congestion levels present in the respective regions.



(a) Texas



(b) Dallas

(c) Houston

Figure 35. Estimated PVTT Measures Using 2018 Data.

8.3.1.1. Rural Freeways

Table 25 to Table 28 list the top 10 segments using 2018 data based on the high estimated PVTT values on rural freeways with different PSLs. Unique id (unique_id) indicates the segment identification. Some operational properties (highway name, or "hwy," district, number of lanes, AADT, and segment length, or "ln_miles") are listed in the table. Six congestion measures (SpdAve, SpdStd, Spd85, TTAve, TTStd, and PVTT) are listed as the specifications of congestion measures for these segments. The interactive tool can be used to develop a comprehensive list of all segments and it can utilize both spatial and temporal features. PVTT measures are higher on 75 mph and 70 mph rural freeways than on 80 mph and 65 mph rural freeways. The segments with high PVTT measures are mostly in Dallas, Houston, and Waco Districts (75 mph and lower). The majority of 80 mph rural freeways with higher PVTT values are in the Lubbock District. All segments listed in Table 25 to Table 28 are rural two-lane freeways.

Table 25. Top 10 Segments with High Estimated PVTT Values on 80-mph Rural Freeways Based on 2018 Data.

#	Unique _id	Hwy	District	Lanes	AADT	Ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	32736	IH0040	Amarillo	2	13315	0.968	63.88	8.71	70.43	19.25	6.2	49.3
2	18311	IH0027	Lubbock	2	10445	0.828	68.7	5.77	74.22	226.29	62.47	28.1
3	20028	IH0027	Lubbock	2	10440	2.004	68.7	5.77	74.22	226.29	62.47	28.1
4	21088	IH0027	Lubbock	2	10440	0.504	68.7	5.77	74.22	226.29	62.47	28.1
5	20027	IH0027	Lubbock	2	10440	2.004	68.7	5.77	74.22	226.29	62.47	28.1
6	18755	IH0027	Lubbock	2	10445	0.11	68.7	5.77	74.22	226.29	62.47	28.1
7	22312	IH0027	Lubbock	2	10445	2.148	68.7	5.77	74.22	226.29	62.47	28.1
8	22544	IH0027	Lubbock	2	10440	1.074	68.7	5.77	74.22	226.29	62.47	28.1
9	22313	IH0027	Lubbock	2	10445	2.148	68.7	5.77	74.22	226.29	62.47	28.1
10	9715	IH0010	San Angelo	2	9574	0.146	69.19	5.63	74	70.77	17.71	25.02

Table 26. Top 10 Segments with High Estimated PVTT Values on 75-mph Rural Freeways Based on 2018 Data.

#	Unique _id	Hwy	District	Lanes	AADT	Ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	16264	IH0020	Dallas	2	50696	0.358	65.92	8.22	69.5	71.97	70.09	97.39
2	17172	IH0020	Dallas	2	50696	0.29	65.92	8.22	69.5	71.97	70.09	97.39
3	17449	IH0020	Dallas	2	50696	0.204	65.92	8.22	69.5	71.97	70.09	97.39
4	14684	IH0020	Dallas	2	50696	1.426	65.87	8.24	69.47	70.85	68.2	95.89
5	17391	IH0020	Dallas	2	50696	0.242	65.91	8.02	69.42	86.8	75.21	90.05
6	15819	IH0020	Dallas	2	50696	0.472	66.25	7.93	69.75	52.38	48.27	89.03
7	11754	IH0020	Dallas	2	50696	0.126	66.25	7.93	69.75	52.38	48.27	89.03
8	13603	IH0020	Dallas	2	50696	0.206	66.25	7.93	69.75	52.38	48.27	89.03
9	13394	IH0020	Dallas	2	50696	0.288	66.12	7.88	69.65	56.82	49.44	86.13
10	15346	IH0020	Dallas	2	44151	0.334	66.71	7.48	69.99	34.92	28.7	82.2

Table 27. Top 10 Segments with High Estimated PVTT Values on 70-mph Rural Freeways Based on 2018 Data.

#	Unique _id	Hwy	District	Lanes	AADT	Ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	32436	IH0045	Bryan	2	40913	0.602	62.14	8.83	66.5	54.62	52.28	95.72
2	32179	IH0045	Bryan	2	40913	0.12	62.14	8.83	66.5	54.62	52.28	95.72
3	5725	IH0010	Houston	2	56083	0.044	62.2	8.44	66.47	140.67	126.76	90.11
4	7911	IH0010	Houston	2	56083	0.03	62.2	8.44	66.47	140.67	126.76	90.11
5	8511	IH0010	Houston	2	56083	0.044	62.2	8.44	66.47	140.67	126.76	90.11
6	9164	IH0010	Houston	2	56083	0.044	62.2	8.44	66.47	140.67	126.76	90.11
7	5404	IH0010	Houston	2	56083	0.046	62.2	8.44	66.47	140.67	126.76	90.11
8	6643	IH0010	Houston	2	50409	0.054	62.2	8.44	66.47	140.67	126.76	90.11
9	10507	IH0010	Houston	2	56083	0.016	62.2	8.44	66.47	140.67	126.76	90.11
10	31227	IH0045	Bryan	2	44811	0.158	62.47	8.53	66.5	19.62	17.54	89.4

Table 28. Top 10 Segments with High Estimated PVTT Values on 65-mph Rural Freeways Based on 2018 Data.

#	Unique _id	Hwy	District	Lanes	AADT	Ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	7400	IH0010	Beaumont	2	45439	0.056	59.54	7.71	63.5	112.87	79.81	70.71
2	5837	IH0010	Beaumont	2	45439	0.054	59.54	7.71	63.5	112.87	79.81	70.71
3	21161	IH0035	Waco	2	63306	0.034	59.71	7.72	63.94	8.18	4.53	55.4
4	26631	IH0035	Waco	2	62624	0.122	61.8	6.08	64.98	17.63	8.68	49.22
5	17957	IH0035	Waco	2	62624	0.118	62.35	6.96	65.5	172.61	84.67	49.05
6	19146	IH0035	Waco	2	62624	4.626	62.35	6.96	65.5	172.61	84.67	49.05
7	19147	IH0035	Waco	2	62624	4.626	62.35	6.96	65.5	172.61	84.67	49.05
8	27617	IH0035	Waco	2	62624	4.624	62.35	6.96	65.5	172.61	84.67	49.05
9	23963	IH0035	Waco	2	62624	0.126	62.35	6.96	65.5	172.61	84.67	49.05
10	27616	IH0035	Waco	2	62624	4.624	62.35	6.96	65.5	172.61	84.67	49.05

8.3.2.1. Rural Freeways

Table 29 to Table 32 list the top 10 segments using 2018 data based on the high estimated PVTT values on urban freeways with different PSLs. The segments with high PVTT measures (PSL higher than 65 mph) are mostly in the Dallas and Houston Districts. Locations with higher PVTT values on 65 mph urban freeways are in the El Paso District. Examining the number of lanes on these urban freeways reveals a range from 2 to 5, which implies a diversity in roadway capacity and potentially varying levels of congestion in different segments. The number of lanes plays a crucial role in determining the capacity of a roadway to handle traffic volume. Generally, a higher number of lanes suggests a greater potential for accommodating vehicles and facilitating smoother traffic flow. However, this does not guarantee immunity from congestion since other factors such as merging points, interchanges, and bottlenecks can still impact travel time variability.

Table 29. Top 10 Segments with High Estimated PVTT Values on 80-mph Urban Freeways Based on 2018 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	34527	IH0069	Houston	4	139504	0.64	62.71	12.52	71.42	86.3	90.47	100.53
2	29786	IH0045	Houston	4	208346	0.208	49.54	18.7	66.65	39.91	37.96	94.8
3	32067	IH0069	Houston	5	220338	1.54	56.5	16.15	68.28	69.89	67.4	93.46
4	37524	IH0635	Dallas	4	202743	0.052	62.29	16.34	74	322.25	293.82	91.18
5	36430	IH0635	Dallas	4	178097	0.052	62.29	16.34	74	322.25	293.82	91.18
6	35264	IH0635	Dallas	4	178097	0.152	62.29	16.34	74	322.25	293.82	91.18
7	30010	IH0045	Houston	4	208346	1.06	50.55	18.13	66.69	23.59	22.28	90.68
8	30158	IH0045	Houston	4	221319	0.064	63.27	14.38	74	449.54	406.18	90.36
9	34093	IH0045	Houston	4	160783	0.268	66.29	10.31	72.55	48.94	59.15	88.76
10	32670	IH0045	Houston	4	193601	1.572	66.29	10.31	72.55	48.94	59.15	88.76

Table 30. Top 10 Segments with High Estimated PVTT Values on 75-mph Urban Freeways Based on 2018 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	30187	IH0069	Houston	5	220338	0.155	49.33	20.87	66.61	67.1	75.58	112.57
2	34335	IH0069	Houston	5	220338	0.49	49.33	20.87	66.61	67.1	75.58	112.57
3	34075	IH0069	Houston	5	220338	0.18	49.33	20.87	66.61	67.1	75.58	112.57
4	34598	IH0069	Houston	5	220338	0.21	54.55	16.86	66.95	76.74	86.1	107.08
5	26829	IH0035W	Fort Worth	2	64697	0.022	61.49	14.16	68.77	25.52	26.59	104.19
6	37432	IH0610	Houston	3	167951	0.87	48.08	21.51	65.42	7.65	7.5	99.97
7	15890	IH0020	Dallas	2	50696	0.166	65.57	8.8	69.32	59.86	59	98.56
8	16402	IH0020	Dallas	2	61768	2.482	65.57	8.8	69.32	59.86	59	98.56
9	13465	IH0020	Dallas	2	61768	0.222	65.57	8.8	69.32	59.86	59	98.56
10	16173	IH0020	Dallas	2	50696	0.218	65.57	8.8	69.32	59.86	59	98.56

Table 31. Top 10 Segments with High Estimated PVTT Values on 70-mph Urban Freeways Based on 2018 Data.

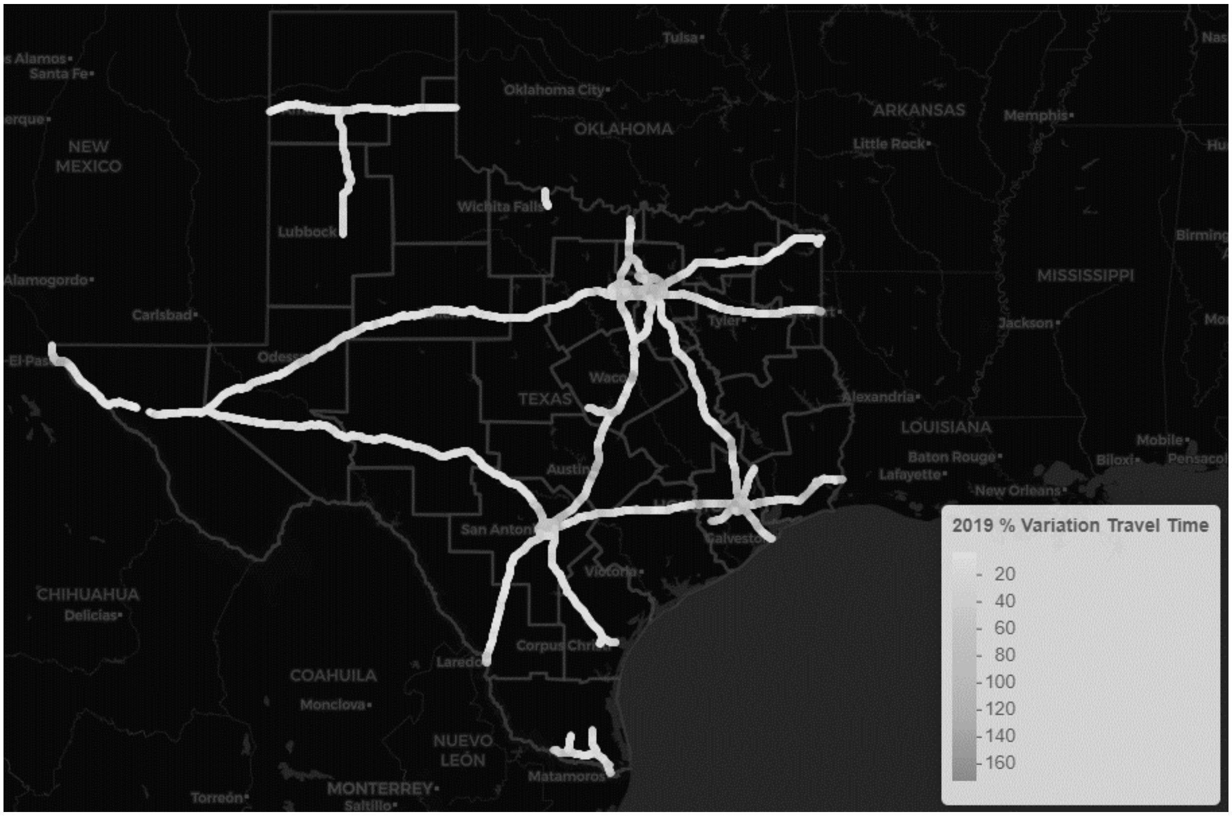
#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	22286	IH0035	Austin	2	178313	0.044	43.84	20.55	63	13.8	16.72	121.08
2	28971	IH0035	Austin	2	178313	0.046	43.84	20.55	63	13.8	16.72	121.08
3	9289	IH0010	San Antonio	4	69363	0.104	62.37	11.91	68.41	25.46	29.14	114.45
4	33866	IH0069	Houston	4	138890	0.044	45.41	21.07	63	1.92	2.16	112.22
5	29843	IH0069	Houston	4	91710	0.088	45.41	21.07	63	1.92	2.16	112.22
6	26032	IH0035	Austin	3	138205	1.455	58.85	11.71	65.47	7.52	8.19	108.94
7	32200	IH0045	Houston	3	208346	0.249	50.6	17.37	63.5	74.26	78.69	105.96
8	30994	IH0045	Houston	3	156215	0.096	50.6	17.37	63.5	74.26	78.69	105.96
9	31528	IH0045	Houston	3	156215	0.033	50.6	17.37	63.5	74.26	78.69	105.96
10	34272	IH0069	Houston	4	139504	0.068	48	19.52	63.86	165.8	169.81	102.42

Table 32. Top 10 Segments with High Estimated PVTT Values on 65-mph Urban Freeways Based on 2018 Data.

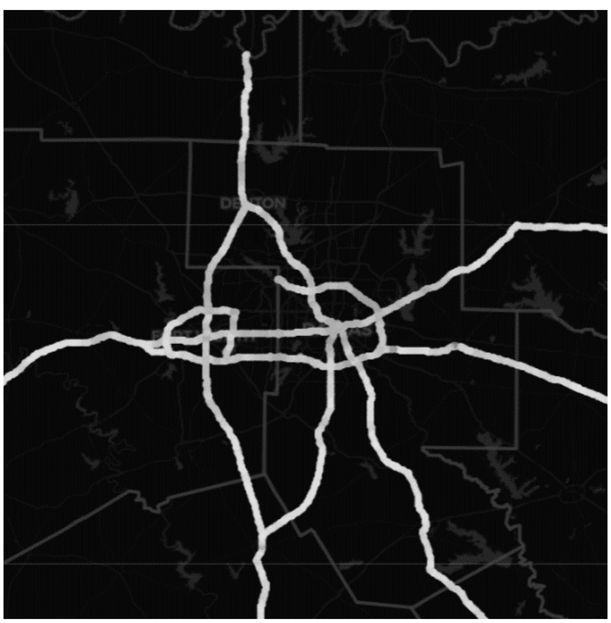
#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	7574	IH0010	El Paso	3	131725	0.36	55.55	12.23	62.46	127.76	141.47	110.73
2	5176	IH0010	El Paso	3	131725	0.309	55.55	12.23	62.46	127.76	141.47	110.73
3	6309	IH0010	El Paso	4	131725	0.264	55.55	12.23	62.46	127.76	141.47	110.73
4	9228	IH0010	El Paso	3	131725	0.207	55.55	12.23	62.46	127.76	141.47	110.73
5	9079	IH0010	El Paso	3	131725	1.344	55.55	12.23	62.46	127.76	141.47	110.73
6	25361	IH0035	Austin	3	129685	0.075	38.06	21.19	60.5	14.68	14.28	97.31
7	6671	IH0010	El Paso	2	102990	0.05	54.65	9.12	60.71	27.69	25.51	95.17
8	6972	IH0010	El Paso	2	102990	0.058	54.65	9.12	60.71	27.69	25.51	95.17
9	10225	IH0010	El Paso	2	102990	0.272	54.65	9.12	60.71	27.69	25.51	95.17
10	10531	IH0010	El Paso	2	102990	0.056	54.65	9.12	60.71	27.69	25.51	95.17

8.3.2 Segments of Interest Based on 2019 Data

The heatmap in Figure 36 illustrates the estimated PVTT measures for different segments using data from 2019. Specifically, Figure 36(a), 36(b), and 36(c) present color-coded segments representing the PVTT measures for Texas, Dallas, and Houston, respectively.



