Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

A DEEP MACHINE LEARNING APPROACH FOR PREDICTING FREEWAY WORK ZONE DELAY USING BIG DATA

by
Abdullah Shabarek

The introduction of deep learning and big data analytics may significantly elevate the performance of traffic speed prediction. Work zones become one of the most critical factors causing congestion impact, which reduces the mobility as well as traffic safety.

A comprehensive literature review on existing work zone delay prediction models (i.e., parametric, simulation and non-parametric models) is conducted in this research. The research shows the limitations of each model. Moreover, most previous modeling approaches did not consider user delay for connected freeways when predicting traffic speed under work zone conditions. This research proposes Deep Artificial Neural Network (Deep ANN) and Convolution Neural Network (CNN) traffic speed prediction models, for upstream freeway segments, including those on connected freeways, under work zone conditions.

The developed models are able to identify the congestion on the connected links in addition to the upstream mainline segments. The models predict traffic speed with work zone conditions based on traffic volume approaching the work zone, speed during normal conditions, work zone capacity, distance from work zone, vertical road gradient, downstream traffic volume and type of freeway segment. Moreover, the previous efforts in non-parametric approaches did not consider a solution to the overfitting problem of Artificial Neural Network (ANN). The proposed Deep ANN and CNN models use a dropout regularization to mitigate the overfitting issues. When comparing the CNN model

to the Deep ANN model and the results of the Work Zone Interactive Management APplication-Planning (WIMAP-P) model, the testing results show higher accuracy with the CNN model compared to the other two models. The CNN model has filters that extract useful inputs from previous layers and reduces the overfitting problems. Dropout regularization technique is used to prevent the co-adaptation of training data. The CNN model is calibrated by varying the number of neurons at each hidden layer, the number of hidden layers, the optimizer algorithm, the filter height and the filter stride. The results indicate that the CNN model outperforms Deep ANN and the model of WIMAP-P in predicting traffic speed under work zone conditions.

While traditional efforts were conducted previously on predicting traffic congestion on the upstream freeway segments, the developed CNN model helps transportation agencies in planning for work zones by including both connected freeways and the upstream segments when predicting traffic speed under work zone conditions. Therefore, transportation agencies can prepare more accurate congestion mitigation plans, and provide more accurate user delay plans.

A DEEP MACHINE LEARNING APPROACH FOR PREDICTING FREEWAY WORK ZONE DELAY USING BIG DATA

by
Abdullah Shabarek

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Transportation Engineering

John A. Reif, Jr. Department of Civil and Environmental Engineering

December 2020

.

APPROVAL PAGE

A DEEP MACHINE LEARNING APPROACH FOR PREDICTING FREEWAY WORK ZONE DELAY USING BIG DATA

Abdullah Shabarek

Dr. Steven I-Jy Chien, Dissertation Advisor — Date
Professor of Civil and Environmental Engineering, NJIT

Dr. Branislav Dimitrijevic, Committee Member — Date
Associate Professor of Civil and Environmental Engineering, NJIT

Dr. Joyoung Lee, Committee Member — Date
Associate Professor of Civil and Environmental Engineering, NJIT

Dr. Lazar Spasovic, Committee Member — Date
Professor of Civil and Environmental Engineering, NJIT

Dr. Taha F. Marhaba, Committee Member — Date
Professor and Chair of Department of Civil and Environmental Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Abdullah Shabarek

Degree: Doctor of Philosophy

Date: December 2020

Undergraduate and Graduate Education:

- Doctor of Philosophy in Transportation Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2020

- Master of Science in Engineering Management,
  New Jersey Institute of Technology, Newark, NJ, 2016

- Bachelor of Science in Civil Engineering,
  Aleppo University, Aleppo, Syria, 2013

Major: Transportation Engineering

Presentations and Publications:

Abdullah Shabarek, Steven Chien, and Soubhi Hadri, "Deep Learning Framework for Freeway Speed Prediction in Adverse Weather" Transportation Research Record, August 2020, https://doi.org/10.1177/0361198120947421

Abdullah Shabarek, Steven Chien, and Soubhi Hadri, "A Deep Learning Framework for Freeway Speed Prediction under Adverse Weather Conditions," The 99th Transportation Research Board Annual Meeting, Washington, D.C., Jan. 2020

Abdullah Shabarek and Steven Chien, "Reducing Delay Cost with a Conditional Transit Priority Control under Connected Vehicles Environment," the 19th Annual NJDOT Research Showcase at The Conference Center at Mercer County College, West Windsor, NJ, October 25, 2017 (poster)

Abdullah Shabarek, Joyoung Lee and Dejan Besenski, "A Hybrid Clustering Method Integrating Genetic Algorithm and Ordered Logit Model for Safety Prioritization of Highway Segments" " Intelligent Transportation Systems in New Jersey. Annual Meeting, New Brunswick, NJ, September 2018 (poster)

To My Family

ACKNOWLEDGMENT

I would like to thank Dr. Steven I-Jy Chien for his mentorship and patience throughout the dissertation. Many thanks to the committee members, Dr. Branislav Dimitrijevic, Dr. Joyoung Lee, Dr. Lazar Spasovic, and Dr. Taha F. Marhaba, for their additions. I would like to thank the members of Intelligent Transportation Research Center for their valuable help in providing some of the required data. Finally, I would like to thank my parents: Waddah and Nahla and my brother Tarek for their unlimited support.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF SYMBOLS

b_i^l		Bias of neuron i at hidden layer l
C	vph	Roadway capacity
C_d	$\$/zone$	Work zone delay cost
C_w	vph	Work zone capacity
D	$veh-hr$	Queuing delay
f_{HV}		Heavy-vehicle adjustment factor
f		CNN filter
i		Index of segments in upstream of work zone
j		Index of time intervals
L_j	mi	Queue length at time interval j
l_i	mi	Length of freeway segment i
l		Index of hidden layers
m		Number of time intervals
N_0		Number of open lanes within a work zone
n		Number of TMC segments upstream a work zone
P_c		Percentage of passenger cars and heavy vehicles

P_t		Percentage of heavy vehicles
Q	vph	Demand flow rate
R	vph	HCM adjustment factor for on-ramps
r^l		Vector of independent Bernoulli random variables at hidden layer l
s_{ij}	mph	Traffic speed under normal conditions for segment i at time j
t_1	hr	Lane closure duration
t_2	hr	Queue discharging time
V_{ij}		Traffic volume for segment i at time j
w_q^l		Weight of neuron q at hidden layer l
y_q^l		Output of each neuron q to hidden layer l
y_{ij}	mph	Actual speed for segment j at time i
$\hat{y}_{ij}$	mph	Predicted speed for segment j at time i
μ_c	$\$/veh-hr$	Value of time for passenger cars
μ_t	$\$/veh-hr$	Value of time for heavy vehicles
τ_{ij}		Congestion status of segment i at time j
$\otimes$		Convolution operation

CHAPTER 1
INTRODUCTION

1.1 Background and Problem Statement

Transportation infrastructures, such as freeways, require frequent maintenance that involves lane closures. With the increase of vehicles-mile traveled, work zones became the second greatest contributor to non-recurrent congestion. Work zone congestion accounts for 24 percent of non-recurrent congestion and 10 percent of the overall congestion (FHWA, 2019). Work zone congestion occurs on the upstream mainline segments and depending on the characteristic of the work zone, traffic volume, and geographic conditions, work zone congestion can spill back to upstream connected freeways.

Transportation systems provide means for passengers and goods movement. With the important role of these systems, work zones are required to maintain and extend the life cycle of the infrastructure. However, work zones with lane closures are accounted for congestion. For this reason, transportation systems aim to predict the congestion upstream work zone and the spillback to other connected freeways. One of the elements in determining work zone congestion is work zone capacity, which is the maximum number of vehicles entering a work zone. When work zone capacity is less than the traffic volume approaching work zone, a queue forms, which leads to reduction in traffic speed upstream work zone area. Other factors include the vertical gradient, traffic speed during normal conditions, traffic volume, and distance of the upstream segment to the work zone. While the effect of each of these factors overestimating upstream traffic speed is not explicit, a non-parametric approach is desirable to predict traffic speed with work zone conditions on the mainline and upstream segments.

Parametric and simulation models are typically used for predicting traffic speed with work zone conditions. Non-parametric approaches can predict work zone speed when historical data is available. Moreover, non-parametric approaches do not assume a distribution of the sampled data. Since real-world data tend not to follow a well-known distribution, the parametric model's approach may reduce the accuracy of the prediction results. Deep Artificial Neural Network (Deep ANN), a non-parametric model commonly used as a prediction model, is a machine-learning technique that are used to identify traffic patterns and traffic speed. Previous studies used ANN to predict the traffic speed upstream work zone on the mainline segments only. Since ANN models include one or two hidden layers, they cannot capture the complexity of larger scale networks that include connected freeways. Additionally, ANN models would be more susceptible to overfitting since some of the overfitting mitigation techniques are hard to be implemented in two hidden layers (Schmidhuber, 2015). With sufficient data, a deep learning approach is more suitable to predict the traffic speed on upstream mainline, ramps, and connected freeways. Furthermore, previous studies suffered from the overfitting problems, when using the traditional ANN models. Therefore, the accuracy of the traffic speed prediction may be affected, when predicting work zone conditions that are not included in the training dataset.

As discussed, traditional ANN models suffered from overfitting problems and did not extend the study to include congestion spillback to connected freeways. Therefore, the CNN structure is adopted in this study to predict traffic speed on the upstream segments, including both the mainline and the connected freeway segments. The proposed CNN model is expected to mitigate the overfitting problems by extracting only important features from the layer it is applied at. The CNN model captures the spatial-temporal

impact of a work zone on the upstream segments. Once the congestion approaches the ramp segments, a spillback of congestion usually occurs on the ramp and connected freeways. This research focuses on the factors affecting the prediction of congestion on upstream ramps and connectors in addition to upstream mainline freeway segments. Moreover, the study recognizes the overfitting problem of Deep ANN models and suggests a dropout technique to prevent the co-adaptation on training data. A numerical evaluation is conducted on Interstate-287 to compare the predicted speed results of the CNN model with previous prediction models.

The developed model in this study can be used for further planning purposes for work zone congestion prediction in which congestion can be predicted on connected freeways in addition to mainline segments. For instance, transportation planners can use the model to predict the delay the queue lengths prior to performing the work zone. Thus, transportation planners can see whether a work zone would produce a congestion spillback to other freeways or not. A work zone congestion that has a spillback on other freeways would produce higher user delay. Therefore, public agencies can increase the coverage in predicting user delay when preparing congestion mitigation plans. This research can be used to predict scenarios in which congestion spill backs on other connected freeways; thus, it can be useful for supporting decisions where to deploy queue warning systems on the upstream connected freeways.

1.2 Objective and Work Scope

The objective of this study is to develop a model that is able to predict the effect of a work zone on the mainline freeway and connected roads using a mass amount of data. Therefore, this research improves the coverage and accuracy of the previous studies in predicting

traffic speed under work zone conditions. In the development of these two models, various parameters of the network are optimized. The limitations of the previous models are discussed. One of the problems in previous models is the overfitting issue, which results in inaccurate prediction values. Dropout is discussed in this research as a mean to mitigate the overfitting problem. Then, two non-parametric models (i.e., Deep ANN and CNN) are developed and evaluated under various work zone, weather, and traffic conditions. The Deep ANN and CNN models utilize various data types including road geometry, work zone data, probe vehicle data, and traffic volume data. Both of the CNN and Deep ANN models use the drop out regularization to overcome the overfitting problem in previous models.

The scope of the work includes predicting traffic speed under work zone conditions on the freeway segments in New Jersey. The scope of the modeling approach is conducted on selected non-parametric approaches (i.e., Deep ANN, and CNN). The proposed CNN model can predict traffic speed upstream of a work zone including both the mainline and the connected freeway segments. Consequently, the predicted work zone traffic speed can include multiple freeways for congestion mitigation plans. Moreover, when a congestion occurs on upstream connected freeways due to work zone, queues can be formed. Thus, transportation agencies can mitigate any safety problems associated with the queue formation on the predicted congested upstream freeway segments.

1.3 Organization

This dissertation is organized into six chapters as shown in Figure 1.1. Chapter 1 focuses on the background and the gaps of the previous work zone speed prediction models. This chapter demonstrates the importance of this research. Chapter 2 discusses the previous work on work zone speed prediction, and reviews the factors affecting the traffic speed.

Chapter 3 discusses the data acquisition from various sources. Moreover, the structures and functionalities of Deep ANN model and CNN model are presented. Chapter 4 discusses the evaluation of Deep ANN and CNN models under various traffic and weather conditions with the database developed in Chapter 3. A case study is discussed in Chapter 5 for demonstrating the applicability of the proposed models. Finally, the research findings are concluded in Chapter 6 in addition to future potential studies.

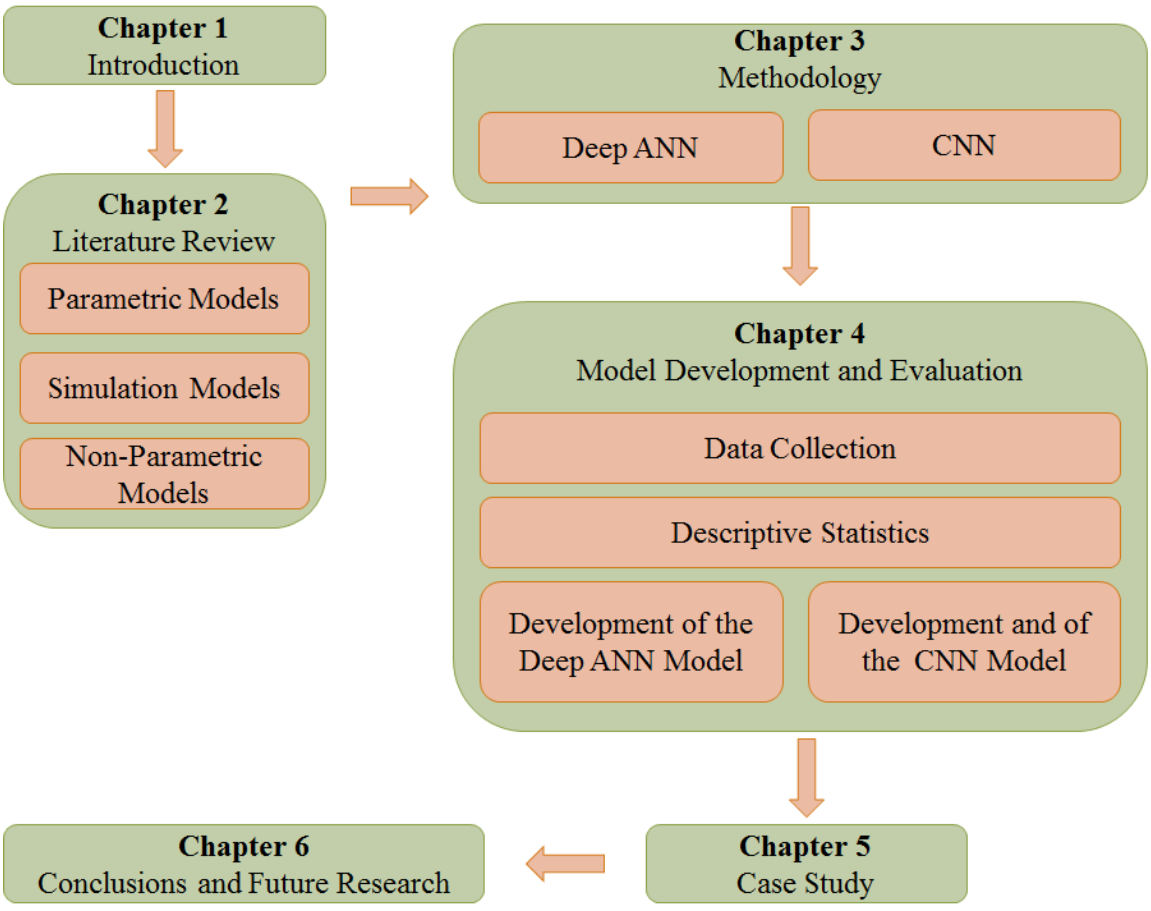


Figure 1.1 Organization of the dissertation.