(a) Texas



(b) Dallas **(c) Houston**

Figure 36. Estimated PVTT Measures Using 2019 Data.

8.3.2.1. Rural Freeways

Using 2019 data, Table 33 to Table 36 list the top 10 segments based on the high estimated PVTT values on rural freeways with different PSLs. It is seen that high PVTT measures are primarily concentrated in the districts of Houston, with one notable exception being 65 mph rural freeways. The Houston Districts experience elevated PVTT values, indicating a higher degree of travel time variability and potential traffic congestion. Notably, when considering rural freeways, the PVTT values differ based on the PSLs. In the case of 65 mph rural freeways, the PVTT values are relatively lower than other rural freeways with higher PSLs. In terms of lane configurations, most rural freeways typically consist of two lanes, excluding rural freeways with a PSL of 65 mph. The prevalence of two-lane rural freeways implies a relatively narrower roadway capacity, which may contribute to potential traffic congestion during peak periods or under high traffic demand. Additionally, recognizing the impact of speed limits on PVTT values underscores the potential benefits of adjusting speed limits on rural freeways to improve travel time consistency and reduce congestion. Moreover, considering the number of lanes on rural freeways aids in assessing the capacity limitations and identifying potential areas for infrastructure enhancements or lane expansions.

Table 33. Top 10 Segments with High Estimated PVTT Values on 80-mph Rural Freeways Based on 2019 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	8563	IH0010	Yoakum	2	56083	0.128	61.28	11.48	66.76	199.24	215.11	107.97
2	8778	IH0010	Houston	2	56083	0.376	59.83	9.37	64.72	141.3	134.01	94.96
3	5404	IH0010	Houston	2	56083	0.046	60.3	9.85	65.45	151.27	141.16	93.32
4	6643	IH0010	Houston	2	50409	0.054	60.3	9.85	65.45	151.27	141.16	93.32
5	9874	IH0010	Yoakum	2	49711	0.56	65.12	9.12	69.35	33	28.36	85.36
6	8061	IH0010	Houston	2	56083	0.702	60.59	8.54	65.11	150.55	121.23	82.58
7	5348	IH0010	Houston	2	50409	0.896	60.19	8.39	64.76	109	91	80.75
8	8958	IH0010	Houston	2	50409	0.306	61.92	7.81	65.94	65.43	48.36	73.91
9	10212	IH0010	Houston	2	50409	0.088	61.92	7.81	65.94	65.43	48.36	73.91
10	9483	IH0010	Houston	2	50409	0.142	61.92	7.81	65.94	65.43	48.36	73.91

Table 34. Top 10 Segments with High Estimated PVTT Values on 75-mph Rural Freeways Based on 2019 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	5922	IH0010	Yoakum	2	56083	0.206	61.28	11.48	66.76	199.24	215.11	107.97
2	6702	IH0010	Houston	2	56083	0.06	61.28	11.48	66.76	199.24	215.11	107.97
3	7499	IH0010	Houston	2	56083	0.182	61.28	11.48	66.76	199.24	215.11	107.97
4	5319	IH0010	Houston	2	56083	0.146	61.28	11.48	66.76	199.24	215.11	107.97
5	7210	IH0010	Houston	2	56083	0.186	61.28	11.48	66.76	199.24	215.11	107.97
6	7000	IH0010	Houston	2	56083	0.252	61.28	11.48	66.76	199.24	215.11	107.97
7	15123	IH0020	Dallas	2	31705	0.434	61.13	14.92	71.42	48.65	57.22	106.69
8	6494	IH0010	El Paso	2	15879	1.616	67.18	11.16	74.17	264.06	18.93	99.17
9	6447	IH0010	Houston	2	56083	0.436	59.36	8.89	63.98	131.33	126.87	96.6
10	8225	IH0010	Houston	2	56083	0.436	59.83	9.37	64.72	141.3	134.01	94.96

Table 35. Top 10 Segments with High Estimated PVTT Values on 70-mph Rural Freeways Based on 2019 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	5628	IH0010	Houston	2	56083	0.25	61.28	11.48	66.76	199.24	215.11	107.9
2	9066	IH0010	Houston	2	56083	0.06	59.36	8.89	63.98	131.33	126.87	96.6
3	10817	IH0010	Houston	2	50409	0.402	59.36	8.89	63.98	131.33	126.87	96.6
4	11023	IH0010	Houston	2	50409	0.22	59.83	9.37	64.72	141.3	134.01	94.96
5	7387	IH0010	Houston	2	56083	0.06	59.83	9.37	64.72	141.3	134.01	94.96
6	5264	IH0010	Houston	2	56083	0.378	59.83	9.37	64.72	141.3	134.01	94.96
7	9259	IH0010	Houston	2	56083	0.576	59.83	9.37	64.72	141.3	134.01	94.96
8	6031	IH0010	Houston	2	50409	0.38	59.83	9.37	64.72	141.3	134.01	94.96
9	9995	IH0010	Houston	2	56083	0.592	59.83	9.37	64.72	141.3	134.01	94.96
10	5725	IH0010	Houston	2	56083	0.044	60.3	9.85	65.45	151.27	141.16	93.32

Table 36. Top 10 Segments with High Estimated PVTT Values on 65-mph Rural Freeways Based on 2019 Data.

#	unique_id	hwy	district	lanes	AADT	ln_miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	16949	IH0020	Dallas	2	31705	0.128	64.14	8.32	68.99	230.44	90.69	39.36
2	16422	IH0020	Brownwood	2	19142	0.562	68.24	4.94	71.94	93.45	23.64	25.49
3	9647	IH0010	Beaumont	3	55080	0.741	66.68	3.18	68.75	140.33	33.63	23.73
4	6680	IH0010	El Paso	2	15100	0.468	62.36	5.76	67	44.37	9.87	22.24
5	9407	IH0010	San Angelo	4	9705	11.228	69.56	5.46	74.22	175.41	35.43	20.21
6	8405	IH0010	Odessa	4	7770	0.692	69.92	5.91	74.74	359.42	70.71	19.62
7	8284	IH0010	Yoakum	2	28582	0.258	68.43	3.67	70.99	143.81	28.08	19.53
8	16341	IH0020	Odessa	2	20700	3.764	69.25	4.67	72.46	110.64	21.63	19.44
9	5442	IH0010	Beaumont	3	55080	3.549	67.64	3.08	69.74	47.34	9.54	19.16
10	11785	IH0020	Abilene	2	19056	2.418	69.15	4.11	72.65	32.27	5.13	15.68

8.3.2.2. Urban Freeways

Using 2019 data, Table 37 to Table 40 list the top 10 segments based on the high estimated PVTT values on urban freeways with different PSLs. Primarily, the segments exhibiting high PVTT measures are concentrated in districts encompassing major metropolitan areas such as Dallas, Houston, Waco, Austin, and El Paso, which indicates that these urban regions face considerable challenges in terms of travel time variability and potential traffic congestion. Furthermore, an analysis of the number of lanes on urban freeways demonstrates a range of two to four lanes. This variation in lane capacity suggests differing levels of roadway infrastructure and potential implications for traffic flow. Although freeways with more lanes generally have a higher potential for accommodating vehicles and facilitating smoother traffic movement, congestion can still occur due to other factors such as bottlenecks, interchanges, or inadequate merging points.

Table 37. Top 10 Segments with High Estimated PVTT Values on 80-mph Urban Freeways Based on 2019 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	20632	IH0035	Waco	3	77498	0.048	63.29	9.57	67.5	27.14	26.12	96.27
2	15730	IH0020	Dallas	3	44293	1.944	64.88	12.24	72.99	67.97	59.68	92.12
3	20504	IH0035	Austin	3	129685	0.066	40.85	19.31	60.98	23.21	21.14	82.69
4	2489	IH0030	Dallas	2	10710	0.324	47.29	15.81	59	19.65	16.01	81.47
5	10938	IH0010	El Paso	3	131725	2.16	59.8	7.97	64.16	106.87	84.61	76.96
6	27399	IH0035W	Fort Worth	2	79171	0.07	56.98	16.31	67.63	26.91	20.35	75.64
7	14261	IH0020	Dallas	4	141084	0.84	62.9	7.81	66.72	39.39	30.78	73.78
8	13339	IH0020	Dallas	4	122147	2.064	65.49	6.26	68.34	54.24	41.71	70.24
9	7674	IH0010	Beaumont	2	59091	0.298	60.47	6.29	64.5	24.2	16.86	68.9
10	9796	IH0010	San Antonio	2	69363	0.83	53.57	15.26	64.94	172.98	119	68.79

Table 38. Top 10 Segments with High Estimated PVTT Values on 75-mph Urban Freeways Based on 2019 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	17607	IH0020	Dallas	3	44293	1.125	63.8	19.05	76.22	60.45	103.68	160.65
2	19344	IH0035	Laredo	3	43529	0.039	30.77	17.26	50	14.99	16.58	110.62
3	11024	IH0010	Houston	3	70110	0.03	60.55	14.1	67.5	31.02	33.11	106.74
4	7574	IH0010	El Paso	3	131725	0.36	58.84	10.41	64	115.19	121.37	105.36
5	9079	IH0010	El Paso	3	131725	1.344	58.84	10.41	64	115.19	121.37	105.36
6	17990	IH0035	Waco	3	115484	0.405	54.09	14.86	65	45.63	46.23	101.3
7	15167	IH0020	Dallas	3	44293	1.011	64.5	13.23	73.26	42.47	52.26	97.3
8	18835	IH0035	Waco	3	115484	0.141	54.28	14.5	64.24	22.17	20.17	92.32
9	20972	IH0035	Waco	3	115484	0.36	54.35	14.45	64.28	21.73	19.78	92.22
10	4976	IH0010	San Antonio	2	159210	0.076	50.13	14.66	60.17	26.64	24.09	90.42

Table 39. Top 10 Segments with High Estimated PVTT Values on 70-mph Urban Freeways Based on 2019 Data.

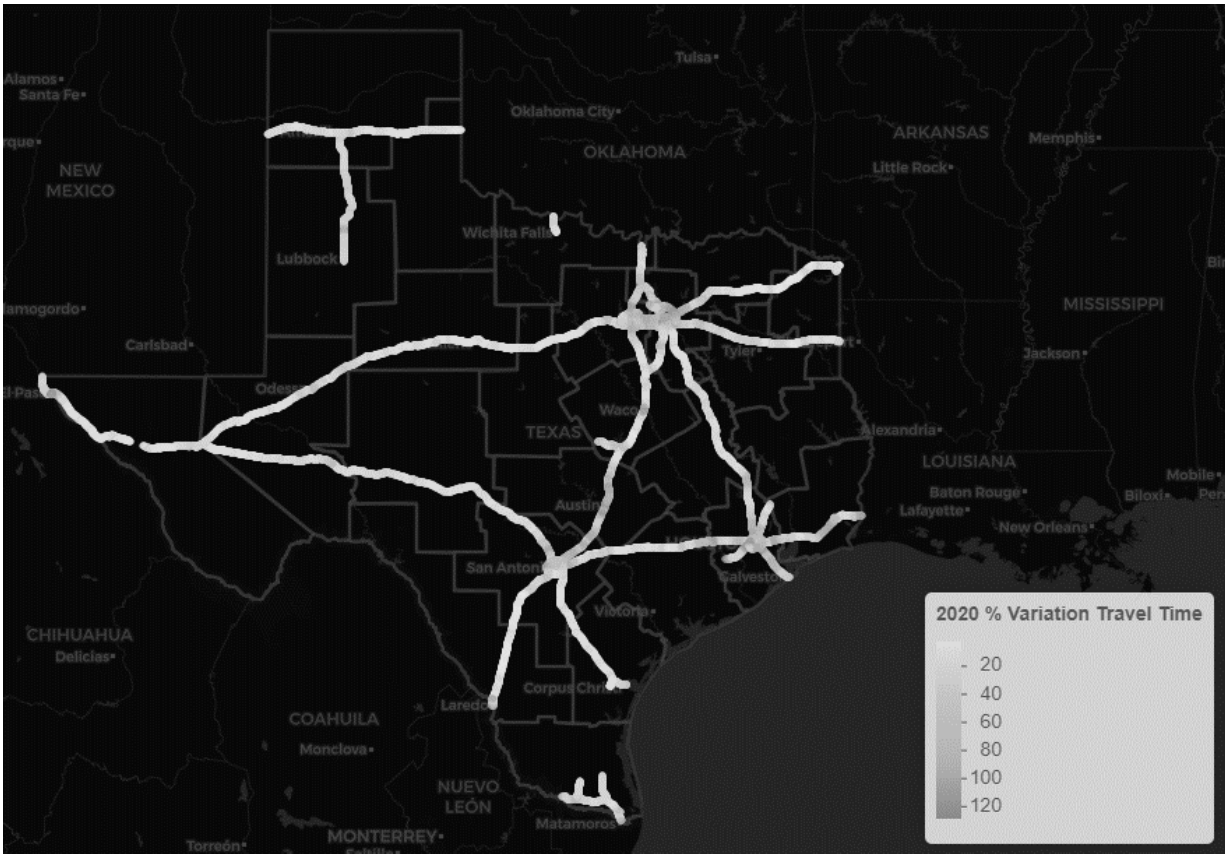
#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	15405	IH0020	Dallas	2	30029	0.016	59.5	21.15	74	58.47	100.56	171.99
2	13633	IH0020	Dallas	3	44293	2.046	64.38	17.15	75.64	88.57	99.13	144.54
3	7031	IH0010	Houston	3	99077	0.15	57.08	17.04	66.47	92.49	119.11	128.79
4	14922	IH0020	Dallas	4	122147	0.048	63.85	9.18	67.5	157.19	181.15	115.25
5	12571	IH0020	Dallas	4	143152	0.5	63.31	10.12	67.94	44.22	48.36	109.35
6	15784	IH0020	Dallas	4	141084	0.84	63.31	10.12	67.94	44.22	48.36	109.35
7	15107	IH0020	Dallas	4	141084	0.136	63.31	10.12	67.94	44.22	48.36	109.35
8	13767	IH0020	Dallas	2	30029	0.032	61.02	15.2	71.5	46.01	58.04	109.13
9	5383	IH0010	Houston	3	70110	0.051	60.55	14.1	67.5	31.02	33.11	106.74
10	8257	IH0010	Houston	3	70110	0.057	60.55	14.1	67.5	31.02	33.11	106.74

Table 40. Top 10 Segments with High Estimated PVTT Values on 65-mph Urban Freeways Based on 2019 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	12417	IH0020	Dallas	4	143152	0.248	63.31	10.12	67.94	44.22	48.36	109.35
2	3729	IH0002	Pharr	3	59774	0.057	54.83	12.32	64	30.76	26.09	84.8
3	20351	IH0035	Austin	3	129685	0.102	41.16	19.24	60.97	69.56	59.42	83.7
4	21063	IH0035	Austin	3	129685	0.927	41.69	18.76	61.07	47.2	40.23	83.27
5	18193	IH0035	Waco	3	48755	2.283	64.42	12.51	73.87	78.78	89.22	81.96
6	17215	IH0020	Dallas	4	149140	0.196	63.26	11.54	69	76.59	61.42	80.2
7	22048	IH0035	Laredo	3	43529	0.147	37.28	14.24	52.3	12.32	11.01	78.46
8	8096	IH0010	San Antonio	2	69363	0.114	56	14.46	65	131.68	103.31	78.45
9	8333	IH0010	El Paso	3	131725	2.115	59.8	7.97	64.16	106.87	84.61	76.96
10	18821	IH0035	San Antonio	2	164373	0.7	53.1	12	61.93	14.65	11.61	76.27

8.3.3 Segments of Interest Based on 2020 Data

The heatmap in Figure 37 depicts the segments and their estimated PVTT measures using data from 2020. Figure 37(a), 37(b), and 37(c) illustrate the color-coded segments representing PVTT measures for Texas, Dallas, and Houston, respectively.



(a) Texas

(b) Dallas **(c) Houston**

Figure 37. Estimated PVTT Measures Using 2020 Data.

8.3.3.1. Rural Freeways

Using 2020 data, Table 41 to Table 44 list the top 10 segments based on the high estimated PVTT values on rural freeways with different PSLs. The districts of Odessa, Yoakum, Houston, and Waco emerge as the dominant districts for rural freeways, with PSLs of 80 mph, 75 mph, 70 mph, and 65 mph, respectively. This distribution of dominant districts raises questions about the underlying factors contributing to these patterns. Factors such as geographic location, population density, economic activities, and transportation infrastructure development may play significant roles in determining the speed limits set for rural freeways in these districts. Furthermore, it is noteworthy that most rural freeways, regardless of the PSL, consist of two lanes. This observation indicates a relatively limited roadway capacity for accommodating traffic volume on these rural freeways, which can have implications for traffic flow and congestion levels, particularly during peak travel periods or when faced with high traffic demand.

Table 41. Top 10 Segments with High Estimated PVTT Values on 80-mph Rural Freeways Based on 2020 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	7925	IH0010	Beaumont	2	45439	0.104	63.05	7.28	67.11	2.58	1.68	64.93
2	10691	IH0010	Beaumont	2	45439	0.104	63.05	7.28	67.11	2.58	1.68	64.93
3	9371	IH0010	San Angelo	2	12969	0.342	68.9	8.4	74.5	27.33	14.16	51.81
4	10415	IH0010	San Angelo	4	12969	0.996	68.9	8.4	74.5	27.33	14.16	51.81
5	7490	IH0010	San Angelo	2	12969	0.338	68.9	8.4	74.5	27.33	14.16	51.81
6	10323	IH0010	Odessa	2	7770	0.074	69.33	6.2	73.99	368.97	150.48	40.78
7	7063	IH0010	Odessa	2	7770	0.45	69.33	6.2	73.99	368.97	150.48	40.78
8	7580	IH0010	Odessa	2	7770	0.05	69.33	6.2	73.99	368.97	150.48	40.78
9	6145	IH0010	Odessa	2	7770	0.882	69.33	6.2	73.99	368.97	150.48	40.78
10	7730	IH0010	Odessa	2	7770	0.364	69.33	6.2	73.99	368.97	150.48	40.78

Table 42. Top 10 Segments with High Estimated PVTT Values on 75-mph Rural Freeways Based on 2020 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	11044	IH0010	Yoakum	2	40152	0.174	67.04	6.97	70.41	21.49	15.1	70.28
2	5315	IH0010	Austin	2	32632	0.018	66.63	6.87	70	2.02	1.33	65.78
3	9874	IH0010	Yoakum	2	49711	0.56	64.78	6	67.85	31.2	18.9	60.41
4	9535	IH0010	Yoakum	2	49711	1.238	64.14	6.38	67.75	31.68	18.43	58.34
5	10113	IH0010	Yoakum	2	56083	0.332	64.59	6.49	68.32	48.91	28.17	57.59
6	10620	IH0010	Yoakum	2	39831	4.122	65.74	6.13	69.5	48.86	26.1	55.86
7	9315	IH0010	Yoakum	2	56083	3.208	64.27	6.44	68.4	38.73	20.12	52.36
8	6234	IH0010	Yoakum	2	56083	3.23	64.22	6.41	68.31	44.95	23.08	52.2
9	19475	IH0027	Lubbock	2	14770	13.074	69.42	8.01	75.47	183.61	93.45	50.91
10	23216	IH0027	Lubbock	2	14770	13.07	69.42	8.01	75.47	183.61	93.45	50.91

Table 43. Top 10 Segments with High Estimated PVTT Values on 70-mph Rural Freeways Based on 2020 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	5725	IH0010	Houston	2	56083	0.044	61.61	7.97	65.5	139.91	124.15	88.73
2	7911	IH0010	Houston	2	56083	0.03	61.61	7.97	65.5	139.91	124.15	88.73
3	8511	IH0010	Houston	2	56083	0.044	61.61	7.97	65.5	139.91	124.15	88.73
4	9164	IH0010	Houston	2	56083	0.044	61.61	7.97	65.5	139.91	124.15	88.73
5	5404	IH0010	Houston	2	56083	0.046	61.61	7.97	65.5	139.91	124.15	88.73
6	6643	IH0010	Houston	2	50409	0.054	61.61	7.97	65.5	139.91	124.15	88.73
7	10507	IH0010	Houston	2	56083	0.016	61.61	7.97	65.5	139.91	124.15	88.73
8	8225	IH0010	Houston	2	56083	0.436	61.86	6.76	65.48	127.92	100.18	77.24
9	8280	IH0010	Houston	2	56083	0.59	61.86	6.76	65.48	127.92	100.18	77.24
10	7810	IH0010	Houston	2	56083	0.378	61.86	6.76	65.48	127.92	100.18	77.24

Table 44. Top 10 Segments with High Estimated PVTT Values on 65-mph Rural Freeways Based on 2020 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	7400	IH0010	Beaumont	2	45439	0.056	60.25	6.46	64.47	105.92	58.87	55.58
2	5837	IH0010	Beaumont	2	45439	0.054	60.25	6.46	64.47	105.92	58.87	55.58
3	10805	IH0010	San Antonio	4	16546	0.156	68.14	4.1	71.5	21.08	2.54	12.05
4	19389	IH0035	Waco	2	62624	0.248	68.04	2.55	69.97	104.8	11.05	10.03
5	20211	IH0035	Waco	2	62624	0.262	68.04	2.55	69.97	104.8	11.05	10.03
6	28003	IH0035	Waco	2	62624	0.178	68.04	2.55	69.97	104.8	11.05	10.03
7	26121	IH0035	Waco	2	62624	0.086	68.04	2.55	69.97	104.8	11.05	10.03
8	21161	IH0035	Waco	2	63306	0.034	65.22	2.61	67.41	18.78	1.87	9.95
9	27360	IH0035	Waco	2	62624	1.138	68.06	2.59	69.98	124.95	12.06	9.44
10	22942	IH0035	Waco	2	62624	0.902	68.07	2.62	69.98	136.52	12.63	9.11

8.3.3.2. Urban Freeways

Table 45 to Table 48, using 2020 data, list the top 10 segments based on the high estimated PVTT values on urban freeways with different PSLs. Notably, the segments with elevated PVTT measures predominantly belong to the Houston, Dallas, and El Paso Districts, indicating potential areas of concern in terms of travel time variations. The number of lanes for urban freeways ranges from two to five.

Table 45. Top 10 Segments with High Estimated PVTT Values on 80-mph Urban Freeways Based on 2020 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	32704	IH0069	Houston	5	255714	0.1	56.55	20.67	74	32.15	35.35	109.97
2	35196	IH0635	Dallas	4	176541	0.044	58.86	20.3	74.23	10.25	10.67	108.91
3	33835	IH0069	Houston	5	171747	0.095	68.97	11.46	77	320.09	347.6	108.6
4	32131	IH0045	Houston	4	208346	0.276	60.03	14.22	71.45	45.2	48.89	108.49
5	34469	IH0045	Houston	4	208346	0.396	60.03	14.22	71.45	45.2	48.89	108.49
6	29584	IH0045	Houston	3	156215	0.042	55.83	13.83	66.48	65.45	66.95	101.41
7	30010	IH0045	Houston	4	208346	1.06	56.94	17.12	70.5	29.69	27.93	93.69
8	30158	IH0045	Houston	4	221319	0.064	68.69	13.69	78	361.21	337.9	93.55
9	1468	IH0045	Houston	2	9600	0.112	63.29	15.08	74.95	66.67	60.63	90.94
10	29786	IH0045	Houston	4	208346	0.208	56.44	17.45	70.9	25.63	23.18	89.9

Table 46. Top 10 Segments with High Estimated PVTT Values on 75-mph Urban Freeways Based on 2020 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	29913	IH0069	Houston	3	159452	0.027	59.73	14.71	70	83.78	94.35	112.61
2	37432	IH0610	Houston	3	167951	0.87	61.76	13.57	70.18	26.27	28.45	108.74
3	31953	IH0045	Houston	4	173504	0.116	57.12	14.42	67.57	14.74	15.24	104.43
4	33503	IH0069	Houston	5	220338	0.08	58.69	18.21	71.4	46.26	46.74	102.05
5	32958	IH0069	Houston	5	220338	0.07	58.69	18.21	71.4	46.26	46.74	102.05
6	31837	IH0069	Houston	5	220338	0.09	58.69	18.21	71.4	46.26	46.74	102.05
7	30741	IH0045	Houston	4	208346	0.96	57.79	16.63	70.17	30.18	30.41	101.47
8	30187	IH0069	Houston	5	220338	0.155	60.21	15.74	71.45	48.89	48.22	101.17
9	34335	IH0069	Houston	5	220338	0.49	60.21	15.74	71.45	48.89	48.22	101.17
10	34075	IH0069	Houston	5	220338	0.18	60.21	15.74	71.45	48.89	48.22	101.17

Table 47. Top 10 Segments with High Estimated PVTT Values on 70-mph Urban Freeways Based on 2020 Data.

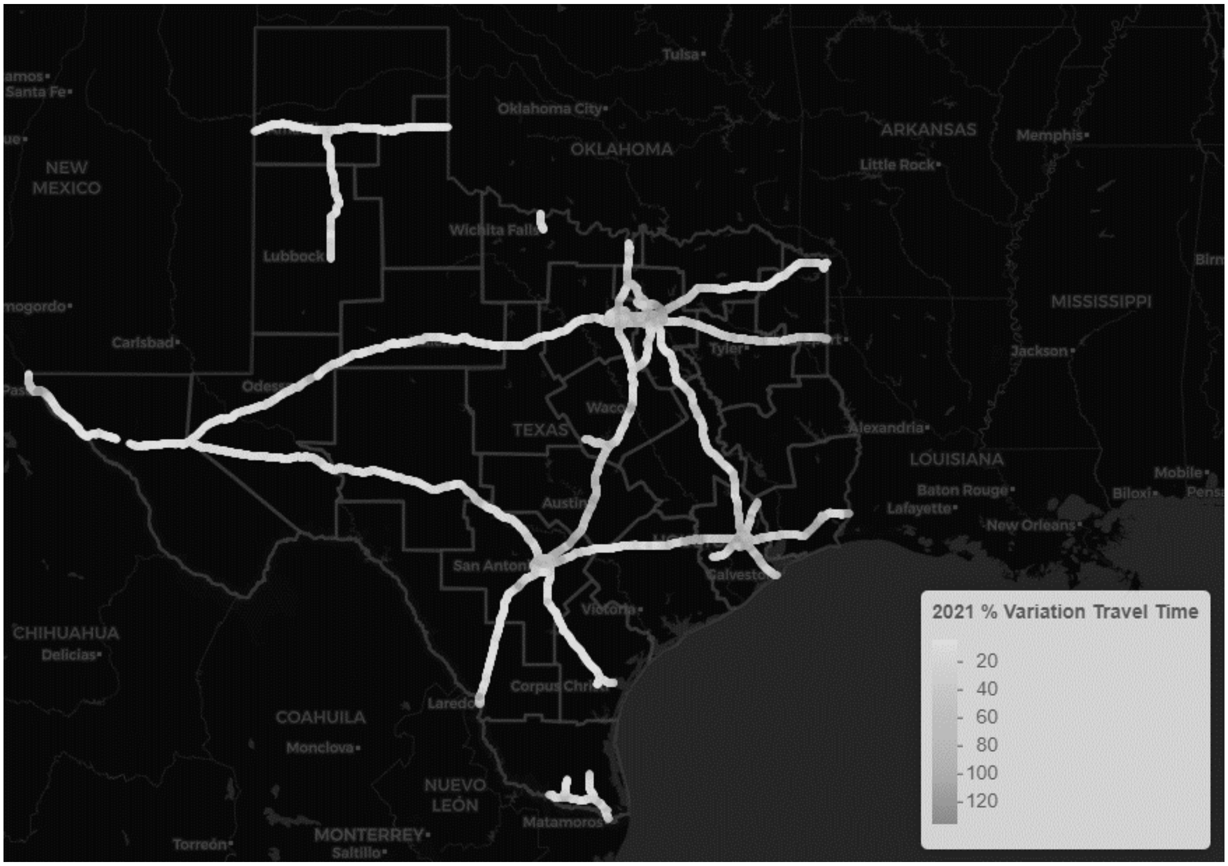
#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	31386	IH0069	Houston	2	24897	0.678	61.18	17.17	73	165.91	189.9	114.46
2	33936	IH0069	Houston	2	24897	0.174	61.18	17.17	73	165.91	189.9	114.46
3	33386	IH0069	Houston	2	24897	3.118	61.18	17.17	73	165.91	189.9	114.46
4	36240	IH0345	Dallas	3	162578	0.474	53.39	15.33	64.15	16.24	18.23	113.15
5	20632	IH0035	Waco	3	77498	0.048	58.37	11.6	63.98	31.51	35.02	111.15
6	2443	IH0345	Dallas	2	32463	0.244	53.19	15.27	63.98	17.49	19.4	110.88
7	2573	IH0345	Dallas	2	32463	0.696	53.19	15.27	63.98	17.49	19.4	110.88
8	32782	IH0069	Houston	3	159452	0.087	56.63	15.93	67.76	64.15	72.12	109.75
9	7031	IH0010	Houston	3	99077	0.15	62.96	9.32	66.99	64.62	70.89	109.71
10	33169	IH0069	Houston	4	159452	0.52	56.52	15.76	67.43	60.57	67.57	108.92

Table 48. Top 10 Segments with High Estimated PVTT Values on 65-mph Urban Freeways Based on 2020 Data.

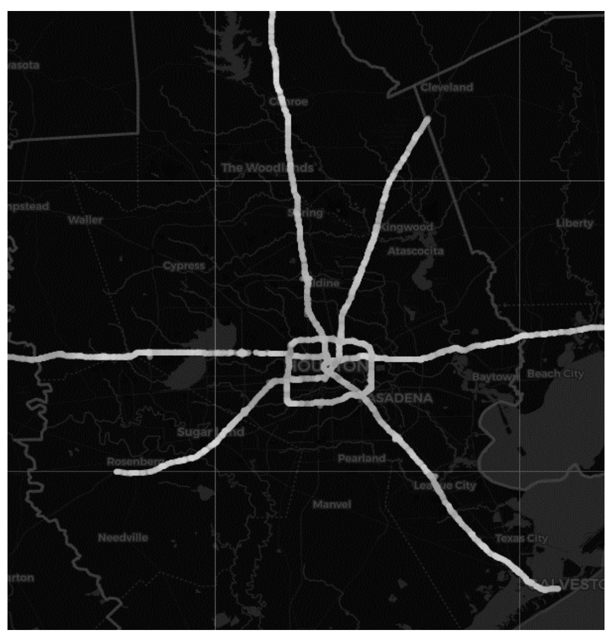
#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	10110	IH0010	Houston	3	152983	0.294	56.68	7.58	60.9	16.04	16.78	105.05
2	35671	IH0110	El Paso	3	32405	0.228	37.49	15.16	52.1	25.06	25.33	104.6
3	8853	IH0010	Houston	3	152983	0.156	57.15	7.73	61.46	20.37	21.18	103.96
4	2530	IH0010	El Paso	2	6945	0.388	56.3	13.3	67.81	27.61	34.38	96.16
5	24990	IH0035E	Dallas	4	161151	0.788	57.45	9.62	63.44	20.46	19.35	94.57
6	38338	IH0110	El Paso	2	32405	0.048	41.82	10.5	50.24	58.39	43.79	93.23
7	37938	IH0110	El Paso	2	43125	0.266	41.82	10.5	50.24	58.39	43.79	93.23
8	37703	IH0110	El Paso	2	43125	0.134	41.82	10.5	50.24	58.39	43.79	93.23
9	25361	IH0035	Austin	3	129685	0.075	44.88	19.07	61.5	27.91	25.92	92.87
10	37295	IH0345	Dallas	5	162578	0.685	50.65	13.57	60.25	10.52	9.91	88.38

8.3.4 Segments of Interest Based on 2021 Data

Figure 38, using 2021 data, depicts a heatmap illustrating the estimated PVTT measures for segments. Subsequently, Figure 38(a), 38(b), and 38(c) showcase the color-coded segments, representing the PVTT measures for Texas, Dallas, and Houston, correspondingly.