CHAPTER 2
LITERATURE REVIEW

This chapter examines the previous research efforts in predicting the traffic speed under work zone conditions, including the parametric approaches, the simulation approaches, and the non-parametric approaches. Then, a research of the available tools and methods is conducted from previous work zone traffic speed predication methods. Finally, this chapter explores the configuration of Deep ANN and CNN in terms of applicable and recommended parameters.

2.1 Work Zone Delay Prediction

This section describes the previous approaches in predicting work zone traffic speed. Work zones usually produce congestion on the upstream segments. Work zone delay is the additional time vehicles need to travel through a work zone segment compared to normal conditions in which a work zone does not exist (Ullman et al., 2011; Weng and Meng, 2013). Predicting a work zone delay is important for transportation agencies for planning purposes. First, this section previews the factors affecting a work zone delay. Second, this section explores the previous research efforts in the main three categories of predicting work zone delay: parametric approaches, simulation approaches and non-parametric approaches.

2.1.1 Factors Affecting Work Zone Delay

There are multiple factors affecting work zone delay (e.g., work zone intensity, work zone starting/ending time, the number of closed lanes, the number available lanes, the traffic

volume approaching work zone, the work zone capacity, the truck percentage, the vertical gradient, the traffic volume downstream an ramp segment, weather conditions)

Work zone delay increases as a work zone intensity increase. Work zone intensities are categorized into three intensity types: low, medium, and high (Karim and Adeli, 2003). High Capacity Manual (HCM 2010) provides work zone intensity as a factor in determining work zone capacity. Additionally, a work zone delay change depends on the work zone starting/ending time. Work zones that occur during the night differ significantly from daytime work zones (Chien and Schonfeld, 2001; Tang and Chien, 2008).

Previous studies indicate the effect of the number of closed lanes and the number of available lanes on a work zone delay (Krammes and Lopez, 1994; Kim et al., 2001; Chung et al., 2012). The increase in the number of closed lanes reduces the available number of lanes for traffic and reduces work zone capacity; as a result, this increases the work zone delay. Furthermore, as the traffic volume approaching a work zone increases, the work zone delay increases (Dudek and Richards, 1982; Krammes and Lopez, 1994; Chien and Schonfeld, 2001; Chien et al., 2002; Tang and Chien, 2008; Du and Chien, 2014). The work zone capacity represents the maximum number of vehicles entering a work zone. Previous research efforts estimate the effect of a work zone capacity in determining the work zone delay (Du et al., 2017; Du et al., 2015).

Trucks have lower speed compared to regular passenger cars. Therefore, truck percentage affects the maximum number of vehicles entering work zone and work zone delay (Du et al., 2015). The vertical gradient also affects the work zone delay, which increases with the increase of vertical gradient, especially as the truck percentage is high (Kim et al., 2001). The increase of traffic flow upstream work zone attributes to the increase

in traffic delay in these upstream segments (Schroeder & Rouphail, 2010). Upstream work zone segments can be classified into three main categories: upstream mainline segments, upstream ramp segments, and upstream connected freeway segments. Depending on the type of upstream segment, the work zone delay can be significantly different (Ullman & Dudek, 2003; Karim & Adeli, 2003). Other factors include weather conditions and driver's behavior. Weather conditions reduce work zone capacity by increasing the headways between the vehicles (HCM, 2010).

2.1.2 Parametric Models

Parametric models are commonly used for predicting traffic speed under work zone conditions. The deterministic approaches, which are parametric approaches, follow the idea in Figure 2.1. The shaded area represents the total work zone queuing delay in (veh-hr). The inputs of the parametric models are roadway capacity during normal conditions C, traffic volume Q, roadway capacity under work zone conditions C_w, the starting/ending time of work zone.

McCoy et al. (1980) defines the work zone delay as the difference between travel times under work zone and normal conditions, which does not consider the condition as the traffic volume is greater than the work zone capacity. Chien and Schonfeld (2001), on the other hand, considered queuing and moving delay but the variation of traffic volume over work zone duration was simplified.

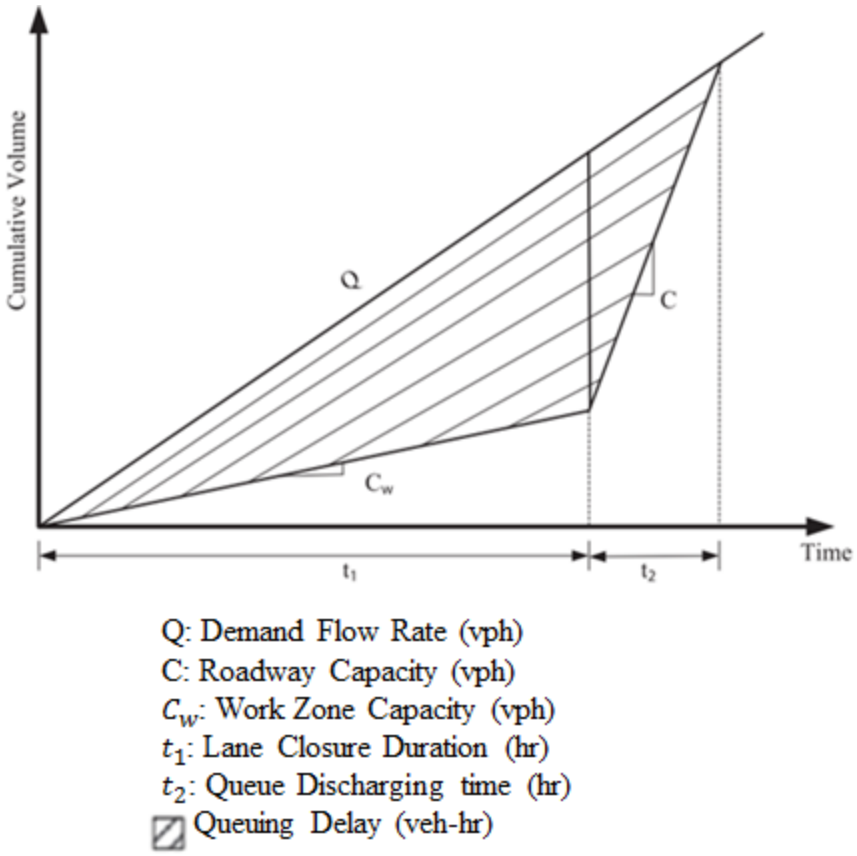


Figure 2.1 A queuing model to determine work zone delay.

Du and Chien (2014) considered a time-variant parametric model to calculate work zone delay considering the effect of heavy vehicles and light conditions. Traffic speed during shoulder closures is reduced through a work zone segment due to the limitation of the work zone speed limit.

The work zone capacity depends on the value of traffic volume approaching work zone compared to the work zone capacity. Du and Chien (2014) modeled the work zone delay considering shoulder use to increase the work zone capacity. It was concluded with this paper the adjustment factors required to adjust the work zone capacity under shoulder usage and various lane closure types. The paper showed that shoulder usage is required to

reduce work zone delay especially during the peak hours' times. These models, however, were not able to capture the spatio-temporal effect of work zones on the upstream segments.

Another parametric approach relies on shockwave theory in which traffic flow is considered similar to fluid flow in terms of its movement (Lighthill and Whitham 1955; Richards, 1956). The shockwave theory utilizes the based on the spatio-temporal traffic flow transition to estimate the queue length. Benekohal et al., (2013) used the shockwave theory to calculate traffic delay and queue length under work zone conditions. The shockwave model uses jam density, speed under normal and work zone conditions, traffic volume, work zone capacity, critical density, and free flow speed to track the congestion spillback on upstream mainline segments (Habtemichael et al., 2015). Thus, developing an work zone congestion prediction using the shockwave theory is challenging due to the scarce of available data that are identical to the shockwave theory's parameters.

2.1.2 Simulation Models

Several simulation models are used to predict work zone delay such as the model used in QuickZone (Chitturi and Benekohal, 2004), FRESIM (Chitturi and Benekohal, 2004), ARENA (Maze and Kamyab, 1999), CORSIM (Chien et al., 2002), PARAMICS (Wang et al., 2002), and VISSIM (Edara et al., 2013). Simulation models are more accurate than parametric models in predicting work zone delay. Earlier models use less calibrated parameters compared to newer simulation models (e.g., driver behavior) (Bloomberg & Dale, 2000).

Chitturi and Benekohal (2004) compared work zone delay results from QUEWZ, FRESIM, and QuickZone against field data collected from 14 freeway segments under

work zone conditions in Illinois. The results suggest that QUEWZ and FRESIM overestimated queuing delay caused by work zones, while QuickZone underestimated delay and queue length.

Maze and Kamyab, (1999) developed a simulation model, ARENA, that predicts work zone delay. Traffic delay is calculated based on the average travel time produced by the simulation model. It was observed that the increase of traffic volume increase work zone delay. ARENA relies on parametric queuing method to calculate work zone delay. It was found that ARENA simulation model underestimates the work zone delay

Chien et al., (2002) developed a model for predicting work zone delay based on simulation data from CORSIM taking into consideration work zone configuration, road geometry, and traffic volume distribution over time. To obtain accurate results, calibration is conducted to match actual work zone conditions. A case study was conducted for work zone on Interstate-80 in New Jersey to show the applicability of the developed CNN model. Moreover, Yang et al. (2008) used CORSIM for predicting work zone delay under saturated and unsaturated traffic conditions. The results show that CORSIM predict work zone delay more accurately under unsaturated conditions comparing to deterministic approaches. However, it was found that CORSIM underestimate work zone delay under saturated conditions. Therefore, deterministic approaches outperform CORSIM under saturated traffic conditions.

Edara et al. (2013) developed a model that uses VISSIM for predicting work zone delay. The results indicate high traffic speed fidelity compared to other simulation approaches. Nevertheless, VISSIM requires extensive network calibration prior to predict work zone delay. Network calibration can be done using travel time or queue length. It was

found that calibrating the network using travel time obtained from probe vehicle data recognizes higher accuracy than queue length calibration. Later, Du et al. (2014) developed a simulation model approach powered by VISSIM to predict work zone capacity under various situation.

Simulation models yield higher accuracy than parametric approaches. However, simulation approaches require extensive calibration for each work zone, which makes them computationally expensive. In addition, simulation models cannot match real-time data in terms of driver behaviors, route choices, and many other factors that cannot be included in the simulation model.

2.1.3 Non-parametric Models

While simulation approaches provide acceptable results for work zone traffic speed prediction compared to parametric approaches, simulation models don not reflect real-work zone traffic speed data. Therefore, non-parametric models are used to capture the effects of multiple parameters on work zone delay in which no mathematical relationship is provided. Artificial Neural Networks (ANN) is imitated from brain neurons functionality. Neurons in the brain is connected at different layers to communicate the information from one part to another (Adeli and Hung, 1995).

In transportation systems, previous research efforts focused on the prediction of traffic flow (Adeli and Hung 1995; Adeli and Park 1998; Adeli 2001; Hasebe et al., 1999; Neubert et al., 1999; Zhang et al., 1997; Park et al. 1998; Suzuki et al. 2000; Du et al., 2014). Park et al., (1998) predicts traffic flow using Radial-Basis Neural Networks (RBFNN) whereas Zhang et al., (1997) developed a Neural Network that uses Back

propagation method to predict traffic flow. Suzuki et al., (2000) developed a back-propagation model that is able to simulate traffic flow in origin-destination networks.

Karim and Adeli (2003) uses RBFNN to predict work zone capacity based on work zone length, work zone layout, number of closed lanes, number of available lanes, heavy vehicle percentage, work zone intensity, work zone vertical gradient, work zone speed, and lane width. Although the results were acceptable, the training data was limited. Thus, the finding of the RBFNN needed more verification. Later, Du et al., (2014) developed a hybrid model of a simulation approach and ANN to predict work zone capacity. Furthermore, a comparison analysis between ANN and Support Vector Machine (SVM) was conducted. The research found that SVM outperforms ANN in predicting work zone capacity. However, the SVM model was trained based on simulated data because of the limitation of the availability of traffic volume data.

As illustrated, previous studies used ANN models for predicting work zone capacity. Yet, these research efforts do not predict work zone traffic speed. Vemuri et al. (1998) uses an ANN with sigmoid function to predict work zone delay. Travel time data was collected from loop detectors as vehicles pass from one detector to the next one. Travel time data is simulated to predict travel time delay; hence, the result of the model needs to be verified with actual travel time delay data. Ghosh-Dastidar and Adeli (2006) used feed-forward ANN model to predict traffic speed under work zone conditions. The research used simulated data and verified the results with five examples. However, this research did not provide a generalized prediction model due to the marginal number of tested samples.

Traffic speed prediction has been modeled using non-parametric approach. Zhang et al., (2020) predicted traffic speed under normal conditions using three dimensions CNN

(3D). Du et al., (2017) developed an ANN model integrated with SVM model to predict traffic speed under work zone conditions. The SVM model predicts work zone capacity using a simulation approach (i.e., VisSim). The study used both actual and simulated data to predict work zone delay on upstream mainline segments. The model used work zone capacity simulated from Du et al., (2014) as a factor in the developed ANN model. The input of the ANN network is weighted speed and distance from freeway. Nevertheless, the research did not consider the complexity of upstream ramp and connected freeway segments. Moreover, the results of the research need to be investigated since it is a mixed model between actual and simulated data.

Non-parametric models are able to predict the traffic speed in less computational efforts compared to simulation approaches. Moreover, parametric models assume a distribution over the prediction function whereas non-parametric approaches do not have a distribution assumption for the trained data (Simar & Wilson, 2000). Random forest models construct a decision tree that is able to predict the output of a given model. The models are commonly used in classification problems. Dogru & Subasi, (2018) used the random forest model to predict the injury levels of accidents. Other studies use the random forest models for transportation prediction purposes (Urbancic et al., 2018; Elhenawy & Rakha, 2017). Non-parametric models have many branches in which CNN model recognizes the best suited model for predicting traffic speed. CNN model has a Max-pooling layer that mitigate the effect overfitting in the training model. Nguyen et al., 2019 found that CNN predicts traffic speed more accurately than other deep learning models due to the ability to extract the inputs in multiple consecutive time frames.

This research utilizes the CNN structure for predicting traffic speed under work zone conditions for upstream mainline and connected freeways. Table 2.1 summarizes the advantages and disadvantages of the parametric, simulation, and non-parametric models. While parametric approaches provide a quick method to determine the outputs, non-parametric approaches and simulation approaches require substantial amount of time finding the optimal structure and calibrating the model's parameters. Non-parametric approaches, when compared to other simulation approaches, can be quickly scaled to multiple scenarios.