(a) Texas



(b) Dallas **(c) Houston**

Figure 38. Estimated PVTT Measures Using 2021 Data.

8.3.4.1. Rural Freeways

Table 49 to Table 52, considering the PSLs and utilizing data from 2021, provide an overview of the highest estimated PVTT values for the top 10 segments of rural freeways. The analysis reveals that the segments with elevated PVTT measures are primarily located within the San Angelo, Yoakum, Beaumont, and Waco Districts. It is noteworthy that most rural freeways consist of two lanes, indicating a consistent trend across the analyzed areas.

Table 49. Top 10 Segments with High Estimated PVTT Values on 80-mph Rural Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	7925	IH0010	Beaumont	2	45439	0.104	62.47	8.41	66.78	2.71	2.41	88.87
2	10691	IH0010	Beaumont	2	45439	0.104	62.47	8.41	66.78	2.71	2.41	88.87
3	6328	IH0010	San Angelo	2	12691	3.094	74.54	7	80.44	33.69	16.82	50.1
4	6327	IH0010	San Angelo	2	12691	3.094	74.86	6.7	80.44	29.6	14.05	49.65
5	9154	IH0010	San Angelo	2	12691	0.038	74.4	6.93	80.47	45.14	22.12	49.01
6	5715	IH0010	San Angelo	2	11476	0.044	74.4	6.93	80.47	45.14	22.12	49.01
7	8650	IH0010	San Angelo	2	11476	0.096	74.4	6.93	80.47	45.14	22.12	49.01
8	11243	IH0010	San Angelo	2	11476	1.004	74.4	6.93	80.46	48.32	23.63	48.99
9	11015	IH0010	San Angelo	2	12691	0.322	74.64	6.85	80.25	210.06	94.35	44.91
10	7743	IH0010	San Angelo	2	12691	0.672	74.64	6.85	80.25	210.06	94.35	44.91

Table 50. Top 10 Segments with High Estimated PVTT Values on 75-mph Rural Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	22843	IH0030	Paris	2	49281	0.038	68.08	6.88	71	21.27	18.73	88.05
2	9874	IH0010	Yoakum	2	49711	0.56	64.15	8.83	69.57	32.78	26.06	78.6
3	10113	IH0010	Yoakum	2	56083	0.332	62.08	10.31	68.99	53.8	42.25	78.54
4	9535	IH0010	Yoakum	2	49711	1.238	63.53	9.23	69.65	33.56	24.83	72.38
5	6234	IH0010	Yoakum	2	56083	3.23	64.34	7.86	69.28	45.95	30.8	69.22
6	9315	IH0010	Yoakum	2	56083	3.208	64.34	7.83	69.25	39.64	27.03	69.21
7	10871	IH0010	Yoakum	2	56083	0.852	63.33	8.79	69.38	52.26	35.92	68.44
8	9011	IH0010	Yoakum	2	56083	0.326	63.33	8.79	69.38	52.26	35.92	68.44
9	8719	IH0010	Yoakum	2	56083	0.84	63.33	8.79	69.38	52.8	36.23	68.43
10	134	IH0020	Tyler	1	886	0.221	71.26	6.1	74.99	24.54	16.41	66.87

Table 51. Top 10 Segments with High Estimated PVTT Values on 70-mph Rural Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	19305	IH0035	San Antonio	3	114081	0.06	66.02	8.69	70.5	44.94	42.59	94.77
2	22638	IH0035	San Antonio	3	112918	0.03	64.6	9.41	69.5	33.69	30.73	91.2
3	9885	IH0010	Beaumont	2	42114	0.026	67.89	7.09	71.99	61.69	55.44	89.86
4	11233	IH0010	Beaumont	2	41700	2.688	68.23	7.21	72.25	62.82	53.74	85.62
5	9135	IH0010	Beaumont	2	41700	2.67	68.23	7.21	72.25	62.82	53.74	85.62
6	18097	IH0035	San Antonio	3	112918	0.51	67.23	7.89	71.25	42.22	36.09	84.85
7	19520	IH0035	San Antonio	3	112918	0.621	67.23	7.89	71.25	42.22	36.09	84.85
8	17819	IH0035	San Antonio	3	112918	0.093	67.23	7.89	71.25	42.22	36.09	84.85
9	19006	IH0035	Austin	3	117230	0.036	67.48	6.97	70.77	39.47	32.91	83.4
10	9976	IH0010	Beaumont	2	42114	3.036	68.04	7.05	71.88	41.52	34.74	83.28

Table 52. Top 10 Segments with High Estimated PVTT Values on 65-mph Rural Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	7400	IH0010	Beaumont	2	45439	0.056	60.84	7.65	65	108.05	74.4	68.85
2	5837	IH0010	Beaumont	2	45439	0.054	60.84	7.65	65	108.05	74.4	68.85
3	21161	IH0035	Waco	2	63306	0.034	68.97	5.2	72.88	18.14	8.14	44.85
4	19389	IH0035	Waco	2	62624	0.248	71.16	5.19	74.75	100.77	16.15	18.78
5	20211	IH0035	Waco	2	62624	0.262	71.16	5.19	74.75	100.77	16.15	18.78
6	28003	IH0035	Waco	2	62624	0.178	71.16	5.19	74.75	100.77	16.15	18.78
7	26121	IH0035	Waco	2	62624	0.086	71.16	5.19	74.75	100.77	16.15	18.78
8	27360	IH0035	Waco	2	62624	1.138	71.23	5.04	74.75	120.1	20	18.09
9	22942	IH0035	Waco	2	62624	0.902	71.28	4.95	74.75	131.2	22.21	17.68
10	17957	IH0035	Waco	2	62624	0.118	71.31	4.89	74.75	139.42	23.85	17.39

8.3.4.2. Urban Freeways

Using 2021 data, Table 53 to Table 56 list the top 10 segments based on the high estimated PVTT values on urban freeways with different PSLs. The segments with high PVTT measures are mostly in the Dallas, Houston, Waco, and El Paso Districts. The number of lanes for urban freeways ranges from two to five. Although urban freeways with a PSL of 80 mph exhibit a distinct pattern, it is worth noting that the remaining roadways with different speed limits tend to display higher PVTT values.

Table 53. Top 10 Segments with High Estimated PVTT Values on 80-mph Urban Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	7400	IH0010	Beaumont	2	45439	0.056	60.84	7.65	65	108.05	74.4	68.85
2	5837	IH0010	Beaumont	2	45439	0.054	60.84	7.65	65	108.05	74.4	68.85
3	21161	IH0035	Waco	2	63306	0.034	68.97	5.2	72.88	18.14	8.14	44.85
4	19389	IH0035	Waco	2	62624	0.248	71.16	5.19	74.75	100.77	16.15	18.78
5	20211	IH0035	Waco	2	62624	0.262	71.16	5.19	74.75	100.77	16.15	18.78
6	28003	IH0035	Waco	2	62624	0.178	71.16	5.19	74.75	100.77	16.15	18.78
7	26121	IH0035	Waco	2	62624	0.086	71.16	5.19	74.75	100.77	16.15	18.78
8	27360	IH0035	Waco	2	62624	1.138	71.23	5.04	74.75	120.1	20	18.09
9	22942	IH0035	Waco	2	62624	0.902	71.28	4.95	74.75	131.2	22.21	17.68
10	17957	IH0035	Waco	2	62624	0.118	71.31	4.89	74.75	139.42	23.85	17.39

Table 54. Top 10 Segments with High Estimated PVTT Values on 75-mph Urban Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	36062	IH0635	Dallas	4	176541	0.224	56.79	17.52	70.9	27.66	31.22	119.07
2	31953	IH0045	Houston	4	173504	0.116	49.04	20.7	66.8	23.29	26.15	112.02
3	37008	IH0635	Dallas	4	176541	3.104	56.36	17.9	70.99	45.1	51.05	111.41
4	32468	IH0045	Houston	3	173504	0.375	50.47	19.54	66.73	17.29	19.36	110.1
5	36399	IH0635	Dallas	4	176541	0.732	57.42	16.38	70	168.54	157.42	109.63
6	30187	IH0069	Houston	5	220338	0.155	49.02	22.99	69.26	83.5	90.14	107.94
7	34335	IH0069	Houston	5	220338	0.49	49.02	22.99	69.26	83.5	90.14	107.94
8	34075	IH0069	Houston	5	220338	0.18	49.02	22.99	69.26	83.5	90.14	107.94
9	26829	IH0035W	Fort Worth	2	64697	0.022	58.45	12.68	66	27.12	29.05	107.12
10	34598	IH0069	Houston	5	220338	0.21	55.34	17.58	69.14	18.47	20.98	105.29

Table 55. Top 10 Segments with High Estimated PVTT Values on 70-mph Urban Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	36240	IH0345	Dallas	3	162578	0.474	49.44	17.6	63.17	20.03	24.33	122.01
2	2443	IH0345	Dallas	2	32463	0.244	49.25	17.56	63	21.6	26.06	120.63
3	2573	IH0345	Dallas	2	32463	0.696	49.25	17.56	63	21.6	26.06	120.63
4	6771	IH0010	Beaumont	2	42970	0.044	64.67	12.41	73	43.05	49.02	113.86
5	24925	IH0035W	Fort Worth	2	64697	0.064	55.09	15.71	65	117.2	128.31	109.48
6	36318	IH0610	Houston	5	145124	0.32	53.52	18.66	67	5.83	6.36	109.12
7	26340	IH0035W	Dallas	2	50190	0.098	60.06	14.22	68.94	87.4	95.3	109.04
8	35915	IH0635	Dallas	4	176541	0.168	56.8	16.26	69.5	59.12	58.62	108.84
9	38217	IH0635	Dallas	4	176541	0.688	56.65	16.23	69.38	33.54	35.53	108.66
10	26629	IH0035W	Fort Worth	2	142075	0.56	53.73	17.95	66.76	85.03	91.16	107.2

Table 56. Top 10 Segments with High Estimated PVTT Values on 65-mph Urban Freeways Based on 2021 Data.

#	unique _id	hwy	district	lanes	AADT	ln _miles	Spd Ave	Spd Std	Spd 85	TT Ave	TT Std	PVTT
1	24374	IH0035	Laredo	3	43529	0.033	42.11	17.4	59	46.75	58.43	124.98
2	2530	IH0010	El Paso	2	6945	0.388	34.79	20.27	52.67	99.65	112.53	118.88
3	35671	IH0110	El Paso	3	32405	0.228	39.23	14.95	52.52	22.94	23.78	108.74
4	25163	IH0035E	Dallas	3	147152	0.06	56.11	13.62	65.19	325.07	349.9	107.64
5	38338	IH0110	El Paso	2	32405	0.048	48.27	13.83	59.74	46.73	49.82	103.52
6	37938	IH0110	El Paso	2	43125	0.266	48.27	13.83	59.74	46.73	49.82	103.52
7	37703	IH0110	El Paso	2	43125	0.134	48.27	13.83	59.74	46.73	49.82	103.52
8	37295	IH0345	Dallas	5	162578	0.685	47.2	15.68	59.39	12.56	12.87	95.41
9	1076	IH0010	Houston	1	15500	0.13	47.71	16.95	59.5	77.38	72.29	93.42
10	24990	IH0035E	Dallas	4	161151	0.788	57.3	12.16	65.55	181.76	191.06	91.25

8.4 CHAPTER SUMMARY

The developed congestion-related decision support tool incorporates AI-based estimation of congestion measures for different segments. Users can access the tool's drop-down panels to select specific segments and choose the congestion measure of interest. Based on this wider range of spatial and temporal selections, the tool displays the estimated congestion measures for the chosen segments. To enhance the user experience and facilitate easy interpretation, the congestion measures are visually represented on a color-coded map. This map utilizes a color gradient ranging from yellow (indicating a low congestion number) to red (indicating a high congestion number). By observing the map, users can quickly assess the level of congestion in different areas and identify segments with varying degrees of traffic congestion. Additionally, the tool can assist transportation authorities and planners in identifying congestion hotspots and prioritizing infrastructure improvements or targeted traffic management interventions. By analyzing the congestion measures for various segments, authorities can allocate resources effectively and implement measures aimed at mitigating congestion and improving overall traffic flow.

In addition to the features of the congestion-related decision support tool, the TTI team also conducted an analysis of the top congested segments based on estimated congestion measures over multiple years. This analysis provided valuable insights and general guidelines regarding segments of interest. The research findings highlight a significant concentration of congestion in urban areas, particularly in major cities such as Houston and Dallas. This observation aligns with expectations since urban locations with high population densities and extensive transportation networks often experience elevated levels of traffic congestion. The presence of numerous economic activities, commuting patterns, and population centers in these urban areas contributes to the heightened congestion levels. However, it is essential to acknowledge that congestion measures exhibit both spatial and temporal variations. Although urban areas generally experience higher congestion levels, the specific segments affected by congestion can differ. Factors such as road network design, traffic demand, and ongoing construction or maintenance projects can influence the localized patterns of congestion within urban areas. Furthermore, the analysis has revealed that congestion measures are influenced by PSLs. Different segments with varying speed limits may exhibit distinct congestion patterns. For instance, segments with higher PSLs may have different congestion dynamics than those segments with lower speed limits. This element highlights the importance of considering PSLs as a contributing factor when evaluating congestion levels and developing targeted strategies to mitigate congestion. Some general observations on rural and urban freeways are listed below:

- Rural Freeways:
    - The prevalence of dominant districts for rural freeways with different PSLs prompts an inquiry into the factors influencing these patterns. Various factors, including geographic location, population density, economic activities, and transportation infrastructure development, likely contribute to the determination of speed limits in these districts. It is crucial to conduct a critical examination of the decision-making processes and considerations that led to the establishment of specific speed limits in each district.
    - A significant observation to highlight is that most rural freeways, regardless of the PSL, are typically designed with two lanes. This characteristic suggests that these roadways have a relatively limited capacity to accommodate high volumes of traffic. Consequently, this limited capacity can lead to potential challenges in traffic flow and increased congestion levels, especially during periods of peak travel or when confronted with high traffic demand.
- Urban Freeways:
    - The capacity of an urban roadway to handle traffic volume is heavily influenced by the number of lanes it possesses. A greater number of lanes generally indicates a higher potential for accommodating vehicles and promoting smoother traffic flow. Nevertheless, it is important to note that having more lanes does not guarantee complete immunity from congestion.
    - Various factors such as merging points, interchanges, and bottlenecks can still contribute to travel time variability despite the presence of multiple lanes.

The concentration of segments with high PVTT measures in major metropolitan areas like Houston, Dallas, and El Paso prompts a thoughtful examination of the effectiveness of current traffic management strategies and infrastructure development in these regions. It is crucial to critically analyze the underlying causes contributing to the observed variations in travel times.

CHAPTER 9: CONCLUSIONS AND RECOMMENDATIONS

9.1 INTRODUCTION

To improve the quality and effectiveness of the Texas surface transportation system, it is important to be able to predict where and when prolonged congestion will start and how it will spread, as well as to track atypical events and estimate their evolution. AI approaches provide a unique opportunity to estimate precise congestion measures by utilizing data from agency-owned sensors, third-party providers, and big enterprise data. The TxDOT 0-7131 project envisions mitigating the current research gap by conducting two major project phases. The first phase can confirm the validity of commercial data sources for planning and operations, while the second involves understanding which AI models/algorithms are most suitable for addressing TxDOT needs based on desirable use cases and data availability. Furthermore, it is important to analyze the required data models and workflows to determine whether it is sustainable to train, test, and validate the proposed AI techniques.

9.2 FINDINGS AND CONCLUSIONS

To understand state DOT practices in freeway congestion reduction using big data, the TTI team developed a survey covering topics such as data collection, analysis, big data platforms, report creation, and future needs. The primary insights are as follows:

- The most used big data sources for freeway congestion reduction are INRIX and Waze, with some agencies exploring data from Streetlight, Wejo, and others.
- Participating agencies universally apply big data to measure travel time, gather real-time information, manage incidents, calculate incident clearance times, evaluate congestion measures, and broadcast emergency alerts.
- The agencies typically engage in continuous collection of traffic volume data, incident information, and traffic crash data. They largely employ cloud platforms like Amazon AWS, Microsoft Azure, and Google Cloud and rely on analytical platforms such as Power BI and Tableau.

In addition, the TTI team conducted an extensive review of state DOT records from 2010–2022. The focus was to identify emergent trends in the realm of congestion performance measures, applications of AI and big data in traffic demand modeling, and the utilization of predictive analytics in monitoring congestion in real time. This review yielded several insightful findings:

- Delay emerged as the most used congestion measure, with the primary data sources being volume and speed metrics.
- State DOT planning activities primarily employed innovative methodologies like K-Means cluster analyses, regression analysis, and RF.
- Leading adopters of these novel applications were found in California, Florida, Minnesota, and China.

The TTI team also developed AI models for forecasting appropriate congestion measures in Texas. The model considered four measures of congestion: SpdAve, SpdStd, Spd85, and PVTT. ML models such as RF, GB, KNN, SVR, ANN, and CB were deployed. Both scenarios of including and excluding crash data as independent variables were explored. The analysis relied on a dataset comprising 28,684 rows of freeway segments. Separate models were constructed for rural and urban land use types. The study also integrated explainable AI, specifically SHAP, to interpret the modeling outcomes. The primary findings from the forecasting model are as follows:

- In the SpdAve model, CB exhibited superior performance, recording the lowest MAE, MSE, and RMSE.
- CB also outperformed other models in the SpdStd and the Spd85 models.
- Notably, the prediction accuracy seemed to improve when models were built based on different PSL levels, indicating the potential value of PSL levels as clustering or filtering options for speed-related congestion measures.
- Model performance was more efficient for rural freeways than for urban ones, which can be possibly attributed to data characteristics varying between the two types of land use.
- A linear trend was identified between observed and predicted congestion performance measures for different subsets based on PSLs. The linearity across most models suggests that AI models are well suited for predicting these values.
- The key factors influencing rural freeways, according to SHAP results, were traffic volume, jobs within a 45-minute auto travel time, k-factor, and median width.
- PDPs for rural freeways showed minimal impact from variables such as VMT traveled, segment length, and crashes (all) on the congestion measures. Conversely, traffic volume, jobs within a 45-minute auto travel time, k-factor, and median width appeared to correlate with congestion measures.
- The main influencing factors for urban freeways, based on the SHAP results, were traffic volume, median width, jobs within a 45-minute auto travel time, k-factor, and the number of lanes.
- For urban freeways, PDPs showed that variables like VMT and segment length had minimal impact on the congestion measures, while a higher number of crashes corresponded with a lower SpdAve. Traffic volume, jobs within a 45-minute auto travel time, median width, and lane numbers were associated with urban roadway SpdAve measures.

The TTI team also developed a prototype decision support tool based on a GIS. This tool estimates and visually represents the status of freeway congestion. The tool is designed to illustrate the congestion profiles of various roadway segments, thereby assisting transportation authorities and planners in identifying congestion hotspots. This identification process aids in prioritizing infrastructure improvements or specific traffic management interventions. By analyzing congestion measures for various segments, authorities can effectively allocate resources and implement measures aimed at reducing congestion and improving overall traffic flow. The key findings are as follows:

- Urban areas, particularly major cities such as Houston and Dallas, experience significant congestion due to factors like high population densities, extensive transportation networks, and a multitude of economic activities.

- Congestion in urban areas is subject to spatial and temporal variations. Different segments experience congestion differently due to factors like road network design, traffic demand, and ongoing construction or maintenance projects.
- Congestion measures are influenced by PSLs. Various segments display unique congestion patterns based on their respective speed limits.
- Segments with higher PSLs may have different congestion dynamics than segments with lower speed limits, which underscores the importance of considering speed limits when assessing congestion levels and formulating mitigation strategies.
- On rural freeways, the prevalence of districts with varying PSLs invites investigation into influencing factors. These factors might include geographic location, population density, economic activities, and transportation infrastructure development.
- Most rural freeways, regardless of the PSL, are designed with only two lanes. This design suggests a limited capacity to handle high traffic volumes, which may lead to potential congestion, especially during peak travel times or periods of high traffic demand.
- The capacity of urban roadways to accommodate traffic volumes is significantly influenced by the number of lanes. A larger number of lanes typically indicate a higher capacity for accommodating vehicles and facilitating smoother traffic flow.
- Despite having more lanes, urban roadways are not entirely immune to congestion. Factors like merging points, interchanges, and bottlenecks can still contribute to variability in travel time and congestion.

9.3 RECOMMENDATIONS

The following recommendations were developed from this study:

- Consider INRIX and Waze as primary data sources for freeway congestion reduction. Evaluate potential benefits of integrating data from Streetlight and Wejo.
- Utilize big data to measure travel time, gather real-time information, manage incidents, determine incident clearance times, assess congestion measures, and deliver emergency alerts.
- Implement a continuous data collection approach that focuses on traffic volume data, incident information, and traffic crash data.
- Leverage cloud platforms like Amazon AWS, Microsoft Azure, and Google Cloud to handle and process the large amounts of collected data.
- Adapt and adopt leading congestion management methodologies from states like California, Florida, and Minnesota, where techniques like K-Means cluster analyses, regression analysis, and RF are employed.
- Apply ML models like CB to predict congestion measures and consider different models for rural and urban areas based on the performance differences noted in the study.
- Pay attention to key influencing factors, such as traffic volume, jobs within a 45-minute auto travel time, k-factor, median width, and the number of lanes, when forecasting congestion on urban and rural freeways.
- Evaluate the potential of integrating explainable AI methodologies, such as SHAP, to interpret the outcomes of the ML models. This process may help researchers understand the influence of various factors on freeway congestion.

- Focus on prioritizing infrastructure improvements and targeted traffic management interventions in urban areas that experience significant congestion, particularly major cities like Houston and Dallas.
- Implement dynamic traffic management strategies such as variable speed limits and ramp metering that adapt to spatial and temporal variations in congestion.
- Review and adjust speed limits in various segments, considering that higher and lower speed limits can influence congestion dynamics differently.
- Consider the addition of lanes or creation of bypasses on rural freeways that are designed with only two lanes to better accommodate high traffic volumes and reduce congestion during peak travel times.
- Identify and improve problematic areas such as merging points, interchanges, and bottlenecks in urban roadways, which can contribute to variability in travel time and congestion despite the presence of multiple lanes.

9.4 CHAPTER SUMMARY

This chapter summarizes the results of this project (TxDOT 0-7131), which was focused on improving Texas surface transportation by predicting and mitigating congestion using AI and big data. The project aimed to address the research gap through two phases—validating data sources and identifying suitable AI models. Findings showed that INRIX and Waze are commonly used for congestion reduction, and cloud platforms like Amazon AWS and Microsoft Azure were favored for data processing. The study developed AI models to forecast congestion measures; notably, the CB model performed well. The team also created a decision support tool based on GIS to identify congestion hotspots. Recommendations include using INRIX and Waze, adopting leading methodologies, and implementing dynamic traffic management strategies in urban areas.

REFERENCES

Achillides, C.D., Bullock, D.M., 2004. Performance Metrics for Freeway Sensors (Web).

Alcolea, A., Paoletti, M.E., Haut, J.M., Resano, J., Plaza, A., 2020. Inference in Supervised Spectral Classifiers for On-Board Hyperspectral Imaging: An Overview. Remote Sens. 12, 534. https://doi.org/10.3390/rs12030534

Amekudzi-Kennedy, A., Clark, R., Wilson, J., Singh, P., 2020. Transportation Performance Management for System Operations: Development of processes, Tools, Measures and Targets (Digital/other).

Azizi, S., Ahmadloo, E., 2016. Prediction of heat transfer coefficient during condensation of R134a in inclined tubes using artificial neural network. Appl. Therm. Eng. 106, 203–210. https://doi.org/10.1016/j.applthermaleng.2016.05.189

Balgowan, R.M., 1987. Evaluation of the Accuracy, Reliability, Effectiveness, Expansibility, and Additional Potential Benefits of the Scan 16 Ef Moisture, Frost and Ice Early Warning System. Final Report.

Baroud, H., Dubey, A., Vazirizade, S.M., Mukhopadhyay, A., 2021. Collaborative Research Project to Coordinate the Data from the CRASH Predictive Analytics Program Between TDOT and TDOSHS (Digital/other).

Bento, A.M., 2021. Using Big Data to Estimate the Environmental Benefits of Congestion Pricing in the Los Angeles Metropolitan Area (Digital/other).

Bertini, R.L., Monsere, C., Eshel, O., Ahn, S., 2008. Using Archived ITS Data to Measure the Operational Benefits of a System-wide Adaptive Ramp Metering System (No. FHWA-OR-RD-09-10/ OTREC-RR-08-04). Portland State University, Oregon Department of Transportation. https://doi.org/10.15760/trec.152

Botchkarev, A., 2018. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. ArXiv Prepr. ArXiv180903006.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brewer, A., Schuerger, J., Hope, J., 2003a. Surface Transportation Security and Reliability Information System Model Deployment: Iflorida Final System Design.

Brewer, A., Schuman, R., Schuerger, J., 2003b. Surface Transportation Security and Reliability Information System Model Deployment: Iflorida Project Work Plan.

Cao, Y., Liu, D., Yin, Q., Xue, F., Tang, H., 2022. MSASGCN : Multi-Head Self-Attention Spatiotemporal Graph Convolutional Network for Traffic Flow Forecasting. J. Adv. Transp. 2022, e2811961. https://doi.org/10.1155/2022/2811961

Çavuşoğlu, Ü., Kaçar, S., 2019. Performance analysis of data mining algorithms for abnormal traffic detection. Acad. Platf. J. Eng. Sci. 7, 2.

Chaudhary, N.A., Farzaneh, R., Shelton, J., Le, M., Zhou, H. “Tracy,” Valdez, G., Venugopal, M., 2018. Potential Use of Ramp Metering as Congestion Management Strategy in the Dallas-Fort Worth Metroplex (Digital/other).

Chen, M., 2010. Travel Time Based Congestion Measures for Freeway Corridors (Digital/other).

Chen, M., Zhang, X., 2017. Collection and Analysis of 2013-2014 Travel Time Data (Digital/other).

Chen, M., Zhang, X., 2015. Methodology Update for Estimating Volume to Service Flow Ratio (Digital/other).

Chen, M., Zhang, X., Green, E., 2015. Analysis of Historical Travel Time Data (Digital/other).

Chinnam, R.B., Murat, A.E., Ulferts, G., 2010. Enabling Congestion Avoidance and Reduction in the Michigan-Ohio Transportation Network to Improve Supply Chain Efficiency: Freight ATIS (Digital/other).

Chiteka, K., Enweremadu, C.C., 2016. Prediction of global horizontal solar irradiance in Zimbabwe using artificial neural networks. J. Clean. Prod. 135, 701–711. https://doi.org/10.1016/j.jclepro.2016.06.128

Christie, D., Neill, S.P., 2022. Measuring and Observing the Ocean Renewable Energy Resource, in: Letcher, T.M. (Ed.), Comprehensive Renewable Energy (Second Edition). Elsevier, Oxford, pp. 149–175. https://doi.org/10.1016/B978-0-12-819727-1.00083-2

Coifman, B., Ponnu, B., 2018. Investigating the Feasibility of Coordinated Ramp Metering Along Freeway Corridors in Ohio (Digital/other).

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27. https://doi.org/10.1109/TIT.1967.1053964

Crawford, J.A., Carlson, T.B., Eisele, W.L., 2011. A Michigan Toolbox for Mitigating Traffic Congestion (Digital/other).

Cunningham, C., Davis, J., Aghdashi, B., Chase, T., Kim, S., Nye, T., 2016. Planning Level Evaluation of the Effects of Ramp Metering on North Carolina Freeways (Digital/other).

De'ath, G., 2007. Boosted Trees for Ecological Modeling and Prediction. Ecology 88, 243–251. https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2

Dutta, N., Fontaine, M.D., 2020. Improving Freeway Crash Prediction Models Using Disaggregate Flow State Information (Digital/other).

Duvvuri, S., Pulugurtha, S.S., 2021a. Researching Relationships between Truck Travel Time Performance Measures and On-Network and Off-Network Characteristics (Digital/other).

Duvvuri, S., Pulugurtha, S.S., 2021b. Researching Relationships between Truck Travel Time Performance Measures and On-Network and Off-Network Characteristics (No. Report 21-08). Mineta Transportation Institute, San José State University, California Department of Transportation. https://doi.org/10.31979/mti.2021.1946

Elefteriadou, L., Chrysikopoulos, G., Lombard, M., 2010. Travel Time Reliability Modeling For Florida (CD-ROM).

Elefteriadou, L., Li, Z., Jin, L., 2013. Modeling, Implementation, and Validation of Arterial Travel Time Reliability (Digital/other).

Elefteriadou, L., Lu, C., Li, Z., Wang, X., Jin, L., 2012. Multimodal and Corridor Applications of Travel Time Reliability (Digital/other).

Elefteriadou, L., Xu, H., 2007. Travel Time Reliability Models for Freeways and Arterials (Print).

Elsafi, S.H., 2014. Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. Alex. Eng. J. 53, 655–662. https://doi.org/10.1016/j.aej.2014.06.010

Establishing Monitoring Programs for Travel Time Reliability, 2014. . Transportation Research Board, Washington, D.C.

Fan, W., Chen, Z., n.d. Predicting Travel Time on Freeway Corridors: Machine Learning Approach.

Fontaine, M.D., Miller, J.S., 2012. Planning for Active Traffic Management in Virginia: International Best Practices and Implementation Strategies (Digital/other).

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal., Nonlinear Methods and Data Mining 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/aos/1013203451

Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22, 1365–1381. https://doi.org/10.1002/sim.1501

Fu, G., 1993. Improved Direct Importance Sampling Method for System Reliability.

Gan, A., Raihan, M.A., Alluri, P., Liu, K., Saha, D., 2016. Updating and Improving Methodology for Prioritizing Highway Project Locations on the Strategic Intermodal System (SIS) (Digital/other).

Ghumman, A.R., Ghazaw, Y.M., Sohail, A.R., Watanabe, K., 2011. Runoff forecasting by artificial neural network and conventional model. Alex. Eng. J. 50, 345–350. https://doi.org/10.1016/j.aej.2012.01.005

Hadi, M., Xiao, Y., Wang, T., Qom, S.F., Azizi, L., Jia, J., Massahi, A., Iqbal, M.S., 2016. Framework for Multi-Resolution Analyses of Advanced Traffic Management Strategies (Digital/other).

Haghani, A., Zhang, Y., 2015. Advanced Volatility Models for Improving Travel Time Prediction (No. NTC2015- SU- R- 06). University of Maryland, National Transportation Center at Maryland.

Haghani, A., Zhang, Y., Hamedi, M., 2014. Impact of Data Source on Travel Time Reliability Assessment (Digital/other).

Hallenbeck, M.E., McCormack, E., Sankarakumaraswamy, S., 2015. Developing a System for Computing and Reporting MAP-21 and Other Freight Performance Measures (No. WA-RD 844.1). Washington State Transportation Center (TRAC-UW), Washington State Department of Transportation.

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer.

Hefner, J.T., Spradley, M.K., Anderson, B., 2014. Ancestry Assessment Using Random Forest Modeling, ,. J. Forensic Sci. 59, 583–589. https://doi.org/10.1111/1556-4029.12402

Hendricks, S., Winters, P., Georggi, N., Varzardoliya, B.K., 2016. Transportation Demand Management Guidance for Corridor Planning Studies (Digital/other).

Hunter, M., Fujimoto, R., Guin, A., Laval, J., Henclewood, D., Suh, W., Huang, Y.-L., Li, Y., Stebar, B., Chari, P., 2012. Real Time Estimation of Arterial Travel Time and Operational Measures through Integration of Real Time Fixed Sensor Data and Simulation (Digital/other).

Idaho Transportation System: 2008 Performance Report (Web), 2009.

Illias, H.A., Chai, X.R., Abu Bakar, A.H., 2016. Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. Measurement 90, 94–102. https://doi.org/10.1016/j.measurement.2016.04.052

Isukapati, I., List, G., 2016. Using Travel Time Reliability Measures With Individual Vehicle Data. pp. 2131–2136. https://doi.org/10.1109/ITSC.2016.7795901

James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning, Springer Texts in Statistics. Springer US, New York, NY. https://doi.org/10.1007/978-1-0716-1418-1_1

Jensen, M., Andersen, D., Marecek, A., Anson, C., 2019. Autonomous Vehicle Feasibility Study – Final Report (Digital/other).