Table 2.1 The Advantages and Disadvantages of Various Modeling Approaches

Model Type	Advantages	Disadvantages
Parametric Models	• Transferability • scalability • Inexpensive computational time	• Assumption of a distribution shape of the data • Difficulties in estimating the temporal and spatial traffic speed accurately by a simple mathematical formula
Simulation Models	• High fidelity for well calibrated models • Requirements for data is less than the other methods.	• Representing a work zone on a specific roadway (Not scalable or transferrable) • Requiring high computation and calibration time
Non-Parametric Models	• Scalability and Extensibility • Less computational time compared to simulation models • The data distribution is not required.	• Requiring more data for model development, training and validation processes • Requires substantial efforts to determine the model structure

Non-parametric approaches have two general purposes: prediction purposes, and optimization purposes. This study is concerned with prediction purposes to predict traffic speed under work zone conditions. Non-parametric models differ in their functionalities

and structures. Table 2.2 demonstrates the difference between Deep ANN models, CNN models, and random forests models. While Deep ANN models provide high fidelity compared to traditional ANN models, Deep ANN models still have problems with overfitting during the training processes. Thus, the CNN structure reduces the overfitting problems by using filters to extract the important features from the model inputs. Other non-parametric approaches (i.e., random forest models) are commonly used for classification problems. However, the problem in this research includes a numerical output that is represented by traffic speed under work zone conditions. Therefore, the CNN structure is better suited for traffic speed prediction compared to other non-parametric models.

Table 2.2 The Advantages and Disadvantages of Various Non-Parametric Models

Model Type	Advantages	Disadvantages
ANN	• Has the ability to predict model outputs with approximately two inputs	• Has accuracy problems when the model inputs exceed two inputs
Deep ANN	• Has high fidelity for more sophisticated models that have more inputs compared to ANN models	• Has overfitting problems that reduces the accuracy of the testing results
CNN	• Uses filters to reduce the overfitting issues of Deep ANN. Thus, improve the accuracy of the results	• Requires substantial efforts in modeling development
Random Forest	• Has high reliability for feature interpretability • Is more suitable for classification problems	• Does not perform well when more input variables are included in the model • Requires much higher computational powers and time for training and testing compared to other Deep ANN models

2.2 Data Collection Technologies

The traffic speed during work zones, in previous studies, was usually obtained from loop detectors. Loop detectors have been used to measure average speed between two points on a roadway segment. However, using loop detectors requires high maintenance and installation costs. With the advancement of big data technologies, new methods have been used to collect traffic speed (e.g., GPS sensors, toll booth sensors, Uber data, mobile devices, floating vehicles). These new technologies provide low-cost big data acquisition tools. Therefore, the proportion of the data provided by loop detectors has been decreasing while new probe vehicle technologies' proportion has been increasing (Burt et al., 2014).

Previous studies investigated the probe vehicle data technologies (i.e., INRIX), and showed that these technologies provide accurate traffic speed data (Elhenawy et al., 2014; Chen & Rakha, 2014; Haghani et al., 2009; Turner & Qu, 2013). One notable project involves studying the probe vehicle data on Interstate-95 corridor, which is located along the eastern coast of the United States. The project found that the INRIX database produce reliable traffic speed data under various scenarios (i.e., accidents). However, one study investigated the probe vehicle data in all roadways in Iowa, over four years span. The results show that INRIX speed is more reliable for Interstate roadways compared to non-interstate ones. Moreover, INRIX speed is more reliable during daytimes between 6:00 AM and 10:00 PM compared to the ones between 10:00 PM and 6:00 AM (Ahsani et al., 2019). One reason for the change in the change in INRIX data is the number of probes available during periods of time. The greater the number of probes available, the higher the confidence score is provided (Eshragh et al., 2017; Ahsani et al., 2019).

Bluetooth is a wireless technology that allows various devices to connect to each other. The Bluetooth technology enables transferring the information over short ranges

(e.g., 10 m, 100 m), depending on the wireless frequency and the Bluetooth hardware type (Bronzi, 2017). The Bluetooth manufacturers produce a unique identification number for each Bluetooth device, also known as, Median Access Control (MAC) address, which has have been implemented in a variety of electronic devices (e.g., vehicles, headphones, smartphones, and watches). Therefore, when a vehicle that is equipped with a Bluetooth passes between known locations of Bluetooth sensors, traffic speed can be captured.

In travel time data acquisition, Bluetooth devices have been used to estimate various transportation performance measurement, which estimate the space mean speed between two known MAC locations based on the time stamps of individual vehicles passing through the locations. There are other applications for Bluetooth sensors including origin-destination studies and queue length estimations. One of the advantages of using Bluetooth sensors is protecting the privacy of the users, as Bluetooth manufacturers do not track their customers through the MAC address. Consequently, the Bluetooth sensors do not recognize the vehicles users, enhancing the privacy of the collected data (Boukhechba et al., 2017).

The Bluetooth sensors, nevertheless, require a sample size to enhance the accuracy of the collected travel time. The required sample size is four percent for a 36,000 Average Daily Traffic (ADT) or greater roadways. When the ADT is lower, the required sample size becomes greater (Puckett & Vickich, 2010). In general, a sample size that ranges between five percent and seven percent would be enough for estimating a reliable travel time (Tarnoff et al., 2009). Another advantages of using Bluetooth sensors is the low cost of production and maintenance compared to other travel time data collection methods.

Toll tags is another technology that relies on floating car concept to predict travel time. The main purpose of deploying toll tags is to collect the tolls from the vehicles without the need to stop and pay. On the other hand, toll tags can be used to calculate travel time between two locations. The toll tags system for travel time data acquisition requires four components: the electronic tag in the floating vehicles, the readers, the antennas, and a central computer to perform the analysis (Wright et al., 2001). A vehicle that has an electronic tag if passes through a toll tag system, a toll identification number is recorded in the system with a specific location and time. When the same vehicle passes through another toll tag system, the system records the location and time, and the central computer calculate the travel time between these two locations. The toll identification number is protected in the system by encoding the number for privacy concerns. The toll tagging system in estimating travel time has been expanding in recent years. One study shows that the floating vehicles data has increased in both the coverage and the granularity over the years between 2013 and 2016 (Ahsani et al., 2019).

In the State of New Jersey, electronic toll tags, 79.6% of the registered vehicles use E-Z pass, which is an electronic toll tag (New Jersey Turnpike Authority, 2020). As the number of E-Z pass users is projected to increase, the accuracy of the travel time prediction using electronic toll tags is projected to increase. One of the disadvantages of toll tags travel time data collection is the coverage area, as some roads are toll-free, and they lack electronic toll tags infrastructure in place.

Another common travel time data acquisition technology is a radar sensor, which is a non-intrusive data technology. The radars are mounted at the side of the roadway, and as vehicles pass by, data is collected. The collected data include traffic volume at each lane,

traffic speed, and traffic density for each lane. Furthermore, the collected data can be calculated for all combined lanes. The collected data are sent to the cloud or a central computer for storage through internet cable. One radar sensor can cover multiple lanes in both directions, replacing the functionality of several loop detectors. The coverage of radar sensor can reach a width up to 250 ft from the pole, depending on the height of the radar sensor and the frequency of the equipment (Nyfors, 2000). The radar sensors are not affected by weather conditions and do not require maintenance, unlike loop detectors that require constant maintenance.

Closed-Circuit Television (CCTV) is another technology for traffic data acquisition, which can be used to collected traffic volume data based on pixeled images to identify vehicles count (Im et al., 2016). CCTV can be used to measure traffic speed by tracking the speed individual vehicles traveling between designated points on the camera screen (Cathey & Dailey, 2005). CCTV technology offers transportation management agencies with an insight of traffic conditions, when the cause of the congestion (e.g., incidents). Multiple research studies have implemented different CNN structure to detect and track vehicles (Chung & Sohn, 2017; Bochinski et al., 2016; Dorai et al., 2016). One type of the CNN structures, named as, You Only Look Once (YOLO), detects vehicles using neural network to a full image, and divides the various regions to detect the vehicles boundaries (Sreekumar et al., 2017). The CCTV technologies face challenges in terms of the accuracy of the vehicle detection method during low-visibility situations (e.g., night conditions, foggy conditions, rainy conditions, snow conditions) (Hahm et al., 2017).

Floating vehicle data are provided through various vendors (e.g., INRIX, TomTom, HERE). These floating vehicle data are provided from GPS sensors, which can capture

traffic speed at one second interval (Mudge et al., 2013). The sources of these floating vehicles data are GPS enabled vehicles (e.g., taxi vehicles, trucks, and smartphone enabled vehicles) (Seymour et al., 2011). INRIX data is divided into three categories: real-time data that has a confidence score of 30, historical data with a confidence score of 10, and a mixed data that has a 20-confidence score (Middleton et al., 2011). INRIX data is reported each one minute for over five million miles over forty countries (INRIX, 2018). The reported speed can be aggregated into 5 minutes, 15 minutes, or 1-hour intervals. Unlike other technologies, the INRIX data do not need any installation or maintenance costs, as smartphone based floating vehicles is increasingly used. On the other hand, INRIX data is more biased toward commercial trips as the data providers are mainly collected from long-hauled trucks and taxi vehicles (Hard et al., 2017).

2.3 Tools for Work Zone Congestion Prediction

This section will describe various tools that are used by transportation agencies to predict the work zone congestion.

Memmott and Dudek (1982) developed a work zone delay prediction model called Queue and User Cost Evaluation of Work Zone (QUEWZ). The model predicts work zone user delay cost on four and six lane multilane highways. Later, a developed model of QUEWZ, QUEWZ-98, estimates work zone capacity based on HCM 2000 (Benekohal et al., 2003). Edara and Cottrell, (2007) identified, through a survey of 19 states, the potential use of QUEWZ for predicting user delay cost due to work zone lane closures. The responses indicate that QUEWZ is an easy tool to be used in addition to giving quick results. However, the responses indicate simplicity in the predicted results as the QUEWZ was calibrated based on Texas freeway segments that has frontage roads. The key

limitations of QUEWZ include network configuration and the ability to adjust upstream ramps to count for diversion routes (Batson et al., 2009).

Memmott and Dudek, (1982) developed a tool for predicting users' delay cost and queue length called Queue and User Cost Evaluation of Work Zone (QUEWZ). The model considers the traffic volume and truck percentage in determining the users' delay costs due to a specific lane closure type. An improved version of QUEWZ, QUEWZ-98, was established to predict work zone capacity based on HCM 2000. Moreover, QUEWZ-98 added the emission costs to the user cost. The simulation model provides an option to determine the optimal work zone schedule time that minimizes road user costs (Benekohal et al., 2003).

QuickZone, an FHWA work zone delay application, is developed to predict work zone delay and maximum queue length (Mitretek System, 2000). The model of QuickZone is a deterministic model, in which all the model inputs are provided in a Microsoft Excel-based model developed. Thus, it is easier to use as a predictive work zone delay tool, compared to other work zone congestion prediction tools. One of the limitations of QuickZone is the limitations of the input parameters.

Wisconsin Department of Transportation developed a tool, Work Zone Capacity Analysis Tool (WZCAT), to predict delays due to work zone lane closures. The tool uses deterministic approaches for estimating work zone delay. Therefore, the tool is not able to have results close to real-time results.

Iteris Performance Management System (iPeMS) was developed for work zone delay prediction. iPeMS integrates real-time data from sensors and other ITS devices and stores them in a big data storage. The data is used for work zone delay prediction and travel

time estimation. The predicted model is based on historical and real-time information (Choe et al., 2002).

Another common tool is the Work Zone Traffic Analysis (WZTA) that is used in the state of Oregon. The tool provides a GIS map to be able to visualize the whole freeway network in Oregon. Oregon Department of Transportation (ODOT) provides the milepost start of the work zone and the milepost end on a selected direction of a selected highway. Moreover, the user specifies the number of closed lanes in the work zone will result in addition to the schedule of the work zone.

The tool uses a parametric approach to predict traffic speed during work zone schedule in the upstream mainline segments. The approach modifies the predicted values depending on studies in the state of Oregon (e.g., seasonal periods, terrain grade, and availability of the information at specific locations). Additionally, the approach accounts for special events timing for additional congestion by updating the calendar of the software accordingly. The tool is published as a web-based application in which the interface is illustrated in Figure 2.3 (Oregon Department of Transportation, 2010).

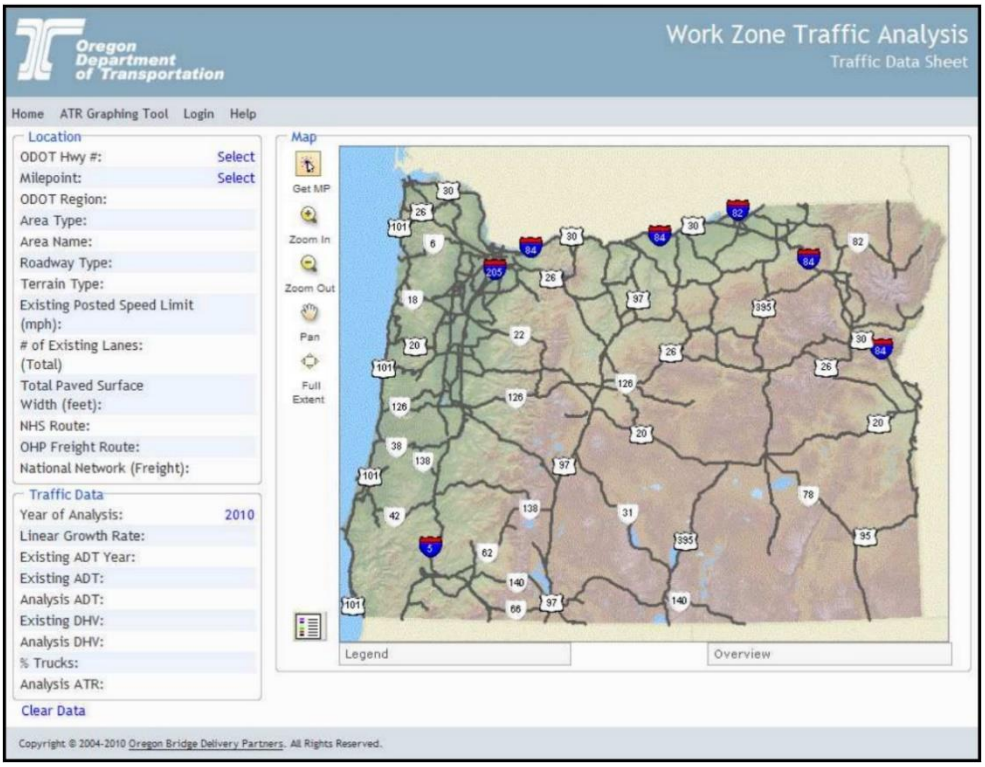


Figure 2.2 WZTA web-based interface.
Source:(Oregon Department of Transportation, 2010).

In the State of New Jersey, Rutgers Interactive Lane Closure Application (RILCA) is used to predict work zone delays at New Jersey Turnpike and Garden State Parkway (Bartin et al., 2012). RILCA, however, does not include real-time data. RILCA is a tool used to provide traffic volume for the routes, between two specified date and time inputs and two points. The tool uses a deterministic approach to schedule short- and long-term work zones and predict the delay costs and queue length accordingly. The queue length is determined when a particular segment has a volume that is higher than the roadway capacity/. One of the advantages that RILCA provides is collecting traffic volume at the toll booths providing transportation agencies with better information. Nevertheless, RILCA uses parametric approaches that can only work for the specified two routes.

Moreover, the tool does not consider delay costs due to congestion spillback on other freeways.

Chien et al. (2016) developed Work Zone Interactive Management Application (WIMAP-P). Figure 2.2 demonstrates the system framework of WIMAP-P. WIMAP-P predicts work zone speed using a data analysis on five different databases: Plan4Safety, OpenReach, NJCMS, NJSLD, and probe vehicle databases. WIMAP-P is based on the model developed in (Du et al., 2017); as a result, the model is a hybrid model of actual and simulation results. WIMAP-P was developed based on the work zone data between 2013 and 2014. Additionally, WIMAP-P predicts work zone speed on the mainline only, without including other connected freeways. Therefore, there is a need for an actual data model that is able predict work zone speed on the mainline and the connected freeways.