Kang, L., Hu, G., Huang, H., Lu, W., Liu, L., 2020. Urban Traffic Travel Time Short-Term Prediction Model Based on Spatio-Temporal Feature Extraction. J. Adv. Transp. 2020, 1–16. https://doi.org/10.1155/2020/3247847

Kassens-Noor, E., Savolainen, P.T., Gates, T.J., Cai, M., Cai, Q., Jashami, H., Megat-Johari, M.-U., Decaminada, T., Herin, G., Zockaie, A., 2022. Evaluation of an Active Traffic Management System with Part-Time Use of the Inside Shoulder (Digital/other).

Kim, J., Moon, J., 2022. Congestion Costs and Scheduling Preferences of Car Commuters in California: Estimates Using Big Data (Digital/other).

Kim, W., Kim, H., Won, M., Chang, G.-L., 2017. Development of a Traffic Management Decision Support Tool for Freeway Incident Traffic Management (FITM) Plan Deployment (Digital/other).

Kwon, E., Jurrens, C., Wright, C., Mahmud, A., 2022. Estimation of Metro Freeway System Reliability and Resilience (Digital/other).

Kwon, E., Park, C., 2018. Development of a Travel-Time Reliability Measurement System (Digital/other).

Kwon, T.M., 2020a. Improve Traffic Volume Estimates From MnDOT’s Regional Traffic Management Center (No. MN 2020-02).

Kwon, T.M., 2020b. Integrate RTMC Vehicle Classification Into the Current Detector Volume Data (No. MN 2020-31).

Lawphongpanich, S., Yin, Y., 2012. Nonlinear Road Pricing (Web).

Lee, M., Jin, X., 2020. Incorporating Reliability Measures into the Freight Project Prioritization Decision Support System (Digital/other).

Liao, C.-F., 2018. Measure of Truck Delay and Reliability at the Corridor Level (Digital/other).

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

List, G., Falcocchio, J.C., Ozbay, K., Mouskos, K.C., 2008. Quantifying Non-Recurring Delay on New York City's Arterial Highways (Web).

Liu, H., Shetty, R.R., 2021. Analytical Model for Traffic Congestion and Accident Analysis (No. Report 21-27).

Liu, X.C., Chen, Z., 2017. Data-Driven Freeway Performance Evaluation Framework for Project Prioritization and Decision Making.

Liu, Z., Tan, H., 2021. Traffic Prediction with Graph Neural Network: A Survey. Presented at the 21st COTA International Conference of Transportation Professionals.

Lomax, T., Lasley, P., Ellis, D., Stockton, B., 2013a. Mobility Investment Priorities Project (Digital/other).

Lomax, T., Lasley, P., Ellis, D., Stockton, B., 2013b. Mobility Investment Priorities Project. Texas A&M Transportation Institute, Texas Department of Transportation.

Ma, X., Wu, Y., Wang, Y., 2011. DRIVE Net: E-science transportation platform for data sharing, visualization, modeling, and analysis. Transp Res Rec J Transp Res Board 37–49.

Macfarlane, G., Copley, M., 2020. A Synthesis of Passive Third-Part Data Sets used for Transportation Planning. Report.

Margiotta, R., Lomax, T.J., Hallenbeck, M.E., Turner, S.M., Skabardonis, A., Ferrell, C., Eisele, W.L., 2007. Guide to Effective Freeway Performance Measurement: Final Report and Guidebook (No. NCHRP Project 3-68). Transportation Research Board.

Martchouk, M., Mannering, F.L., Singh, L., 2010a. Travel Time Reliability in Indiana (Digital/other).

Martchouk, M., Mannering, F.L., Singh, L., 2010b. Travel Time Reliability in Indiana (No. FHWA/IN/JTRP 2010-8). Purdue University, Indiana Department of Transportation, Federal Highway Administration.

Mattar, M.A., Alazba, A.A., Zin El-Abedin, T.K., 2015. Forecasting furrow irrigation infiltration using artificial neural networks. Agric. Water Manag. 148, 63–71. https://doi.org/10.1016/j.agwat.2014.09.015

McLellan, J.W., 2013. Quantifying the Impact of New Freeway Segments (Digital/other).

Metropolitan Freeway System 2013 Congestion Report (Digital/other), 2014.

Mia, M., Dhar, N.R., 2016. Prediction of surface roughness in hard turning under high pressure coolant using Artificial Neural Network. Measurement 92, 464–474. https://doi.org/10.1016/j.measurement.2016.06.048

Millar, R., 2016. WSDOT's Handbook for Corridor Capacity Evaluation. Washington State Department of Transportation.

Miller, J.S., 2022. A Model to Forecast Peak Spreading (No. FHWA/VCTIR 12-R11). Virginia Department of Transportation, Federal Highway Administration.

Miller, J.S., 2012. A Model to Forecast Peak Spreading (Digital/other).

Minnesota Tolling Study Report Modern Tolling Practices and Policy Considerations (Digital/other), 2018.

Miovision, 2020. TrafficLink Platform [WWW Document]. URL https://miovision.com/resources/trafficlink-platform/.

Mohamed, H.I., 2013. Design of alluvial Egyptian irrigation canals using artificial neural networks method. Ain Shams Eng. J. 4, 163–171. https://doi.org/10.1016/j.asej.2012.08.009

Molan, A., Murugesan, N., Shams, A., Tortora, C., Rahman, F., Loh, J., Pande, A., 2020. Evaluation of Coordinated Ramp Metering (CRM) Implemented By Caltrans. Mineta Transp. Inst. Publ. https://doi.org/10.31979/mti.2020.1812

Monsere, C.M., Bertini, R.L., Ahn, S., Eshel, O., 2008. Using Archived ITS Data to Measure the Operational Benefits of a System-Wide Adaptive Ramp Metering System (Print).

Morris, P., Parikh, G., 2022. The Tipping Point: What COVID-19 Travel Reduction Tells Us About Effective Congestion Relief (Digital/other).

Moses, R., 2015. Twenty-Four Hour Peaking Relationship to Level of Service and Other Measures of Effectiveness (Digital/other).

Najumudeen, K., Bin, M.E.A., 2020. Traffic Accident Analysis and Prediction Using The NPMRDS.

Nasr, M.S., Moustafa, M.A.E., Seif, H.A.E., El Kobrosy, G., 2012. Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT. Alex. Eng. J. 51, 37–43. https://doi.org/10.1016/j.aej.2012.07.005

National Academies of Sciences, Engineering, and Medicine, 2014. Establishing Monitoring Programs for Travel Time Reliability (No. SHRP 2 Report S2-L02-RR-1). Transportation Research Board, Washington, D.C. https://doi.org/10.17226/22612

Olayode, I.O., Tartibu, L.K., Okwu, M.O., 2021. Prediction and modeling of traffic flow of human-driven vehicles at a signalized road intersection using artificial neural network model: A South African road transportation system scenario. Transp. Eng. 6, 100095. https://doi.org/10.1016/j.treng.2021.100095

Ong, G.P., Noureldin, S., Sinha, K.C., 2010. Automated Pavement Condition Data Collection Quality Control, Quality Assurance, and Reliability (Digital/other).

Özdoğan-Sarıkoç, G., Sarıkoç, M., Celik, M., Dadaser-Celik, F., 2023. Reservoir volume forecasting using artificial intelligence-based models: Artificial Neural Networks, Support Vector Regression, and Long Short-Term Memory. J. Hydrol. 616, 128766. https://doi.org/10.1016/j.jhydrol.2022.128766

Paleti, C., Peeta, S., Sinha, K.C., 2014. Identifying Strategies to Improve Lane Use Management in Indiana (Digital/other No. 9781622603107).

Pandey, V., Juri, N.R., 2018a. Using National Performance Management Research Data Set for Corridor Performance Measures: A US-281N Corridor Case Study. Transp. Res. Rec. 2672, 257–267. https://doi.org/10.1177/0361198118796972

Pandey, V., Juri, N.R., 2018b. Using National Performance Management Research Data Set for Corridor Performance Measures: A US-281N Corridor Case Study. Transp. Res. Rec. J. Transp. Res. Board 2672, 257–267. https://doi.org/10.1177/0361198118796972

Peeta, S., Kumar, A., Sharma, S., 2011. Post-Processing Techniques to Enhance Reliability of Assignment Algorithm Based Performance Measures (Web).

Petalas, Y., Ammari, A., Georgakis, P., Nwagboso, C., 2017. A Big Data Architecture for Traffic Forecasting Using Multi-Source Information.

Pirouzmand, A., Kazem Dehdashti, M., 2015. Estimation of relative power distribution and power peaking factor in a VVER-1000 reactor core using artificial neural networks. Prog. Nucl. Energy 85, 17–27. https://doi.org/10.1016/j.pnucene.2015.06.001

Pozdnoukhov, A., 2018. Travel Demand Nowcasting (Digital/other).

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18. Curran Associates Inc., Red Hook, NY, USA, pp. 6639–6649.

PTV, 2022. PTV OPTIMA, Real-time traffic management software for multimodal networks [WWW Document]. URL http://vision-traffic.ptvgroup.com/en-us/products/ptv-optima/.

Pulugurtha, S.S., Penmetsa, P., Duddu, V.R., 2017. Monetizing Reliability to Evaluate the Impact of Transportation Alternatives (Digital/other).

Qiu, B., Fan, W., 2021. Travel time forecasting on a freeway corridor: a dynamic information fusion model based on the random forests approach.

Refai, H.H., Bitar, N., Kaleia, M.S., 2017. National Performance Management Research Dataset (NPMRDS) - Speed Validation for Traffic Performance Measures (Digital/other).

Rhode Island Statewide Model Update, 2016. . Tech. Pap.

Saidi, S.J., 2020. Comparison of Machine Learning and Statistical Approaches for Predicting Travel Times in the Oklahoma Highway System.

Schneider, H., Hutchinson, C., Pfetzer, E., 2019. Highway Construction Work Zone Safety Performance and Improvement in Louisiana (Digital/other).

Schrank, D., Eisele, B., Lomax, T., 2019. Urban Mobility Report 2019 (Digital/other).

Seeherman, J., Anderson, P., 2017. Measuring the influence of recurring sporting events on freeway characteristics (Digital/other).

Sharma, A., Ahsani, V., 2019. Assessing the Impact of Game Day Schedule and Opponents on Travel Patterns and Route Choice using Big Data Analytics (Digital/other).

Sharma, A., Wang, S., Khattak, A., 2016. Active Traffic Management Case Study: Phase 1 (Digital/other).

Shelton, J., Morgan, C., Warner, J., Valdez, G., Sharma, S., Pesti, G., Le, M., Balke, K., 2020. Strategies for Managing Freight Traffic Through Urban Areas: Technical Report (Digital/other).

Siemens, 2022. Traffic Prediction | Strategic Management and Coordination [WWW Document]. URL https://new.siemens.com/global/en/products/mobility/road-solutions/traffic-management/strategic-management-and-coordination/traffic-prediction.html.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Sobolewski, M., Polum, T., Morris, P., Loos, R., Anderson, K., 2014. Pilot Testing of SHRP 2 Reliability Data and Analytical Products: Minnesota [supporting dataset].

Southworth, F., Gillett, J., 2011. Trucking in Georgia: Freight Performance Measures (Web).

Sriraj, P.S., Metaxatos, P., Rueter, J., Barnum, A., Bernal, M., 2017. Integration of Transportation for Improved Mobility (Digital/other).

Srivastava, A., Chitturi, M.V., Ahn, S., Rafferty, P., 2018a. Analytical Methods for Work Zone Travel Time Reliability (Digital/other).

Srivastava, A., Chitturi, M.V., Ahn, S., Rafferty, P., 2018b. Analytical Methods for Work Zone Travel Time Reliability (No. InTrans Project 06-277). University of Wisconsin, Iowa Department of Transportation, Federal Highway Administration.

Stevanovic, A., Mitrovic, N., 2019a. Development of Traffic Map Evaluation Tool for TMC Applications.

Stevanovic, A., Mitrovic, N., 2019b. Development of a Traffic Map Evaluation Tool for TMC Applications. Florida Atlantic University, Florida Department of Transportation.

Swanson, B., Culp, J., 2022. Forecasting Travel-Time Reliability (Digital/other).

Texas A&M Transportation Institute, 2020. Analysis Procedures and Mobility Performance Measures 100 Most Congested Texas Road Sections. College Station, TX.

The 2010 Congestion Report (Web), 2010.

Tian, Z., Wang, A., Xu, H., 2020. Developing a Quality of Signal Timing Performance Measure Methodology for Arterial Operations (Digital/other).

Twin Cities Metro Freight Initiative: Performance Management Framework (Digital/other), 2011.

URBANLOGIQ, 2022. Transportation Data Analytics Platform for Government [WWW Document]. URL https://www.urbanlogiq.com/traffic.

Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. Springer, New York, NY. https://doi.org/10.1007/978-1-4757-3264-1

Varaiya, P.P., 2008. Causes of Freeway Productivity Decline and the Opportunities for Gain: A Quantitative Study, Final Report. ed. University of California, Berkeley.

Wang, Y., Ban, X., Cui, Z., Zhu, M., 2019. An Artificial Intelligence Platform for Network-wide Congestion Detection and Prediction using Multi-Source Data, in: C2 SMART Report. Seattle, WA.

Wang, Y., Ma, X., Xiao, S., Malinovskiy, Y., Corey, J., Henrickson, K., 2013. Digital Roadway Interactive Visualization and Evaluation Network Applications to WSDOT Operational

Data Usage (No. WA-RD 823.1). Smart Transportation Applications and Research Laboratory, University of Washington.

Wang, Y., Yu, R., Lao, Y., Thomson, T., 2010. Quantifying Incident-Induced Travel Delays on Freeways Using Traffic Sensor Data: Phase II (Digital/other).

Wang, Y., Zhang, W., Henrickson, K., Ke, R., Cui, Z., 2016. Digital Roadway Interactive Visualization and Evaluation Network Applications to WSDOT Operational Data Usage (Digital/other).

Washburn, S.S., Ko, B., 2007. Travel Time Reliability and Truck Level of Service on the Strategic Intermodal System Part B: Truck Level of Service (Print).

Waycare, 2022. AI-driven Mobility Solutions [WWW Document]. URL http://waycaretech.com/.

Williams, B.M., Chase, R.T., Xu, Y., Kim, S., Craft, D., Rouphail, N.M., Encarnacion, A., 2013. Mobility and Reliability Performance Measurement (Digital/other).

Williams, B.M., Rouphail, N.M., Kim, S., Song, T.J., 2016. Incident Management Assistance Patrols – Assessment of Benefits/Costs, Route Selection, and Prioritization (Digital/other).

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30, 79–82. https://doi.org/10.3354/cr030079

Xu, R., Luo, F., 2021. Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest. Saf. Sci. 135, 105125. https://doi.org/10.1016/j.ssci.2020.105125

Yang, J., Li, J., Wei, L., Gao, L., Mao, F., 2022. Spatiotemporal DeepWalk Gated Recurrent Neural Network: A Deep Learning Framework for Traffic Learning and Forecasting. J. Adv. Transp. 2022, e4260244. https://doi.org/10.1155/2022/4260244

Yang, T., Asanjan, A.A., Welles, E., Gao, X., Sorooshian, S., Liu, X., 2017. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resour. Res. 53, 2786–2812. https://doi.org/10.1002/2017WR020482

Yao, W., Qian, S., 2021. From Twitter to Traffic Predictor: Next-Day Morning Traffic Prediction Using Social Media Data. Transp. Res. Part C Emerg. Technol. 124, 102938. https://doi.org/10.1016/j.trc.2020.102938

Zhang, T., Lin, W., Vogelmann, A.M., Zhang, M., Xie, S., Qin, Y., Golaz, J.-C., 2021. Improving Convection Trigger Functions in Deep Convective Parameterization Schemes Using Machine Learning. J. Adv. Model. Earth Syst. 13, e2020MS002365. https://doi.org/10.1029/2020MS002365

Zhang, W., Zhu, F., Lv, Y., Tan, C., Liu, W., Zhang, X., Wang, F.-Y., 2022. AdapGL: An adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks. Transp. Res. Part C Emerg. Technol. 139, 103659. https://doi.org/10.1016/j.trc.2022.103659

Zhang, W.-B., 2008. Study of Integrated Corridor Management for San Francisco Bay Area I-880 Corridor. University of California, Berkeley.

Zhang, X., Zhao, M., Appiah, J., Fontaine, M.D., 2021a. Methods to Analyze and Predict Interstate Travel Time Reliability (Digital/other).