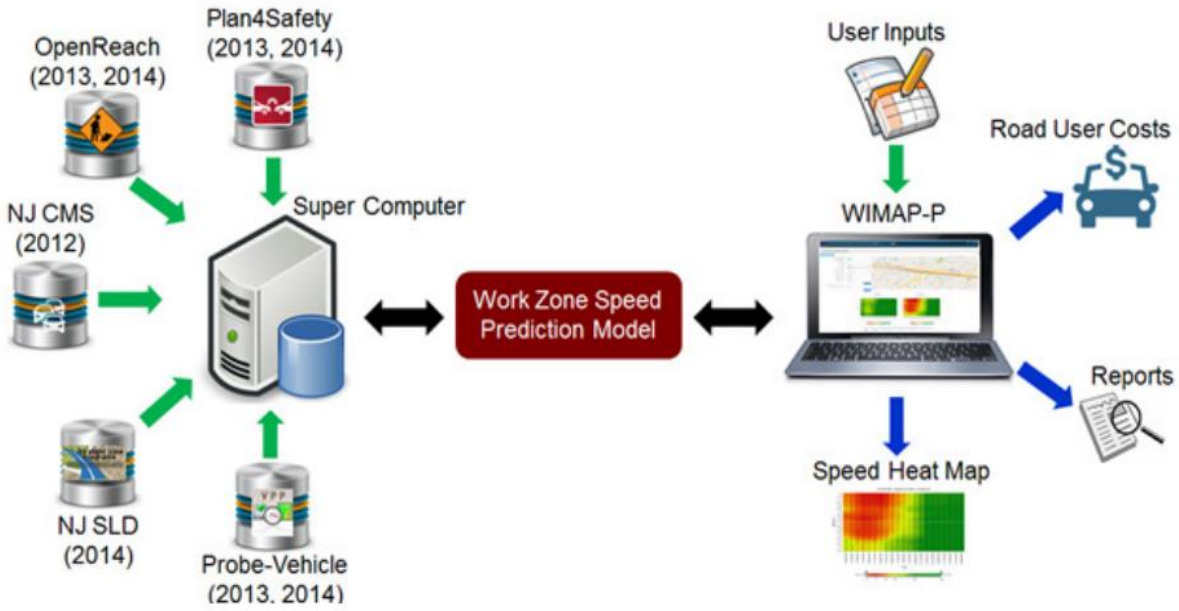


Figure 2.3 System framework for WIMAP-P.
Source: (Chien et al., 2016)

Table 2.3 summarizes the commonly used tools by various sponsored agencies to predict traffic speed due to work zone lane closures.

Table 2.3 The Inputs, output and Modeling approaches of Various Work Zone Congestion Prediction Tools

Tool	Inputs	Outputs	Modeling Approach
FlagSim	Time and location of work zone	• Traffic volume • Queue length • Delay	Parametric
Web-based Work Zone Traffic	Time and location of work zone	• Delay cost • Queue length	Parametric
Lane Closure Decision Support System (LCDSS)	Time and location of work zone.	• Queue length	Parametric
WIMAP-P	Time, location of work zone, and values of time.	• Delay cost • Queue length • Predicted traffic speed	Non-parametric
RILCA	Time and location of work zone only for the Garden State Parkway and New Jersey Turnpike.	• Queue length • Delay	Parametric

2.4 Deep Learning

The structure of ANN varies depending on each type of problem. Deep learning is a type of ANN with two or more hidden layers (Weston et al., 2012). Recent study developed a new deep machine learning approaches for predicting crash severities (Yang et al., 2018). Other studies use deep machine learning for predicting the number of Uber pickups (Wang et al., 2018). All these studies indicate that deep machine learning models produce better results than typical artificial networks.

Elisseeff & Paugam-Moisy, (1997) recommends using a number of neurons at each hidden layer that are twice the number of neurons in the previous layer. Moreover, the increase of the number of neurons and the number of hidden layers contributes to the overfitting problem in ANN (Moody, 1992).

To train a neural network, a loss function is defined to calculate the difference between the model and the actual results. Based on the difference value, sets of weights in the neural network are calculated. The optimizer updates the calculated sets of weights with every training epoch. A simple optimizer is the gradient descent method (Bottou, 2012). However, one of the problems with the gradient descent is being slow to achieve the optimal solution or never achieve the optimal solution (i.e., vanishing gradient descent) (Hanin, 2018). Reducing the number of training epochs contributes to the mitigation of the overfitting problem (Panchal et al., 2011). There are several optimizers that improve the accuracy of the traditional stochastic gradient descent functionality: Adagrad (Duchi et al., 2011), Adam (Kingma & Ba, 2014), Adadelta (Zeiler, 2012), and RMSProp (Mukkamala and Hein, 2017). The best optimizer yields the most accurate results.

While deep learning models can be formulated in various ways, simple structures may yield low accuracies. On the other hand, more complex configurations may not be suitable for smaller sample size during the training phase. Therefore, the CNN and Deep ANN structures may be promising based on the sample size in the database. Other complex structures (e.g., RNN, Long-Short Term Memory) are deemed to perform better when more data is available (Shabarek et al., 2020).

ANN suffer from overfitting problems. Figure 2.3 illustrates the difference between overfitting, underfitting, and good fitting problems. Figure 2.3 (a) shows a hypothetical data values (y) over (x). Figure 2.4 (b) demonstrates an underfitting model that does not fit the model well. On the other hand, Figure 2.3 (c) shows an overfitting model that recognizes less error but is not able to capture the relationship. Therefore, when an input is provided into the model for prediction, the model would show higher error than the trained data. Finally, a model that recognizes higher error than the overfitted problem, but represents an acceptable model is shown in Figure 3.2 (d).

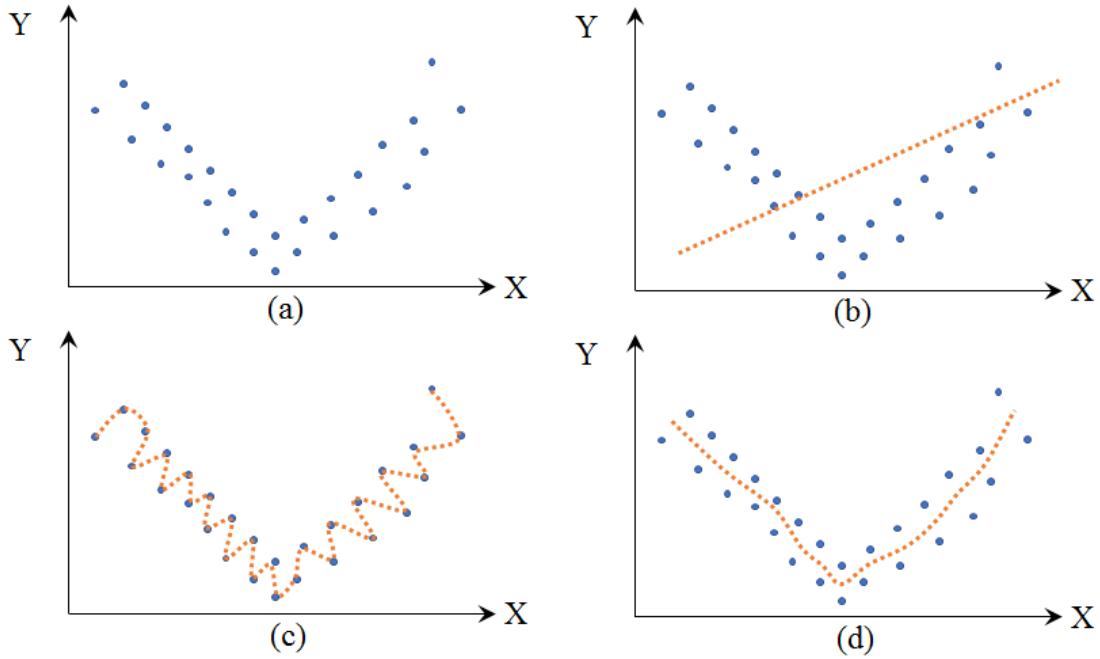


Figure 2.4 Overfitting, underfitting, and good fitting demonstration.

Dropout is a regularization technique that is applied in hidden layers for the purpose of reducing the overfitting problem (Lambert et al., 2018). Figure 2.4 shows a neuron network with dropout and without dropout. More mathematical demonstration is shown in Section 3.2.

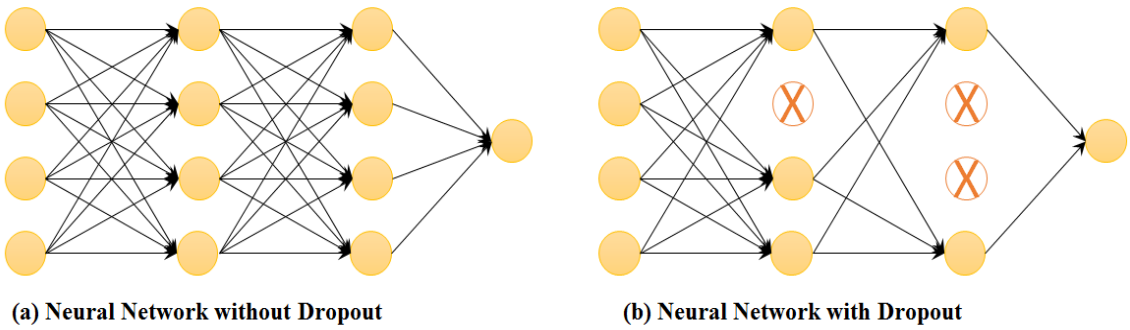


Figure 2.5 ANN architecture with and without dropout regularization.

Deep Machine learning approaches has improved over the recent years in transportation applications. Ma et al. (2015) uses deep machine learning to predict short-term speed based on microwave sensors. The prediction model is compared to other Neural Network models and shows less prediction errors. Hou & Edara, (2018) developed a Convolutional Neural Network (CNN) for predicting traffic speed in a network scale. The research indicate that CNN makes more accurate prediction on a network scale than other deep machine learning approaches. Pu et al. (2018) suggests a dropout regularization to overcome the limitation of overfitting in Artificial Neural Networks for predicting the decision on vertical gradient in railway systems. However, previous studies do not only use actual work zone speed on both mainline and connected freeways using deep machine learning and do not apply measurements for overfitting reduction.

This study extends from the existing body of literature in the following ways. First, the study uses only actual work zone information in prediction models. Second, this research aims on applying deep machine learning approaches to predict work zone speed not only on the mainline but also on the connected freeways. Third, this study mitigates

the effect of overfitting problems in ANN by applying dropout regularization in the Deep ANN.

2.5 Error Measurement Indexes

The evaluation of deep learning models is used in this research to choose the optimal deep learning model. There are several common evaluation indexes (e.g., Mean Absolute Error (MAE), and Root Mean Absolute Error (RMSE), and Mean Absolute Percentage Error (MAPE)). MAE and MAPE are used to identify the absolute error as it is shown in Equations (2.4) and (2.5) respectively, in which y_{ij} represents the observed value, $\hat{y}_{ij}$ the predicted value, n is the sample size.

$$MAE = \frac{1}{n} * \sum_{i=1}^{m} \sum_{j=1}^{n} |y_{ij} - \hat{y}_{ij}| \tag{2.4}$$

$$MAPE = \frac{1}{n} * \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{|y_{ij} - \hat{y}_{ij}|}{|\hat{y}_{ij}|} * 100 \tag{2.5}$$

On the other hand, RMSE is used to identify the actual error as it is shown in Equation (2.6).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - \hat{y}_{ij})^2}{m * n}} \tag{2.6}$$

$\hat{y}_{ij}$: Predicted work zone speed for segment j at time i

y_{ij}: Actual work zone speed for segment j at time i

n: The number of TMC segments upstream work zone

m: The number of time intervals upstream work zone

While using MAE and MAPE, provides a constant weight for all errors, RMSE penalizes the errors as they deviate from the mean and therefore is more restrictive for model evaluation (Chai, & Draxler, 2014). RMSE is preferred to evaluated speed prediction models (McKeen et al., 2005; Savage et al., 2013) when the data primarily contain less congested situations, whereas MAE and MAPE can be used when the congestion dominates traffic speed data (Kim et al., 2018). For instance, a traffic speed error of 10 mph is more critical from transportation point of view, when the error occurs in the speed bin below 30 mph. The 10-mph error might be less critical for speed bin that is higher than 60 mph. As a result, for model development that primarily contains non-congested situations, RMSE would be preferable. In work zone model developments. RMSE is recommended as work zone data does not primarily contain traffic speed with congestion (Du et al., 2017; Yu et al., 2016).

2.6 Summary

This chapter discussed the literature review of the models used in predicting work zone congestion using parametric, simulation, and non-parametric approaches. Based on the literature the non-parametric approaches do not assume a distribution of the data when training the model. Thus, the non-parametric approaches provide more accurate results compared to simulation and parametric approaches, when there is enough data to be used for training. Most of the studies have investigated the applicability of the non-parametric approaches in predicting traffic speed due to work zone congestion but did not investigate the effect of work zone on the connected freeways. With the predicted speed being extended to cover both the upstream mainline segments in addition to the upstream connected freeways, transportation agencies can predict user delay on both the mainline

and the connected freeways. Thus, a developed mitigation plan that includes any congestion spillback on other freeways can be developed.

Based on the literature review, as the number of variables increase, the structure of the ANN becomes deeper, increasing the overfitting problem. Therefore, the CNN structure, along with the dropout, can reduce the effect of the overfitting. Previous studies did not consider the problem of overfitting when predicting traffic speed under work zone conditions. Reducing the overfitting problem of traditional ANN models will improve the accuracy of the result. In this study, two main non-parametric approaches are developed and compared: Deep ANN model and CNN models. An understanding of the functionality of Deep ANN and the parameters need to be optimized is required for the model development and evaluation. Moreover, the CNN structures build on the optimal structure of Deep ANN; thus, it requires the optimal Deep ANN structure for the CNN development and evaluation.

CHAPTER 3
METHODOLOGY

This chapter explains the general structure of Deep ANN and CNN. Additionally, this chapter explains the integration of dropout as an overfitting mitigation method.

3.1 Deep ANN Model

This section discusses the general structure of Deep Artificial Neural Network (Deep ANN). Deep ANN can be used to predict traffic speeds on the roadways with work zones. Deep ANN has high fidelity for more sophisticated models that have more inputs compared to ANN models.

The general structure of the model uses, in its first step, the back propagation for Deep ANN development. Back propagation is a training algorithm that includes two steps. First, feed forward is applied through the connection of the network. Second, the error that is calculated at the propagated stage back. Deep ANN has a more complex structure than ANN in which the number of layers exceeds two layers. ANN uses kernel functions in the learning algorithm (Vapnik, 2013). However, Deep ANN use more complex learning algorithms that are able to achieve a lower minimum error compared to kernel machine functions (Schmidhuber, 2015).

The number of hidden layers and the number of neurons at each hidden layer is determined through analysis, by finding the minimum value of Root Mean Square Error (RMSE).

Dropout regularization creates a new neural network that is thinned from the actual network (Srivastava et al., 2014). To understand the concept beyond dropout, assume an ANN. The index of hidden layers and neurons are l and q, respectively. Given the number

of neurons Q_l at a hidden layer l, $y_{q'}^{l+1}$, the output of neuron q' on a hidden layer $l+1$ is shown in Equation 3.1 in which y_q^l is the output of neuron q on a hidden layer l, w_q^l and b_q^l are the weight and bias of each neuron q on a hidden layer l, respectively.

$$y_{q'}^{l+1} = \sum_{q=1}^{Q_l} \left(w_q^l * y_q^l + b_q^l \right) \tag{3.1}$$

When dropout regulation is applied at an ANN, a vector of independent Bernoulli random variables is created for each layer l. Figure 3.1 demonstrates the difference between standard neural network and a neural network with drop out regularization. Each element in the created vector has a weight probability to be multiplied by either 0 or 1. Equations (3.2) explains how each element r_q^l for layer l at neuron q is assigned to values of 0 or 1, in which ρ_q is the probability of assigning a value of zero to the element r_q^l (i.e., dropout ratio).

$$r_q^l = \begin{cases} 1 \; for \; 1 - \rho_q \\ 0 \; for \; \rho_q \end{cases} \tag{3.2}$$

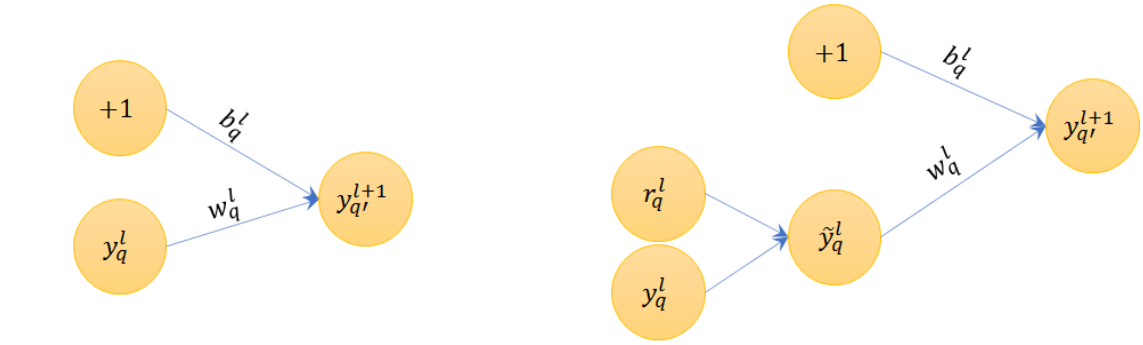
As formulated in Equation (3.3), a thinned ANN output $\tilde{y}_q^l$ is the product of r_q^l and y_q^l. A neuron q on layer l is dropped out if r_q^l is equal to zero. Otherwise, it will stay in the ANN.

$$\tilde{y}_q^l = r_q^l * y_q^l \tag{3.3}$$

Equation (3.4) illustrates the layer output under dropout regularization. Dropout ratio should not exceed the 0.5 ratio and should be minimal in the first hidden layer

(Lambert et al., 2018). In this study, dropout regularization is applied at ratio of 0.25 for the first hidden layer and 0.5 ratio for all other hidden layers.

$$y_{q'}^{l+1} = \sum_{q=1}^{Q_l} \left(w_q^l * \tilde{y}_q^l + b_q^l \right) \tag{3.4}$$



(a) Neural Network without Drop out **(b) Neural Network with Drop out**

Figure 3.1 ANN structure with and without drop out regularization.

Overfitting imposes an issue in deep learning. The dropout helps mitigating the effect of overfitting by randomly deleting some weights, so the network does not remember the old path, increasing the chance of deleting the overfitted coefficients. Figure 3.2 illustrates a general Deep ANN structure.

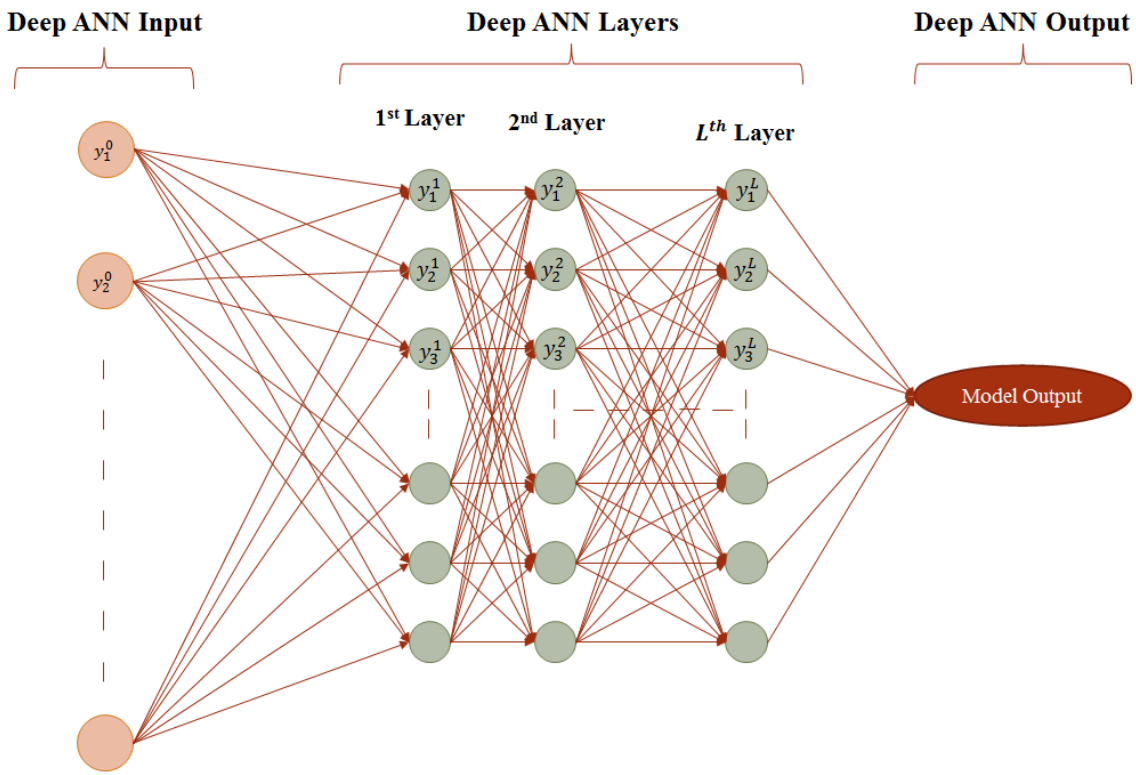


Figure 3.2 General structure of Deep ANN.

The Deep ANN can still have over fitting problems that reduces the accuracy of the testing results. Therefore, a more comprehensive model algorithm is required to mitigate the over fitting problem. A CNN model would be optimal in reducing the overfitting issue by using filters. The filters select the important features of a layers through applying filters. The next section explains the CNN model.

3.2 CNN Model

$f \otimes g_l$ is a convolution function. The convolutional function $\otimes$ is a shape function between filter f at a location τ and a hidden layer g_l. The filter f_l is a matrix that convolutes with l over stride size z, as formulated in Equation (3.5), in which Z is the total number of strides and τ is the index of the filter's location in the hidden layer.

$$f_l \otimes g_l = \int_0^z f(\tau) l(z-\tau) d(\tau) \quad (3.5)$$

The Convolution function identifies the important features in the network by extracting a reduced sized matrix from neuron layers. The output of convolutional layer will be the input of the next neuron layer. The model inputs and output have not changed from Deep ANN model. Since Convolutional layers are built on neuron layers, the number of layers and the number of neurons at each layer are determined through the Deep ANN model. A convolutional layer is applied at the first neuron layer matrix that has the dimensions of model input and number of neurons in the first hidden layer. A filter that has the width of neuron layer matrix and a height of h convolutes over the neuron layer with strides z. The output of the convolutional layer has the width of the model input. However, the height of the matrix would be the number of neurons at a layer l divided by strides z. More illustration is provided in Chapter 4. Figure 3.3 demonstrates a general unconfigured CNN network.

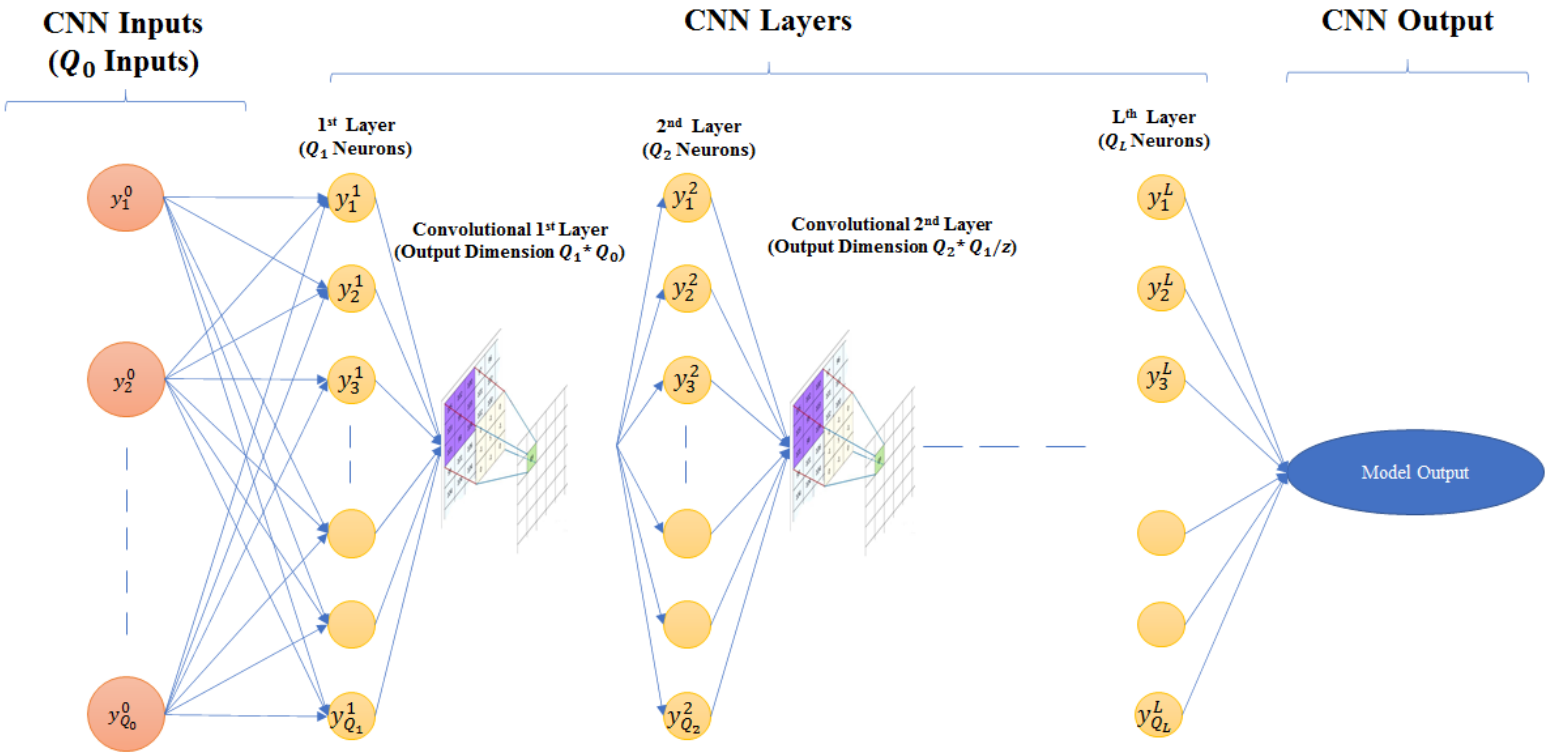


Figure 3.3 General structure of CNN.

The CNN structure is able to extract important features from hidden layers by using filters that are applied at each layer. CNN filters have heights and strides that needs to be optimized depending on the type of problem. Figure 3.4 shows how a filter, denoted in yellow shadowing, is applied on a matrix of hidden layer. The output of the convolutional layer would be a reduced sized matrix that mitigates the effect of overfitting.

Figure 3.4 The functionality of the CNN on a hidden layer.

3.3 Summary

This study shows the general structure of the Deep ANN model and the CNN model. These models require data, which is discussed in detail in Chapter 4. The structure of the CNN model uses the structure of the Deep ANN in addition to the convolution function applied at each hidden layer. Dropout is explained as method to overcome the overfitting problem in the deep learning models.

CHAPTER 4
MODEL DEVELOPMENT

Roadway work zones include shoulder and/or lane closures. These closures lead to increase in travel time of vehicles traveling upstream work zone. To mitigate the delays associated with the roadway users, developing a traffic speed prediction model is recommended for public agencies. A prediction model is needed to capture the travel time reduction due to work zones not only on upstream mainline segments, but also on the upstream connected freeways. The prediction model is required to predict work zone impact over space and time upstream work zone area. Therefore, this chapter explains the database development that is required for model inputs. After that, this chapter introduces the model formation that is used for predicting work zone speed. The framework of the model's development and evaluation is provided in Figure 4.1.

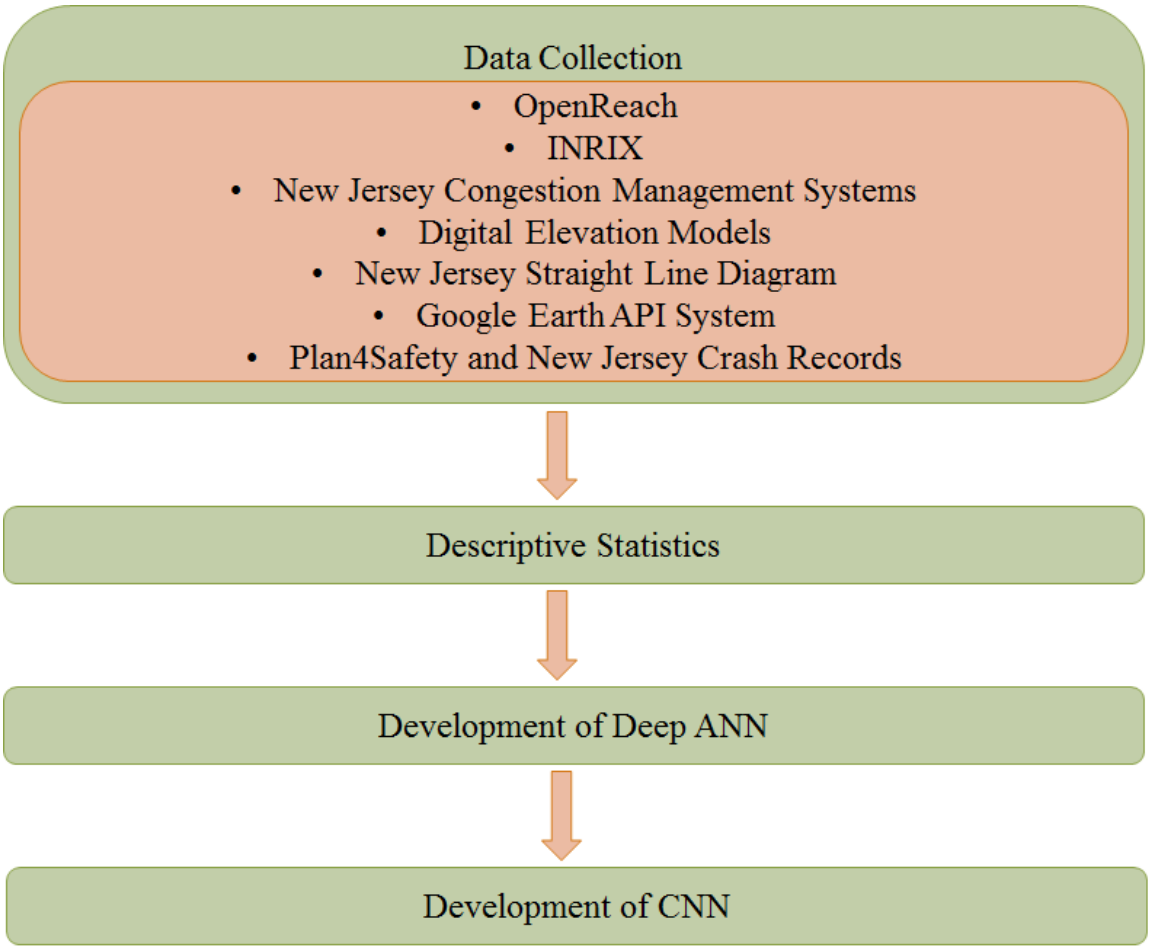


Figure 4.1 The framework of the model development and evaluations.