Zhang, X., Zhao, M., Appiah, J., Fontaine, M.D., 2021b. Methods to Analyze and Predict Interstate Travel Time Reliability (No. FHWA/VTRC 22-R2). Virginia Transportation Research Council, Virginia Department of Transportation.

Zhang, Z., Yuan, Y., Yang, X. (Terry), 2020. A Hybrid Machine Learning Approach for Freeway Traffic Speed Estimation. Transp. Res. Rec. J. Transp. Res. Board 2674, 68–78. https://doi.org/10.1177/0361198120935875

Zhou, L., Zhang, S., Yu, J., Chen, X., 2019. Spatial-Temporal Deep Tensor Neural Networks for Large-Scale Urban Network Speed Prediction. IEEE Trans. Intell. Transp. Syst. PP, 1–12. https://doi.org/10.1109/TITS.2019.2932038

Zlatkovic, M., Zhou, X., 2017. Simplified Web-Based Decision Support Method for Traffic Management and Work Zone Analysis (Digital/other).

Zwahlen, H.T., Oner, E., 2006. Improved Work Zone Design Guidelines and Enhanced Model of Travel Delays in Work Zones. Phase I: Portability and Scalability of Interarrival and Service Time Probability Distribution Functions for Different Locations in Ohio and the Establishment of Improved Work Zone Design Guidelines (Print).

APPENDIX A: SURVEY QUESTIONNAIRE

Texas A&M Transportation Institute (TTI) is leading research on the application of AI in freeway congestion reduction for the Texas Department of Transportation (TxDOT). This survey is in support of TxDOT Project 0-7131, **Leveraging Artificial Intelligence (AI) Techniques to Detect, Forecast, and Manage Freeway Congestion**.

This survey will take approximately **10-15 minutes** to complete, and it will allow you to save your work and return to finish at a later time.

This survey is to be completed by DOT Agency staff with knowledge of freeway congestion and traffic monitoring, including:

- Current agency practices of freeway congestion management
- Datasets used for congestion reduction
- Big data platforms or servers
- Dashboard of freeway congestion or similar tools

This survey should be completed by persons in each Agency. Responses from multiple persons in one Agency will be combined. The answers you provide will be synthesized with the information collected from other transportation professionals. Your answers will remain confidential to the extent allowed or required by law. We greatly appreciate your participation in this study. If you have questions, or concerns regarding this study, you may contact the principal investigator, **Ioannis Tsapakis**, at i-tsapakis@tti.tamu.edu. The survey is only open to individuals 18 years and older. Your answers may need to be clarified, or further information may be collected by TTI.

1. Do you agree to participate in this survey?
    a. Yes
    b. No → THANK AND TERMINATE

Agency Information

1. Participant information
    a. Name:
    c. Professional Title:
    d. Agency Name:
    g. City:
    h. State/Province:
    i. Postal Code:
    j. Country:
    k. Phone Number:
    l. Email Address:
2. What is the principal task(s) your office focuses on? Please check all that apply.
    a. Congestion measures
    b. Travel time

c. Incident management
d. Incident clearance time
e. Real-time information
f. Emergency alert and crisis information
g. Other (specify):

Data Collection

3. Do you use private vendor data from the following data vendors? Please check all that apply.
    a. INRIX
    b. AirSage
    c. Wejo
    d. Waze
    e. Streetlight
    f. Strava
    g. Twitter
    h. Other (specify):
    i. None
4. Indicate the type of information your agency acquires via the following major private data vendors. Please check all that apply.

Type of Information	INRIX	AirSage	Wejo	Streetlight	Strava	Others
Congestion measures	○	○	○	○	○	○
Travel time	○	○	○	○	○	○
Incident management	○	○	○	○	○	○
Incident clearance time	○	○	○	○	○	○
Real-time information	○	○	○	○	○	○
Emergency alert and crisis information	○	○	○	○	○	○
Other (specify below)	○	○	○	○	○	○

List any other relevant information you want to add.

Enter text in box.

5. Indicate what type data your agency collects. Please check all that apply.

Type of Information	Continuous	Once in a week	Once in a month	Once in a year	Others
Traffic volume data	○		○	○	○
Traffic volume data by vehicle type	○		○	○	○
Travel time	○		○	○	○
Speed	○		○	○	○
Travel time by vehicle type	○		○	○	○
Other (specify below)	○		○	○	○

List any other relevant information you want to add.

Enter text in box.

Datasets and Methods

6. Do you use following data for congestion reduction methods?
   a. Short duration traffic volume data
   b. Traffic volume data by vehicle type
   c. Travel time by vehicle type
   d. Speed by vehicle type
   e. Weather data
   f. Incident information
   g. Traffic crash data
   h. Social media feed
   i. Other (specify):

7. Please add any datasets and variables, which would be effective in congestion reduction related matrix development.

Enter text in box.

8. Please provide us information on the application of AI in freeway congestion reduction by your agency. If not applicable, please type NA.

Enter text in box.

9. Please provide us information on applied AI algorithms or relevant report or code in freeway congestion reduction by your agency. If not applicable, please type NA.

Enter text in box.

Big Data Platforms

10. Do you use any Big Data platforms for your congestion management projects?
    a. AWS
    b. Google Cloud
    c. Microsoft Azure
    d. Other (specify):

11. Do you use any Enterprise Software Platforms for your congestion management projects?
    a. C3.ai
    b. Qubole
    c. Alteryx
    d. SparkCognition
    e. ASAPP
    f. Other (specify):

12. Do you use any Cloud Powered Analytics for your congestion management projects?
    a. Snowflake
    b. Cloudera
    c. Databricks
    d. Other (specify):

13. Does your agency maintain a dashboard to show the congestion-related analytics?
    a. Yes
    b. No.

14. [IF Q12=YES] If yes, please provide the link(s). List any other relevant information you want to add.

Enter text in box.

Reports and Guidance Document

15. Please upload or provide link of any report or guidance document your agency developed on freeway congestion reduction.

Upload link and text box

16. Please upload or provide link of any relevant and noteworthy report or guidance document.

Upload link and text box

Lessons Learned and Future Needs

17. What are the congestion reduction goals for your agency over the next 1–3 years?

Enter text in box.

18. Please describe any lessons learned by your agency that could be useful to other agencies.

Enter text in box.

19. Please describe any suggestions on improving congestion reduction on freeways in the future.

Enter text in box.

20. Please describe any challenges your agency is facing to incorporate big data platforms in freeway congestion reduction.

Enter text in box.

21. Do you know of any peer transportation agencies that may be interested in this survey?
    a. Yes
    b. No → THANK AND TERMINATE

22. Please provide contact information for the peer transportation agency that may be interested in this survey.
    a. Agency name:
    b. Contact person name:
    c. Contact email:

23. [FOR SURVEY COMPLETES] Thank you for completing our survey! Your input is greatly appreciated. If we need more information or clarification, we may need to contact you in the future. Please feel free to contact the principal investigator at s-das@tti.tamu.edu with any questions in the meantime.

APPENDIX B: STATE DOT PERFORMANCE METRICS AND FORMULAS

Congestion performance measures and formulas were listed within reports, guides, and manuals of eight states identified during the novel applications search provided in Table 57. This appendix provides more information on formulas associated with congestion metrics to ascertain potential avenues forward in predictive analytics and AI strategies for traffic congestion forecasting.

Table 57. Congestion Performance Measures and Associated Formulas.

Performance Metric	Equation	DOT
Delay per Mile	$\text{Delay per Mile (annual hours per mile)} = \dfrac{\left(\underset{(minutes)}{\text{Actual Travel Time}} - \underset{(minutes)}{\text{Free-Flow Travel Time}}\right) \times \underset{(vehiles)}{\text{Vehile Volume}} \times \underset{(\frac{persons}{vehicle})}{\text{Vehicle Occupancy}} \times \dfrac{hour}{60\ minutes}}{\text{Road Miles}}$	TxDOT (Texas A&M Transportation Institute, 2020)
Total Delay	$\text{Total Segment Delay (person} - \text{minutes)} = \left[\underset{(minutes)}{\text{Actual Travel Time}} - \underset{(minutes)}{\text{Free} - \text{Flow Travel Time}^3}\right] \times \underset{(vehicles)}{\text{Vehicle Volume}} \times \underset{(persons/vehicle)}{\text{Vehicle Occupancy}}$	TxDOT (Texas A&M Transportation Institute, 2020)
Congestion Cost	$\text{Congestion Cost} = \underset{\text{Vehicle Cost}}{\text{Annual Passenger}} + \underset{\text{Cost}}{\text{Annual Truck}}$ $\underset{\text{Vehicle Cost}}{\text{Annual Passenger}} = \left[\underset{\text{Hours of Delay}}{\text{Annual Passenger Vehicle}} \times \underset{(\frac{persons}{vehicle})}{\text{Vehicle Occupancy}} \times \underset{(\frac{\$X}{hour})}{\text{Value of Person Time}}\right] + \left[\underset{\text{Consumed by Passenger Vehicles}}{\text{Annual Gallons of Excess Fuel}} \times \underset{(\$Y/gallon)}{\text{Price per Gallon of Gasoline}}\right]$ $\underset{\text{Truck Cost}}{\text{Annual}} = \left[\underset{\text{Hours of Delay}}{\text{Annual Truck}} \times \underset{(\$Z/hour)}{\text{Value of Time for Trucks}}\right] + \left[\underset{\text{Consumed by Trucks}}{\text{Annual Gallons of Excess Fuel}} \times \underset{(\$Y/gallon)}{\text{Price per Gallon of Diesel}}\right]$	TxDOT (Texas A&M Transportation Institute, 2020)
Throughput Productivity	$\text{Throughput Productivity} = \begin{cases} 1, speed \geq maximum\ throughput\ speed \\ 1 - \dfrac{V}{V_0}, speed < maximum\ throughput\ speed \end{cases}$ Where, V_0: $optimul\ throughput$ V: $5 - minute\ flow\ rate$	WSDOT (Wang et al., 2013)

Performance Metric	Equation	DOT
Commute Congestion Cost	$Commute\ congestion\ cost = \sum(Average\ travel\ time_{\ 5\ min} - Travel\ time_{\ threshold\ speed}) \times Traffic\ volume_{\ 5\ min} \times Cost\ per\ minute$	WSDOT (Wang et al., 2013)
Hours of Travel Delay	$Hours\ of\ travel\ delay = \left(\frac{Vehicle\ miles\ traveled}{Travel\ speed}\right) - \left(\frac{Vehicle\ miles\ traveled}{Threshold\ speed}\right)$ WSDOT uses maximum throughput speed (85% of PSL) as the threshold in order to measure delay relative to a highway's most efficient operating condition.	WSDOT (Wang et al., 2013)
Annual Hours of Vehicle Delay	$AHD = \sum_{i=1}^{n}(M_i\ /\ S_i) - \ (M_i\ /\ S_T)$ Where i means the ith weekday in a year, n is the total number of weekdays in a year, and correspondingly, M_i and S_i are the VMT and travel speed on the ith weekday of the year.	WSDOT (Wang et al., 2013)
Annual Cost of Vehicle Delay	$C = (C_H \times CPI_{change}) \times H$ Where C is the annual cost of vehicle delay, its unit is dollars, CPI means consumer price index and CPI_{change} represents the difference in CPI between 2 years, and H is the hours of travel delay.	WSDOT (Wang et al., 2013)
BI	$Buffer\ Index = \left[\frac{95th\ percentile\ Travel\ Time - Average\ Travel\ Time}{Average\ Travel\ Time}\right] \times 100\%$	INDOT (Martchouk et al., 2010b)
PTI	$Planning\ Time\ Index = \frac{95th\ percentile\ Travel\ Time}{Travel\ Time\ Based\ on\ Free-Flow\ Speed}$	INDOT (Martchouk et al., 2010b)
Misery Index	$Misery\ Index = \frac{Average\ travel\ time\ for\ the\ longest\ 20\%\ of\ trips - Average\ Travel\ Time}{Average\ Travel\ Time}$	INDOT (Martchouk et al., 2010b)
Chi-Squared Statistic	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ Where χ^2 is the chi-squared statistic. When the ratio of sample variance and population variance is near 1, then χ^2 statistics will be close to degrees of freedom $(n-1)$, and the null hypothesis holds. $1. Null\ Hypothesis: H_0: s_0^2 = \sigma_0^2$ $2. Altenative\ Hypothesis: H_a: s_0^2 \neq \sigma_0^2$ Where s^2 is the sample variance and σ^2 is the population variance (to test the magnitude of travel time variance between different runs).	INDOT (Martchouk et al., 2010b)
Skew-Width Measures	Skew-width measures: λ^{Skew} is defined as the ratio of the difference between the 90th and the 50th percentile travel times to the difference between the 50th and the 10th percentile travel times. λ^{Var} is defined as the ratio of the difference between the 90th and the 10th percentile travel times to the 50th percentile travel time. Larger magnitudes of λ^{Skew} implies higher probability travel times of road to be extreme (either high or low). Larger magnitudes of λ^{Var} implies a wider distribution of travel times with respect to its median (the 50th percentile travel time).	Caltrans (Duvvuri and Pulugurtha, 2021b)

Performance Metric	Equation	DOT
	$\lambda^{Skew} = \frac{(90th\ percentile\ travel\ time - 50th\ percentile\ travel\ time}{50th\ percentile\ travel\ time - 10th\ percentile\ travel\ time}$ $\lambda^{Var} = \frac{(90th\ percentile\ travel\ time - 10th\ percentile\ travel\ time}{50th\ percentile\ travel\ time}$	
Delay	$Delay = VHT - (Count, veh) * \frac{Distance\ between\ stations, mi}{Free\ Flow\ Speed, mph}$ $Vehicle\ Hours\ Traveled\ (VHT) = (Count, veh) * (Distance\ between\ stations, mi)/(Speed, mph)$	ODOT (Bertini et al., 2008)
BTI	$Average\ Buffer\ Hours\ per\ Trip = Buffer\ Time\ Index\ (Fraction\ Congested) * (Vehic - Hours\ Traveled / Vehicle\ Trips).$	TxDOT (Lomax et al., 2013b)
Median Travel Time	$\tau_{median}(t) = median\left(\tau_i(t)\right)\forall i$ Where $\tau_i(t)$ is the travel time function observed on the ith day, representing the experienced travel time on the corridor as a function of time of day (t). Similarly, $\tau_{median}(t)$ represents the median travel time function obtained by calculating the median of travel times observed on all days for that particular time of day.	TxDOT (Pandey and Juri, 2018b)
Deviation of Travel Time of an *ith Day* from a Typical Day	The deviation of travel time function of an ith day from a typical day was measured by calculating mean absolute error (MAE_i) and root mean square (RMS_i) error. . $MAE_i = \frac{\sum_t \lvert \tau_{median}(t) - \tau_i(t) \rvert}{n}$ $RMS_i = \sqrt{\frac{\sum_t (\tau_{median}(t) - \tau_i(t))^2}{n}}$ Where n is the number of days being analyzed. For this day-to-day variation analysis, all weekdays from Jan 1, 2015, to Aug 31, 2016, were considered, and thus the value of n is 436. The data were cleaned only using the first level of cleaning to retain the outliers occurring during particular days.	TxDOT (Pandey and Juri, 2018b)
SV	SV, σ_r^2 , is a one-sided variance metric that uses a reference value r instead of the mean as the basis for the calculation, and only observations x_i that are greater or less than that reference value are used: $\sigma_r^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - r)^2\ and\ \sigma_r = \sqrt{\sigma_r^2}\ \exists x_i \geq r$ To differentiate travel time observations based on reliability from system loading, such as congestion, use the SV because the SV is sensitive to how travel times are distributed above the minimum value.	USDOT (National Academies of Sciences, Engineering, and Medicine, 2014)
Capacity for Vehicles per Hour *	$Base\ capacity = 2{,}200 + 10 \times (FFS - 50)$ $Capacity = Base\ capacity \times Lanes \times f_{hv} \times PHF \times f_p \times f_g$ Where, FFS = Free-flow speed; f_{hv} = the heavy vehicle adjustment factor depending on facility type, vehicle mix, and grade; PHF = the peak hour factor, calculated by the ratio of the peak 15-minute flow rate to the average hourly flow rate; and f_p and f_g = adjustment factors for driver population and grades provided by VDOT's Traffic Engineering Division.	VDOT (X. Zhang et al., 2021b)
LOTTR*	$TTRM = 100 \times \frac{\sum_{i=1}^{R} SL_i \times AV_i \times OF_j}{\sum_{i=1}^{T} SL_i \times AV_i \times OF_j}$ Where, $TTRM$ = travel time reliability measure;	VDOT (X. Zhang et al., 2021b)

Performance Metric	Equation	DOT
	SL_i = the segment length of interstate (or non-interstate) National Highway System (NHS) reporting segment i; AV_i = annual traffic volume of reporting segment i, calculated as AADT × Directional factor × 365 (366 for leap year), where the directional factor is the factor splitting AADT by direction with the default value of 0.5; OF_j = occupancy factor for vehicles on the NHS within a specified geographic area j within the state/metropolitan planning area; R = total number of interstate (or non-intertstate) reporting segments with an LOTTR value below 1.50 for all four time periods; and T = total number of interstate (or non-interstate) NHS reporting segments.	
Travel Time Index	$TTI = \frac{Mean\ peak\ period\ travel\ time}{Free - flow\ travel\ time}$	VDOT (X. Zhang et al., 2021b)
VHD	$Vehicle - Hours\ of\ delay = (Actual\ Travel\ Time - Threshold\ Travel\ Time) * Volume$ The four different threshold-based delay measures recommended by the NCHRP 08-98 are • $Delay_{Free\ flow}$. • $Delay_{Speed\ limit}$. • $Delay_{MaxEfficiency}$. • $Delay_{Target}$.	WSDOT (Hallenbeck et al., 2015)
Annual Hours of Person Delay	$Annual\ hours\ of\ person\ delay = (Average\ hours\ of\ daily\ vehicle\ delay \times Average\ vehicle\ occupants) \times Work\ days\ per\ year$	WSDOT (Hallenbeck et al., 2015)
Delay per User	$Delay\ per\ user_{Year;Corridor} = \frac{Annual\ hours\ of\ person\ delay}{(Average\ daily\ traffic\ volume \times Average\ vehicle\ occupants)}$	WSDOT (Hallenbeck et al., 2015)
Percent Lane Miles Delayed	$Percent\ lane\ miles\ delayed = \left(\frac{Lane\ miles\ with\ speed < 85\%\ of\ posted\ speed}{Total\ lane\ miles\ _{region}}\right)$	WSDOT (Hallenbeck et al., 2015)
Percent Lane Miles Congested	$Percent\ lane\ miles\ congested = \left(\frac{Lane\ miles\ with\ speed < 70\%\ of\ posted\ speed}{Total\ lane\ miles\ _{region}}\right)$	WSDOT (Hallenbeck et al., 2015)
Vehicle Delay Per Capita	$Vehicle\ delay\ per\ capita = \left(\frac{Annual\ hours\ of\ vehicle\ delay\ _{region}}{Population\ _{region}}\right)$	WSDOT (Millar, 2016)
Work Zone Impact Ratio	$WZ\ impact\ ratio: WZIR_{TTR\ measure} = TTRMeasure_{WZ}/TTRMeasure_{baseline}$	WisDOT (Srivastava et al., 2018b)
Prediction Interval	$(u_t - Z_{\alpha/2}\sigma_t, u_t + Z_{\alpha/2}\sigma_t)$ Where u_t is the predicted mean, $Z_{\alpha/2}$ denotes the standard score corresponding to the cumulative probability level of $\alpha/2$, and σ_t is the prediction variance from a volatility model.	MDOT (Haghani and Zhang, 2015)

<table>
<tr><th>Performance Metric</th><th>Equation</th><th>DOT</th></tr>
<tr><td>HVR</td><td>Formula: $HVR = V_{i,j}/V_{h_{i,j}}^{d}$;
Where, $V_{h_{i,j}}^{d}$—historical volume at time instant i and day d at the microwave vehicle detection system (MVDS) station j.
Description: This performance measure compares the observed and expected volume. It quantifies deviations of the current from the historical average volume at a particular MVDS station. The historical average is computed based on 1 year of past volume data.</td><td>FDOT (Stevanovic and Mitrovic, 2019b)</td></tr>
<tr><td>HTTR</td><td>Formula: $HTTR = TT_{i,j}^{d}/ TT_{h_{i,j}}^{d}$;
Where, $TT_{h_{i,j}}^{d}$—historical travel time at time instant i along segment j.
Description: This performance measure compares the observed and expected travel time. It quantifies deviations of the current from the historical travel time along a particular segment.</td><td>FDOT (Stevanovic and Mitrovic, 2019b)</td></tr>
<tr><td>Cumulative Volume Distribution Function</td><td>Derivation Method: All volumes ($V_{i,j}^{d}$) at time -i, day -d, at station- j will be taken into account to construct the cumulative volume distribution function. $F(v_{i,j}^{d})$ shows the probability of having the volume that is less or equal to $v_{i,j}^{d}$ where $v_{i,j}^{d}$ is the reported (or most recent) volume.
$F(v_{i,j}^{d}) = P(v_{i,j}^{d} \leq v_{i,j}^{d} \,|t = t_i \text{ and } s = s_j|)$
Expected Values: 0.5 $+/-$0.3 (in the described example it would be slightly less than 0.5).</td><td>FDOT (Stevanovic and Mitrovic, 2019b)</td></tr>
<tr><td>Cumulative Travel Time Distribution function</td><td>Derivation Method: All travel times ($TT_{i,j}^{d}$)) at time -i, day -d, along segment- j will be used to construct the cumulative travel time distribution function. This distribution function will be used to compute $F\left(tt_{i,j}^{d}\right)$.
$F(tt_{i,j}^{d})$ = P($TT_{i,j}^{d} \leq tt_{i,j}^{d} \,|t = t_i \text{ and } s = s_j|$)
Where $F(tt_{i,j}^{d})$ shows the probability of having travel time less than or equal to $tt_{i,j}^{d}$; $tt_{i,j}^{d}$ is the observed (or current) travel time.</td><td>FDOT (Stevanovic and Mitrovic, 2019b)</td></tr>
<tr><td>K-Cluster-Based Volume-Based Traffic Profile</td><td>Derivation Method: Historical volumes at single MVDS stations are used to develop the representative profiles for each cluster. Then, for a current volume obtained in real time, the method is assigned to one of the four traffic profiles that the observed volume most likely belongs to.
Output values: {Light Traffic, Moderate Traffic, High Traffic, Peak Traffic}
<table>
<tr><th>Sources</th><th>RT</th><th>Hist</th><th>Frequency</th><th>Freeway</th><th>Arterial</th></tr>
<tr><td>MVDS</td><td>Volume (V_{ij})</td><td>Volume $V_{h_{i,j}}^{d}$ & Thresholds</td><td>V_{ij}—1 min
$V_{h_{i,j}}^{d}$ -15 min</td><td>P-V S, C-V/C</td><td>P-V S, C-V/C</td></tr>
</table></td><td>FDOT (Stevanovic and Mitrovic, 2019b)</td></tr>
<tr><td>K-Cluster Travel Time-Based Traffic Profile</td><td>Derivation Method: Travel times along a segment is used to develop the representative profiles for each cluster. Then, current travel time is assigned to one of the four traffic profiles.
Output values: {Light Traffic, Moderate Traffic, High Traffic, Peak Traffic}
<table>
<tr><th>Sources</th><th>RT</th><th>Hist</th><th>Frequency</th><th>Freeway</th><th>Arterial</th></tr>
<tr><td>RITIS</td><td>(estimated) TT (TT_{ij})</td><td>$TT_{h_{i,j}}^{d}$ & Thresholds</td><td>TT_{ij} – 1 min
$TT_{h_{i,j}}^{d}$ -15 min</td><td>S, C</td><td>S, C</td></tr>
</table></td><td>FDOT (Stevanovic and Mitrovic, 2019b)</td></tr>
</table>

<table>
<tr><th>Performance Metric</th><th>Equation</th><th>DOT</th></tr>
<tr><td>ML-Based Predicted Performance Measures</td><td>Description: This PM predicts the volume that might be expected at an MVDS station in the next 5–30 minutes (short-term prediction).
Derivation Method: For given current and past volume data and corresponding temporal information (e.g., day and hour), the algorithm performs short-term predictions and returns the expected volume for the next 5 to 30 minutes. Various predictions methods ranging from moving averages to NNs and SVR will be evaluated, and the best performing method will be deployed.
Inputs: V_{ij}, $V_{i-5,j}$, $V_{i-10,j}$, t_i, d_i; V_{ij}- Volume at time i and station j. t- time, d-day
Output: $V_{i+5,j}$, $V_{i+10,j..}$ $V_{i+30,j}$
<table>
<tr><th>Sources</th><th>RT</th><th>Hist</th><th>Frequency</th><th>Freeway</th><th>Arterial</th></tr>
<tr><td>MVDS</td><td>V_{ij}</td><td>$V_{i-5,j}$,
$V_{i-10,j..}$</td><td>1 min</td><td>S, C</td><td>S, C</td></tr>
</table>
</td><td>FDOT (Stevanovic and Mitrovic, 2019b)</td></tr>
</table>

Note: S = Segment, C = Corridor, P = Point (MVDS station). * Performance metric is for both urban and rural facility type. All other performance metrics are only for urban facility type.

APPENDIX C: DATA DICTIONARY

Table 58. Data Dictionary.

Variable Code	Description
unique_id	Unique id representing individual segment
hwy	Name of the highway
District	District name
County	County name
ru_f_syste	Functional classification
num_lanes	Number of through lanes
adt_adj	Adjusted AADT
k_fac	Peak factor
d_fac	Directional distribution factor
Trk_aadt_p	Truck AADT pct
dhv	Truck design hourly volume pct
ln_miles	The length of the road segment in miles
tt_1721	TTAve over 5-year period (2017–2021)
P_WrkAge	Percent of population that is working aged 18 to 64 years
Pct_AO1	Percent of one-car households in CBG
Workers	Count of workers in census block group (CBG) (home location)
TotEmp	Total employment
D1C	Gross employment density (jobs/acre) on unprotected land
D2A_JPHH	Jobs per household
D3A	Total road network density
D3AAO	Network density in terms of facility miles of auto-oriented links per square miles
D5AR	Jobs within 45 minutes auto travel time, time decay
D5CR	Proportional accessibility to regional destinations—Auto: Employment accessibility expressed as a ratio of total core based statistical area (CBSA) accessibility
D5CRI	Regional centrality index—Auto: CBG [D5cr] score relative to max CBSA [D5cr] score
D5CE	Proportional accessibility to regional destinations—Auto: Working age population accessibility expressed as a ratio of total CBSA accessibility
D5CEI	Regional centrality index—Auto: CBG [D5ce] score relative to max CBSA [D5ce] score
SpdAve	Average speed
SpdStd	Standard deviation of speed
Spd85	85^{th} percentile speed determined
PSL	Post speed limit
TTAve	Average travel time
TTStd	Standard deviation of travel time
PVTT	Percent variation of travel time

APPENDIX D: VALUE OF RESEARCH

The TTI team conducted a value of research (VOR) analysis of TxDOT Research Project 0-7131 to produce an estimate of the benefit that the project will likely yield for TxDOT. The temporal scope for this analysis is an 11-year period (labeled as years 1–11), starting with the beginning of the 2-year project. The value of the project is described in terms of NPV and cost-benefit ratio (CBR), which are computed using economic discounting formulas.

The primary objective of TxDOT Research Project 0-7131 is to reduce congestion of urban and rural freeways in Texas. The project developed an interactive tool of congestion forecasting by applying AI models. The TTI team focused the VOR analysis on the overall benefits of safety and mobility improvement based on the advanced forecasting models and decision-making from the tool outcomes. In addition, recommendations from big data and AI platform investigations will also be considered benefits.

METHODOLOGY

The TTI team used a VOR template provided by TxDOT to compute the NPV and CBR measures. The template requires the following items:

- Project budget: $297,204.
- Project duration: 2.00 years.
- Expected value duration: 11 years (convention chosen by TxDOT).
- Discount rate: 3 percent (default value assumed by TxDOT).
- Expected value per year: $698,850.00.

The project's expected value per year is estimated based on savings obtained from reduced crashes. The analysis method is described in the following sections.

Concept

To conduct the VOR analysis, the following steps were taken:

1. Determine the reduced crash frequency by severity by comparing the expected crash outcomes from the AI models.
2. Provide approximation of congestion-related mobility benefits.
3. Provide approximation of big data and AI platform investigation-related benefits.
4. Apply the procedure to estimate the expected value of the research.

Input Data

The VOR analysis was conducted by randomly selecting 50 miles of urban freeways. The calculated benefit for 9 years is $5,293,596.00.

Crash Cost

The TTI team derived crash severity distribution proportions from the sample considered in TxDOT Research Project 0-7131. These proportions for the following crashes are as follows:

- Fatal (K): 1.11 percent.
- Incapacitating injury (A): 1.11 percent.
- Non-incapacitating injury (B): 2.22 percent.
- Possible injury (C): 8.89 percent.
- Property damage only (PDO): 86.67 percent.

The TTI team used the following to calculate the benefit from this project:

- First, the TTI team used crash costs from TxDOT's Highway Safety Improvement Program guideline. The crash value is $3.7 million for K and A crashes. The B crash value is $520,000. Second, the National Safety Council values of $155,000 and $51,000 for C and no injury (O) crashes, respectively, were used. The benefit from crash reduction is calculated as $4,938,000.00.
- The benefit from mobility improvement is considered to be $750,000, and the benefit from AI and big data explorations is considered to be $200,000.00.

Cost

The TTI team used an annual maintenance cost of $0 for analysis based on the assumption that TxDOT would provide the same amount of periodic maintenance and monitoring for the reconfigured sites as for the sites in their existing condition.

RESULTS

The TTI team conducted the VOR analysis using the SRPW program and obtained an annual VOR estimate of $698,850. This value represents the benefit that can be obtained if the results of the research project are used to analyze 50 miles of urban freeways.

Figure 39 summarizes the VOR calculations. The payback period for Research Project 0-7131 is 0.43 years, and the CBR is 16.30.

The findings shown in Figure 39 are as follows:

- The benefits included in the VOR calculations include only those incurred by TxDOT. In reality, other agencies (e.g., local and county agencies within Texas and other state DOTs) will be able to implement and benefit from the published findings from the project.
- The estimated benefits include crash reduction, mobility improvement, and big data/AI platform explorations, based on the assumption of tool and model usage. TxDOT will likely receive additional benefits that are more difficult to quantify.
- The VOR analysis focused on urban freeways. Both rural and urban freeway facilities may also realize similar benefits from the application of these project results. The estimated VOR, NPV, and CBR would increase if these sites were included in the analysis.

	Project #	0-7131		
	Project Name:	Leveraging Artificial Intelligence (AI) Techniques to Detect, Forecast, and Manage Freeway Congestion		
	Agency:	TTI	**Project Budget**	$ 297,204
	Project Duration (Yrs)	2.00	**Exp. Value (per Yr)**	$ 698,850
Expected Value Duration (Yrs)		10	**Discount Rate**	3%
Economic Value				
Total Savings:	$5,293,596		**Net Present Value (NPV):**	$ 4,843,994
Payback Period (Yrs):	0.43		**Cost Benefit Ratio (CBR):**	16.30

Years	Expected Value
1	-$170,852
2	-$126,352
3	$698,850
4	$698,850
5	$698,850
6	$698,850
7	$698,850
8	$698,850
9	$698,850
10	$698,850
11	$698,850

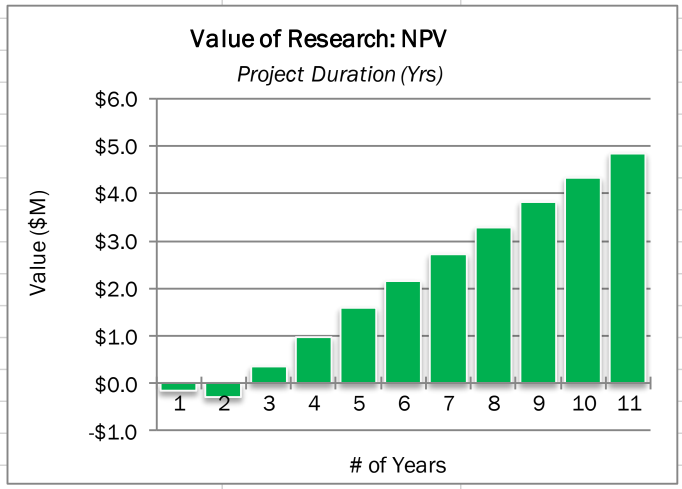


Figure 39. VOR Analysis Results.