4.1 Data Collection

This section presents the database sources and the procedure in which databases is fused. Previous studies indicate that an accurate prediction model for work zone effect would require a significant amount of big data (Edara and Cottrell, 2007; Du et al., 2017). In work zone congestion prediction problem, it is required to obtain information about work zone location and timing, road geometry, traffic volume, traffic speed, and incident occurrence. Thus, the developed databases are categorized and categorized into the followings:

- Work zone data: Work zone data includes information about work zones such as work zone location, time, starting milepost, ending milepost, number of closed lanes.

- Road geometry data: Road geometry data includes data regarding road type, number of lanes, ramp connection to mainline, connected freeways, vertical gradient.
- Traffic volume data: Traffic volume is collected for New Jersey through a big data source (New Jersey Congestion Management Systems). The analyzed data includes truck percentage, traffic volume, and traffic volume at ramps.
- Floating car data: Floating car data includes the space mean speed for freeway and ramps segments under work zone conditions and under normal conditions.
- Crash records data: Crash records data includes the location of crashes and the time crashes occurred.

The databases are combined and merged through big data analysis, to ensure homogeneity in the data inputs. The databases are used to report actual work zone conditions and the associated model inputs that are used for model development and evaluations. The needed databases for model development are illustrate d in Figure 4.2.

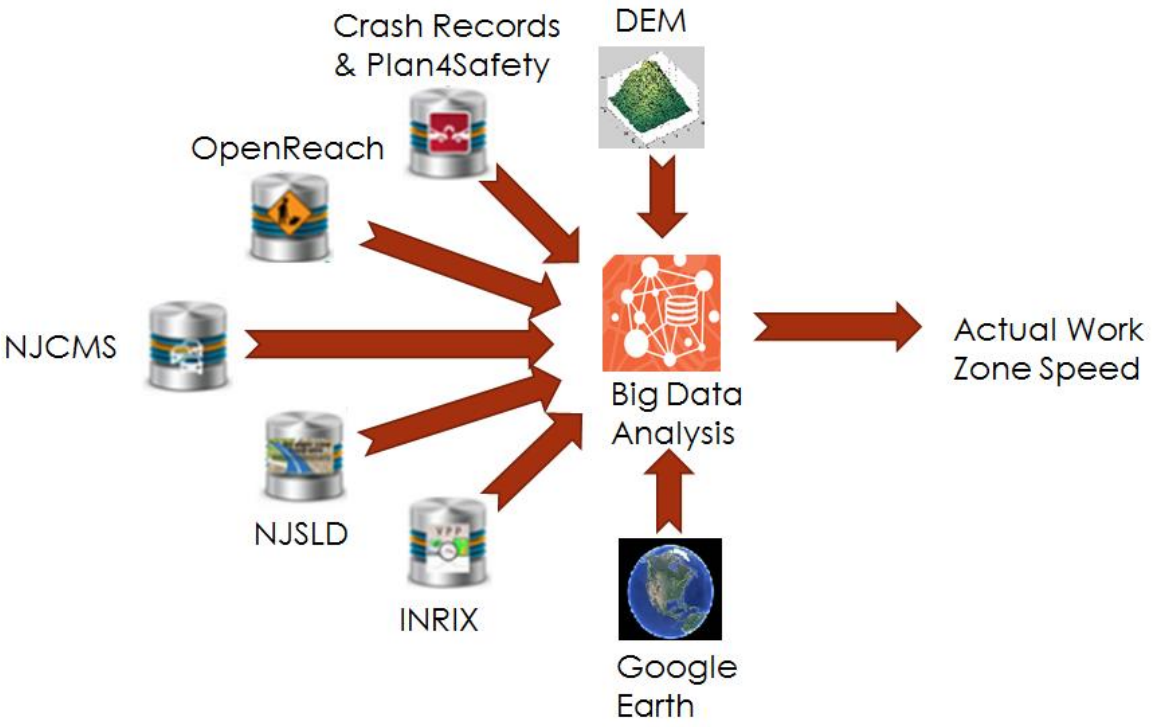


Figure 4.2 Data sources for model development.

Data are combined through big data analysis to obtain a final version of the data that includes all the information from these databases. The combined database is used in model development and evaluation.

The databases in which the data is collected from are as explained in detailed in the following sections.

4.1.1 OpenReach

OpenReach (CoVal Systems., 2016) is a dynamic event reporting system of work zones and accidents. The required work zone data from OpenReach includes work zone starting and ending time, work zone location, and number of closed lanes. OpenReach database is the result of a collaboration of 16 agencies in New Jersey, Connecticut, and New York. OpenReach database includes three main categories: work zone information, incident information, and special events information.

OpenReach database is updated, by Traffic Operations Center (TOC), on the 511NJ website to reflect any incident or work zone occurrence. Figure 4.3 demonstrates an example of work zone date, work zone time, and work zone location at a given location. The information also includes the agencies responsible for the work zone and the date they updated the information on OpenReach (511NJ, 2020). In this study, OpenReach data is analyzed for all work zones between July 2014 and September 2019.

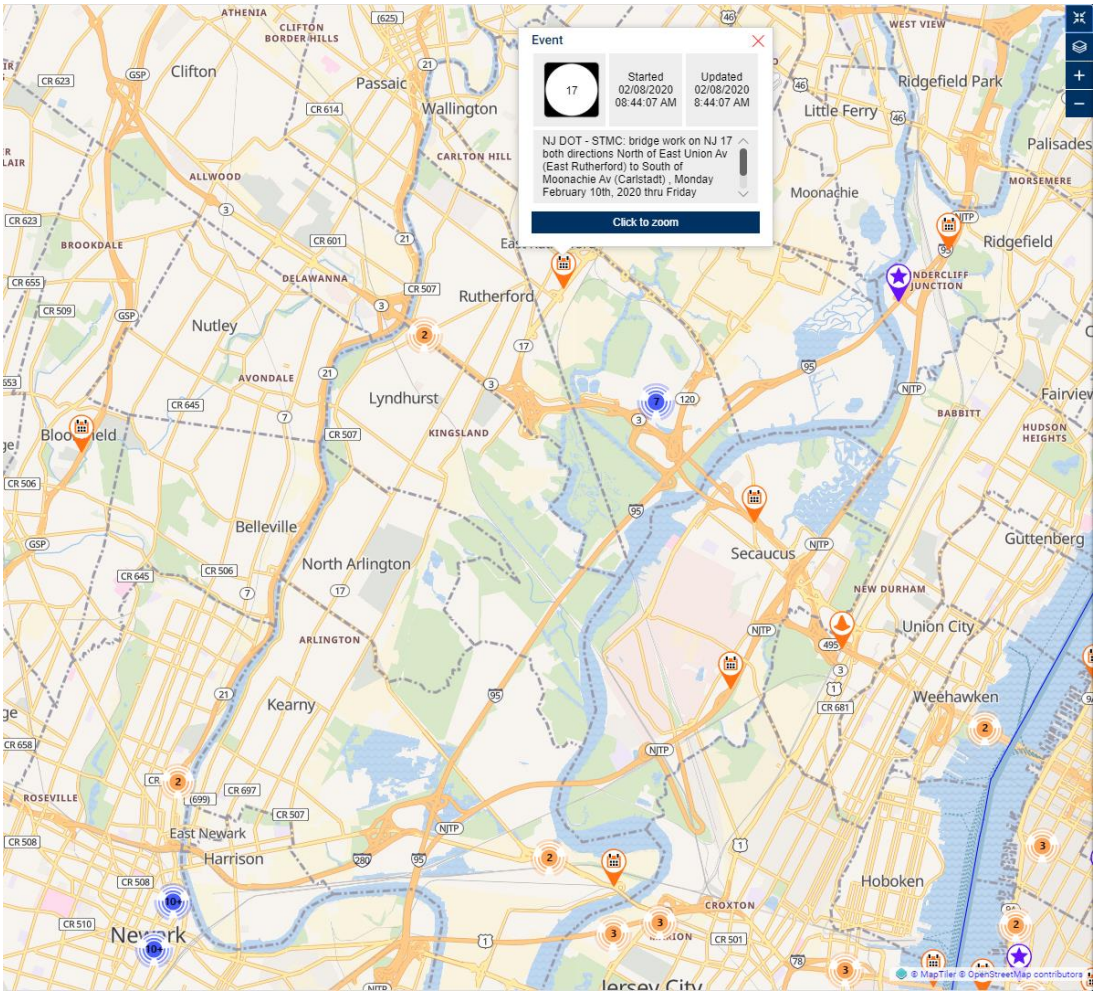


Figure 4.3 Sample of Real-time work zone data illustration through the 511NJ website.
Source: (511NJ, 2020)

4.1.2 New Jersey Straight Line Diagram (NJSLD)

NJSLD (New Jersey Department of Transportation., 2014) is a database that is developed by New Jersey Department of Transportation (NJDOT), which includes geometric information about all roadways in New Jersey. Roadways are identified in NJSLD through Standard Road Identification (SRI) system. The database includes information about roadway milepost locations, total number of lanes. roadway class, and traffic direction.

NJSLD database includes around one million segments in the State of New Jersey (NJGIN, 2020). Figure 4.4 demonstrates an example of one segment of Interstate-78 in New Jersey and the information presented from NJSLD dataset.

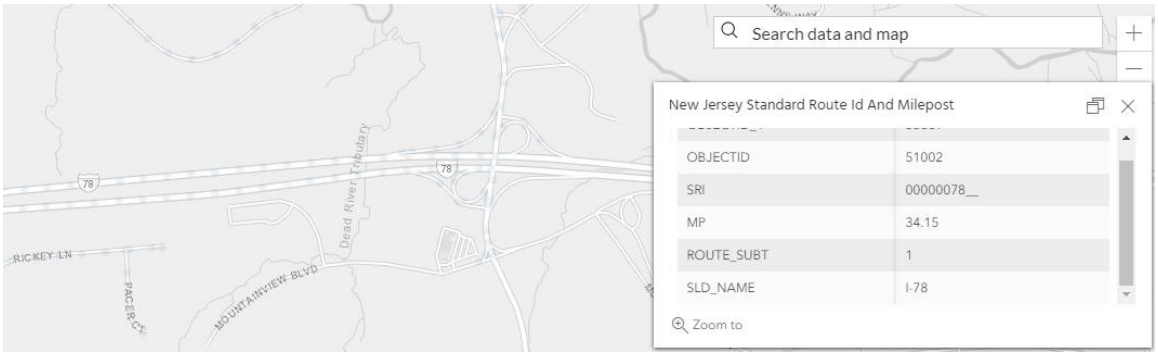


Figure 4.4 Example of NJSLD data records.
Source: (NJGIN, 2020)

4.1.3 Digital Elevation Model (DEM)

DEM is an elevation system developed by United States Geological Survey (USGS) that is used to find landscape value (i.e., the elevation of specific points) on a given terrain (United States Geological Survey, 2018). DEM data are developed using topographical data, spot heights, and a software package called ANUDEM (Wilson et al., 2000). ANUDEM provides a grid elevation map from drainage points (Hutchinson, 2011)

The DEM data is used to determine the vertical gradient of roadways, by calculating the difference in the elevation between two points and dividing the difference by the segment length. Figure 4.5 shows a sample of DEM database visualization (Satellite Imagery Corporation, 2020).

Figure 4.5 Sample of DEM elevation heat map.
Source: (Satellite Imagery Corporation, 2020)

4.1.4 Google Earth API System

Google Earth API is developed by Google corporate (Google., 2018). It provides information regarding ramp and connected freeways related to freeway segments. Therefore, it is used to develop a database that includes mainline network and the associated ramp and connected freeway segments. The Google Earth API system can provide information about the longitude and latitude of the connection points between the ramps and the freeways. However, a data analysis is required to combine these coordinates with the Milepost system. Thus, other databases can be merged into the developed database.

4.1.5 New Jersey Congestion Management System (NJCMS)

NJCMS (New Jersey Department of Transportation., 2015) provides information regarding traffic volume and truck percentage in the state of New Jersey. NJCMS is a system software that provides NJDOT with various performance measurements (e.g., level of service,

volumes to capacity ratios, delays, and travel speed)The database is used to provide work zone capacity, which is an input in the deep machine learning model.

The NJCMS covers an overall of 7,129 miles of roadway segments in the State of New Jersey. This study includes all freeway segments in New Jersey; therefore, the developed model uses 1,562 miles of NJCMS data, which are distributed over 1,227 segments. One major issue when using NJCMS with other data sources is that NJCMS uses the Milepost coordinates in identifying the segments. However, some other databases use the global coordinate systems, making the matching between the two systems important for data analysis.

4.1.6 INRIX Database

INRIX data provides space mean speed data. The probe-vehicle data used in the model is reported from INRIX speed database (INRIX., 2019). INRIX identify segments through Traffic Message Channel (TMC). There are more than 1700 freeway TMCs and more than 600 ramp TMCs in the state of New Jersey. The collected data includes 4 billion records of freeway segments and 1.3 billion records of ramp segments. The data duration is from July 2014 to July 2018.

INRIX data is provided for all the interstate segments, the Turnpike and the Garden State Parkway in New Jersey. Figure 4.6 demonstrates an INRIX coverage of the studied segments in New Jersey. For the development of the model, the data is aggregated on 15-minute intervals for each roadway segment.

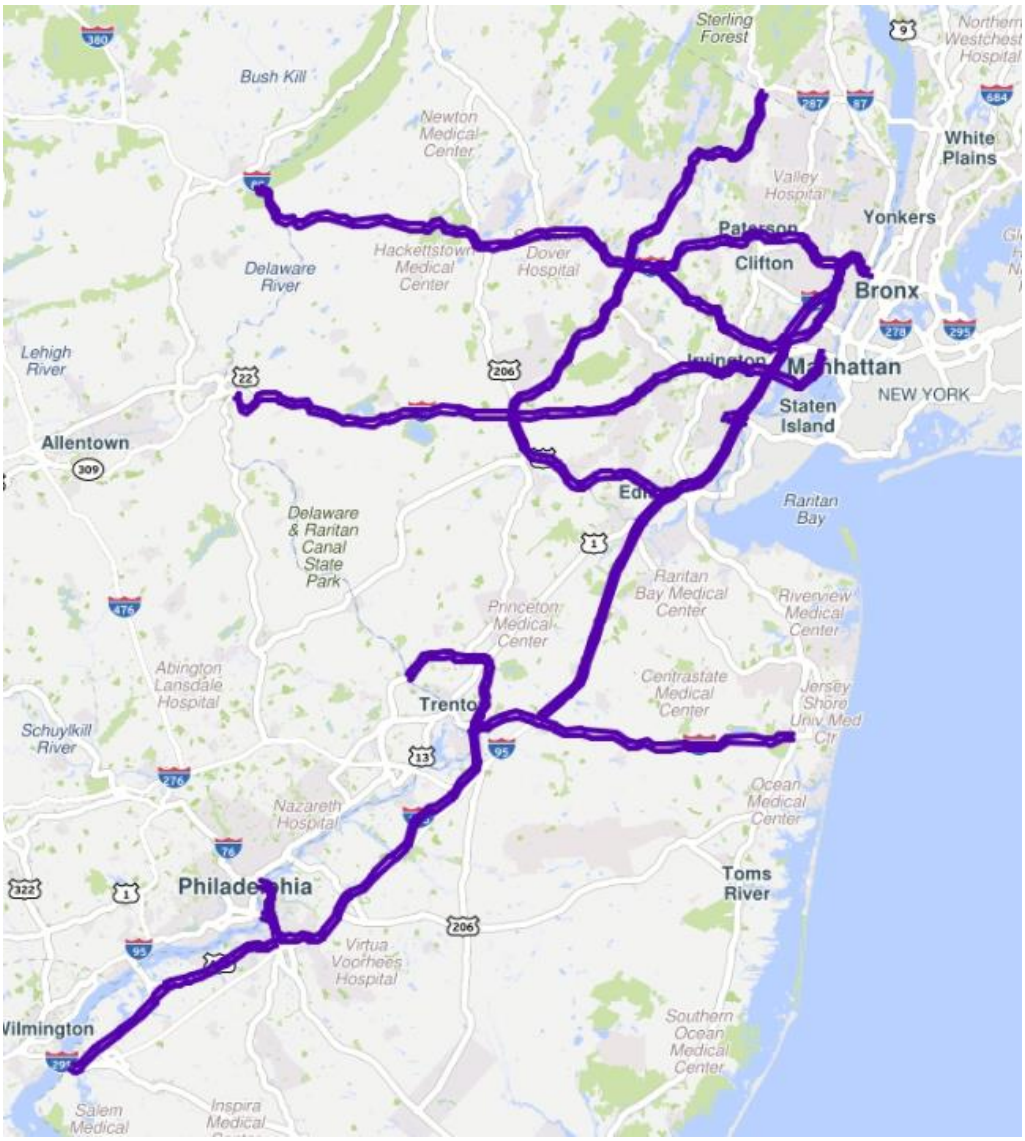


Figure 4.6 INRIX coverage of Interstate roadways in New Jersey.
Source: (INRIX 2019)

4.1.7 Plan4Safety and New Jersey Crash Records

Plan4Safety and New Jersey Crash Records report crash accidents time and location Transportation Safety Research Center. (2016); New Jersey Department of Transportation. (2018). New Jersey Crash Records data was developed from 511NJ website. Accidents were reported from Plan4Safety, but Plan4Safety was stopped in 2016. Hence, the New Jersey Crash Records was substituted during the years 2017 and 2018. Figure 4.7 shows a

sample of the 511NJ website showing real-time incidents including a description of the accident, the exact location, and the time the incident is reported (511NJ, 2020).

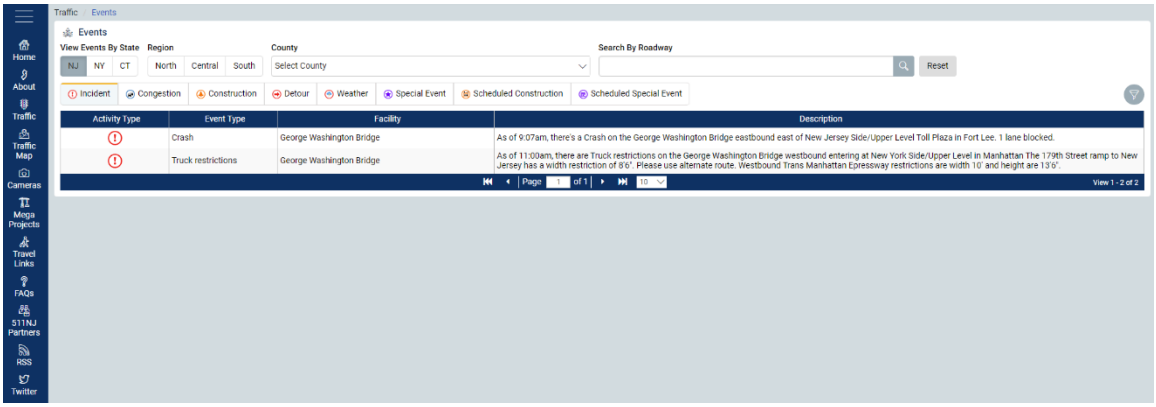


Figure 4.7 Sample of real-time incident of 511NJ website.
Source: (511NJ, 2020)

The purpose of these two databases is to exclude work zones with accidents records, downstream and upstream work zone, as the purpose of the model is to predict work zone congestion without any additional accident congestion.

4.2 Descriptive Statistics

To develop a model that is able to identify work zone congestion, work zone information is collected between 2014 and 2019. Work zone information that are useful for predicting upstream mainline, ramps, and upstream segments are listed in section 2.1.1 in the literature review. Among collected work zones, work zones with accident records are excluded using Plan4Safety database and Crash Records as it is illustrated in section 3.1. In New Jersey, in the periods between 2014 and 2019, there is around 40,000 work zones. Only around 5,500 work zone have occurred on the interstate roadways. The selected work zone include only work zones that have a duration less than 24 hours, and with full information (i.e., verified location, and time). Consequently, after screening the 5,500 work zone, there is

822 work zones with complete information. The analysis is conducted on 822 work zones with complete information in New Jersey between 2014 and 2019. The complete information includes the exact work zone location and work zone time, excluding work zones with accidents 10-miles on the upstream and downstream segments during the work zone duration. Work zones vary in terms of the number of lane closures and the available lanes for traffic. Table 4.1 demonstrates the distribution of the 822 work zones in terms of lane closure type.

Table 4.1 The Selected Number of Work Zones for Model Development.

Lane Configuration at Each Direction	Shoulder Closure	One Lane Closure	Two Lane Closure	Total
Two Lane Freeway	23	119	NA	142
Three Lane Freeway	122	394	21	537
Four Lane Freeway	20	106	17	143
Total	165	619	38	822

Table 4.2 provides the descriptive statistics of TMC data by routes. The number of freeways TMC segments is 1,733, extending over 1,561.5 miles of freeway roads in NJ, in which the total route length is 1,561.5. The average TMC length varies between 0.31 and 1.22 with an average standard deviation of 1.35.

Table 2.2 Descriptive Statistics of Freeways TMC Data by Route

Route Name	**Number of Work Zones**	**Avg. TMC Length (miles)**	**SD of TMC Length (miles)**	**Route Length (miles)**	**Number of TMCs**
Garden State Parkway	205	0.89	1.14	393	441
I-76	2	0.31	0.22	9.4	30
I-78	113	1.02	2.98	198.8	195
I-80	187	0.78	1.17	153.4	197
I-195	22	1.14	1.23	70.8	62
I-278	6	0.52	0.25	4.6	9
I-280	32	0.4	0.37	33.4	84
I-287	59	0.84	1.15	144.6	173
I-295	153	0.75	0.83	135.3	180
I-676	1	0.36	0.21	9.4	26
New Jersey Turnpike	42	1.22	1.63	408.8	336
Total	822	0.9	1.35	1,561.5	1,733

4.3 Deep ANN Development

The proposed model uses probe vehicle data to capture traffic speed on the network. Probe-vehicle database provides one input and the output of the model depending on the work zone time, and the location. The output is the speed during work zone conditions for segment (i) at time interval (j). The input of the model is the average monthly speed during normal conditions for the same day for segment (i) at time interval (j). Traffic volume data in work zones are typically provided through vehicle counting. However, the scarce availability of the traffic counts during work zone conditions would not be feasible to be included in the model. Therefore, an available vehicle counts through a big data source, New Jersey Congestion Management Systems (NJCMS), is used in the model. The model assumes that traffic volume and truck percentage are given through historical data based on Average Annual Daily Traffic (AADT) from NJCMS. Moreover, the model calculates

work zone capacity using Highway Capacity Manual, HCM (2010), approach as it is shown in Equation 4.1.

$$C_w = (1600 + I) * f_{HV} * N_0 - R \quad (4.1)$$

where C_w: Work zone Capacity (vph);

f_{HV}: Heavy vehicle adjustment factor explained in (HCM);

I: The adjustment factor for type and intensity of work activity (vphpl)

N_0: The number of open lanes within a work zone; and

R: HCM manual adjustment for on-ramps (vph).

The Deep ANN considers eight inputs to predict speed with work zone conditions for segment (i) at time (j): Traffic volume approaching work zone at time j, Traffic speed during normal conditions, Traffic volume on the mainline downstream interchange on at time (j), Vertical gradient of segment (i), Work zone capacity, Distance of segment (i) to work zone, traffic volume of segment (i) at time (j). As it is illustrated in the literature review, traffic volume approaching work zone at time (j) in addition to work zone capacity is correlated to speed reduction upstream work zone. Some of model inputs are retrieved from the datasets directly, others are obtained from other sources (e.g., HCM formula for work zone capacity). The increase of traffic volume approaching a work zone increases the congestion upstream work zone whereas the decrease in work zone capacity is attributed to the increase of congestion upstream work zone. The model classifies upstream segments into three types: upstream mainline segments, upstream ramp segments, upstream

connected freeway segments as it is shown in Figure 4.8. The model includes prediction of the work zone impact up to 10-miles upstream work zone segment.

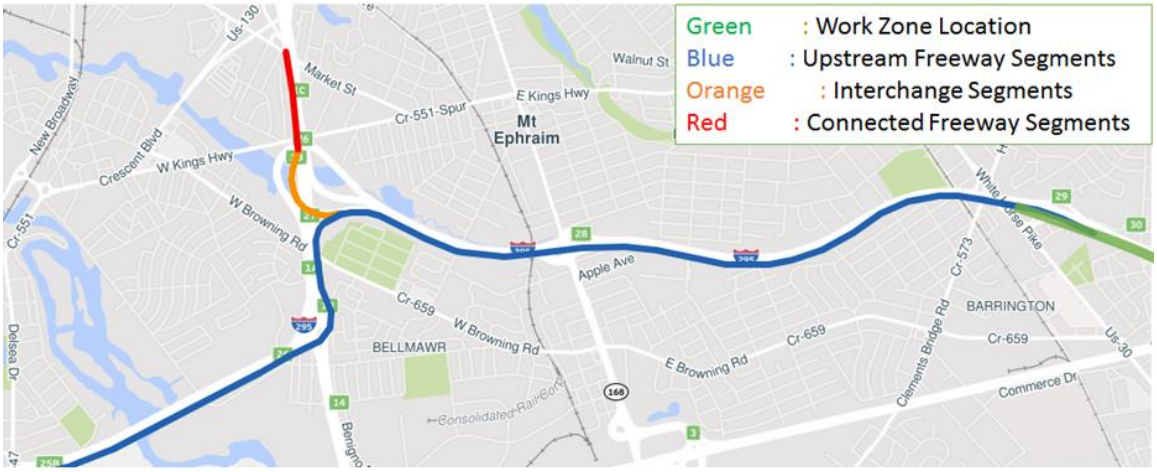


Figure 4.8 Work zone on Interstate-295and adjacent road network.

The model includes factors affecting ramp and connected freeways only in order to predict when work zone congestion hits upstream ramps and connected freeways. Figure 3.2 demonstrates the general input-output of the suggested model. The model inputs are identified to affect traffic speed during work zone conditions throughout Chapter 2.

The optimized Deep ANN structure is determined by its performance which yields the least RMSE. The RMSE is calculated based on the TMC segmentations of the INRIX data.

To find the optimized structure, a set of scenarios are set based on number of hidden layers and number of neurons with each hidden layer, which are illustrated in Table 4.2. In this study, 5 layers are chosen as the maximum number of layers because increasing the number of layers results in overfitting problems. Therefore, throughout the analysis, six layers would yield higher errors, setting a trend of overfitting pattern in the data.

The results in Table 4.5 indicate that a Deep ANN with 4 layers that have 128 neurons in the first layer, 256 neurons in the second one, 512 neurons in third one, and 1024 neurons in the fourth one recognizes 5.9 mph RMSE value. The optimizer used in these structures is Adam, as it is recognized as a superior optimizer in the literature. However, to investigate its effectiveness in the suggested model, three other Deep ANN optimizers are analyzed on the optimal Deep ANN structure. A grid search analysis is conducted to find the optimal optimizer. To find the optimal structure, a grid search analysis is used. The number of layers in the grid search analysis is within the range of 3 and 5 and the number of neurons in the first hidden layer is one of the following: 128, 256, or 512. The following layers have a number of neurons twice the number of neurons in the previous layer. The grid search is conducted with a variety of optimizers. Based on grid search analysis, Table 4.3 shows the RMSE results of using RMSprop, Adagrad, Adadelta, and Adam as optimizers on the suggested Deep ANN structure. From the analysis, Adadelta and Adagrad has similar performance, but since Adam has the least RMSE, Adam optimizer is selected for model development and evaluation. The optimal structure represents the number of neurons at each hidden layer.

Table 4.3 RMSE of the optimal structure with various Optimizers.

Optimizer Name	Optimal Structure	RMSE (mph)
RMSprop	256/512/1024/2048	6.4
Adadelta	128/256/512	6.3
Adagrad	256/512/1024/2048	6.0
Adam	128/256/512/1024	5.8

The RMSE for the Deep ANN model is 5.8 mile per hour. However, to get more insight of the applicability of the model regarding each lane-closure type, a sample of each lane-closure type is selected for model development. Table 4.4 indicates the RMSE value

for the selected sample size of each lane-closure type. The results show that one-lane closure in general yields the minimum RMSE indicating that the predicted model have more accurate results for this type of lane closures. Three different number of hidden layers structure are analyzed: 3 layers, 4 layers, and 5 layers. Structure 1, 2, 3, 4 have 16, 32, 64, 128 neurons in the first hidden layer respectively. The number of neurons in the next hidden layers is twice the number of neurons of the previous hidden layer, as it is suggested by (Elisseeff & Paugam-Moisy, 1997). Testing results are proceeded on 15% of work zone database Table 4.4 shows testing sample size for each lane-closure type. 52% of the total 822 work zone database has ramp spillback at connected freeways.

Table 4.4 Testing Sample Size and the number of TMC links

Testing Sample Size (Number of TMC Links)	Shoulder Closure	One Lane Closure	Two Lane Closure	Total
Two Lane Freeway per each Direction	3 (96)	18 (446)	NA	21 (542)
Three Lane Freeway per each Direction	18 (754)	59 (1730)	3 (66)	80 (2,550)
Four Lane Freeway per each Direction	3 (111)	16 (522)	2 (62)	21 (695)
Total	24 (961)	93 (2,698)	5 (128)	122 (3,787)

The results, in Table 4.5, indicate that a deep Artificial Neural Network with 4 layers that have 128, 256, 512, 1024 neurons in the first, second, third and fourth layers respectively is the optimal structure of Deep ANN, and it recognizes 5.8 mph RMSE value. Figure 4.9 shows the structure of the Deep ANN model. The activation function that is used in the training is Rectified Linear Unit (ReLU) as it is widely

recommended to be used as a non-linear function with Deep ANN and CNN layers (Schmidt-Hieber, J., 2020). It is not computationally expensive and simply outputs the maximum value between zero and the input value.

Table 4.5 RMSEs with Different Deep ANN Structures.

Structures	Number of Neurons at each hidden layer (RMSE in mph)		
	3 Layers	4 Layers	5 Layers
Structure 1	16/32/64 (10.3)	16/32/64/128 (9.4)	16/32/64/128/256 (9.1)
Structure 2	32/64/128 (6.2)	32/64/128/256 (6.3)	32/64/128/256/512 (6.2)
Structure 3	64/128/256 (6.0)	64/128/256/512 (6.1)	64/128/256/512/1024 (6.2)
Structure 4	128/256/512 (6.1)	128/256/512/1024 (5.8)	128/256/512/1024/2048 (6.1)

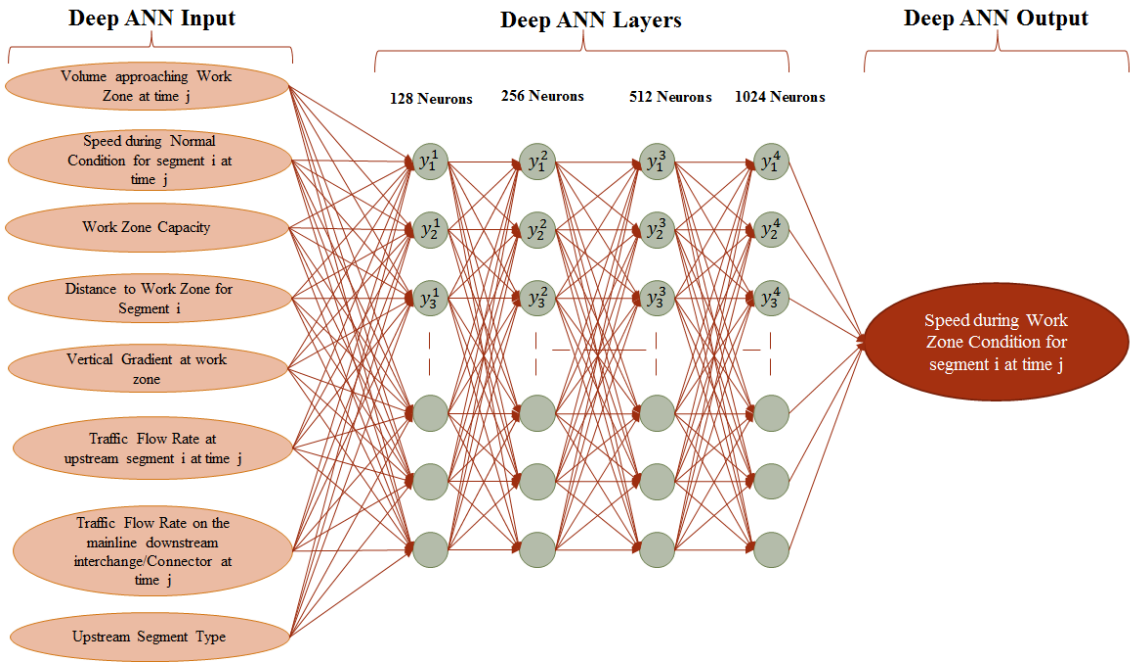


Figure 4.9 The structure of the Deep ANN model.

4.4 CNN Development

Convolutional Neural Networks are deep learning approach used for estimating model outputs based on different outputs. The convolutional function $\otimes$ is a shape function that is the product of filter f and input layer g over strides z as it is illustrated in Equation (3.8).

Convolution identifies the important features in the network by extracting a reduced sized matrix from neuron layers. The output of convolutional layer will be the input of the next neuron layer. To demonstrate how convolution occurs, the model input from Deep ANN model, consists of 7 variables and based on the optimal number of neurons in the first layer, we have 128 neurons at the first layer and 7 input variables. Since Convolutional layers are built on neuron layers, the number of layers and the number of neurons at each layer are determined through the previous Deep ANN model we conducted before. As we previously found in step 2, the optimal number of layers for our problem is 4 in which the first, second, third, and forth layers include 128, 256, 512, and 1,024 neurons respectively. A convolutional layer is applied at the first neuron layer matrix that has the dimensions of model input and number of neurons in the first hidden layer (7*128). A filter (f) that has the width of neuron layer matrix (7) and a height of (h) convolutes over the neuron layer with strides (z). The output of the convolutional layer has the width of the model input, which is 7. However, the height of the matrix would be 128 divided by strides (z). Figure 4.10 demonstrates an example of filter height of 3 and stride size of 2. Note that each hidden layer is reshaped and padded and the activation function that is used in the training is ReLU, which is suitable for a regression output (i.e., traffic speed).

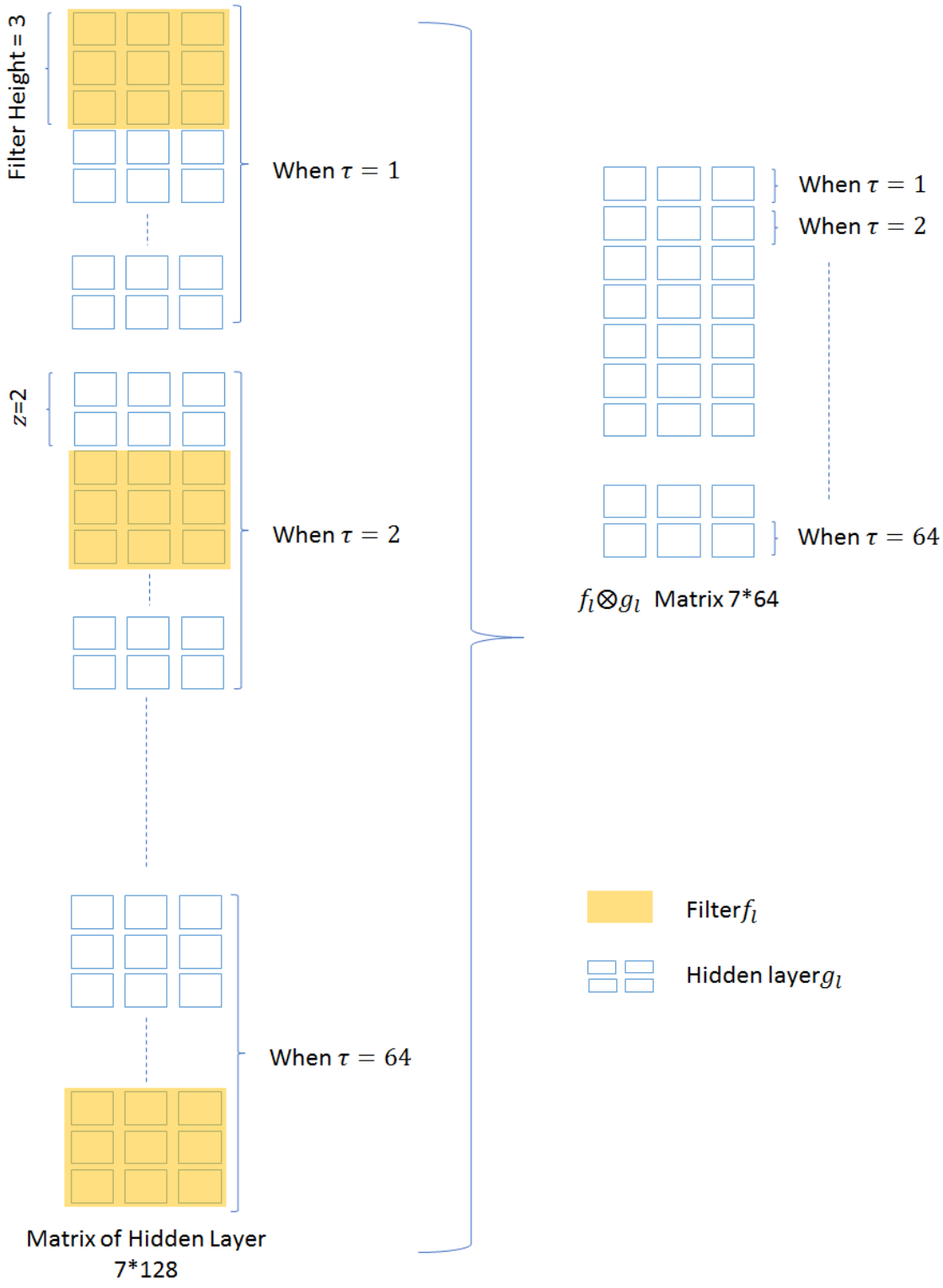


Figure 4.10 The CNN mechanism example.

The purpose of CNN model is to predict the speed during work zone for segment i at time j. The general structure of CNN model is demonstrated in Figure 4.11.

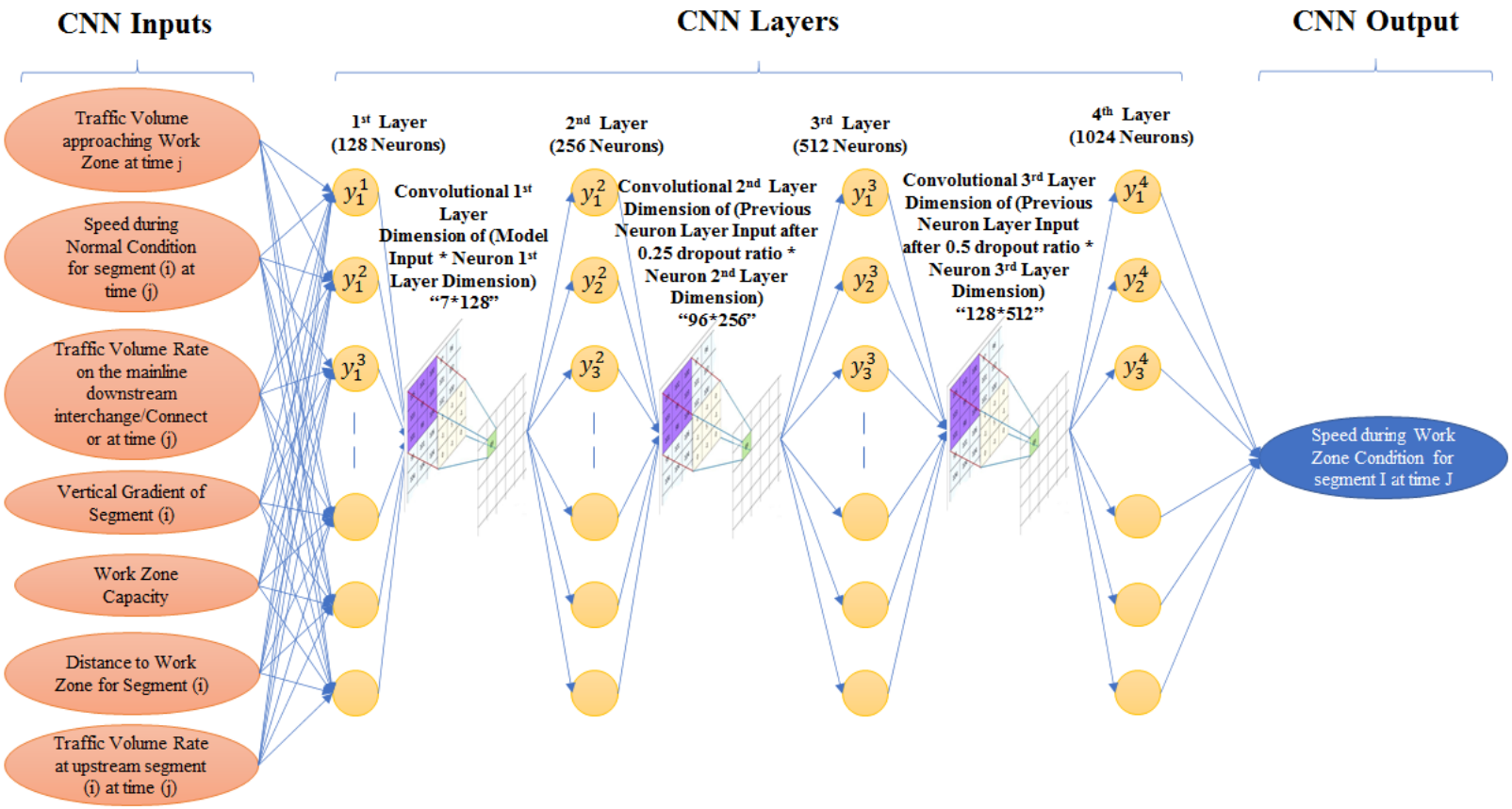


Figure 4.11 Structure of the proposed CNN model.

From the previous discussion, we have not determined the filter height and filter stride. These values are determined through analysis and based on recommended ranges specified in the literature review. Filter stride is recommended to be less or equal to filter height whereas Filter height is recommended to be less than the number of variables (Goodfellow, 2016). Table 4.6 demonstrates the testing results to determine the optimal filter height and filter stride. The testing results are obtained from 15% of available work zone database as it is indicated in Table 4.4.

Table 4.6 CNN Model Results

RMSE value "mph"	Filter Height =2	Filter Height =3	Filter Height =4
Stride Size = 1	5.6	5.6	5.6
Stride Size = 2	5.5	5.7	5.7
Stride Size = 3	-	5.6	5.7
Stride Size = 4	-	-	5.8

The results indicate that the optimal CNN structure has a filter height of 2 and stride size of 2 in which RMSE value is equal to 5.5 mile per hour. CNN model has lower RMSE values than the Deep ANN model (5.8 mile per hour), and therefore is able to predict traffic speed under work zone conditions with less error.

4.5 Summary

This research implements the models discussed in Chapter 3 for the prediction of traffic speed under work zone conditions. The data sources are discussed in detail with the coverage of each data source. Two deep learning models are developed and evaluated: Deep ANN and CNN.

The Deep ANN model requires optimization of the number of hidden layers and the number of neurons at each hidden layer. It is found that four layers with the structure 128/256/512/1024 yields the optimal solution. The optimized structure of Deep ANN is

used for the CNN model development. In the CNN model, the filter height and stride need to be optimized. It is found that a filter height of 3 and stride size of 2 yield the optimal solutions. Based on the results, the CNN model yields more accurate results compared to the Deep ANN.

CHAPTER 5
CASE STUDY

5.1 Background

The developed CNN model predicts traffic speed on the mainline segments and the connected freeways. To illustrate the applications of the developed CNN model, a work zone location that has a connected freeway segment close to the work zone location is selected. The selected location has a congestion spillback due on other freeway segments. Therefore, the location of the 1-mile work zone is Interstate-287 between Milepost 39 and Milepost 38 in which one lane closed over four lane freeway at the southbound direction. I-287 Southbound has multiple junction areas with various routes 10 miles upstream work zone (i.e., Route-10, I-80, Route 202, and Route-46). However, INRIX data only covers the TMCs in the junction area of I-80. Thus, this study only includes I-80 as a connected route.

The selected work zone duration is from 3:00 PM till 09:00 PM on 07/08/2015. This study considers both the mainline segments and the connected freeways when predicting traffic speed under work zone conditions. Figure 5.1 demonstrates the general configuration of work zone location. The green links represent the selected work zone location whereas the blue links represent upstream mainline segments, and the orange links illustrate ramp links, connecting the connected freeways to the mainline freeway. The red links are the connected freeway segments. In this case study, the connectors are Interstate-80, Westbound and I-80 Eastbound.

The importance of adding connected freeways is to account for the congestion spillback from one freeway to another. Therefore, users delay due to work zone conditions

will change. The coverage of the analysis includes 10-miles of upstream segments including both the mainline segments and the connected segments, separately. Moreover, the work zone congestion prediction span extends two-hours post the ending of work zone, to account for any residual delays from previous time steps. Consequently, is the heat maps are shown is between 2:30 PM (i.e., 30 minutes before the starting time of work zone to observe any congestion prior to the work zone starting time) and 11:00 PM. The inputs of the model include traffic speed and traffic volume over the specified period of time in addition to the number of closed lanes and the number of lanes at the upstream segments.

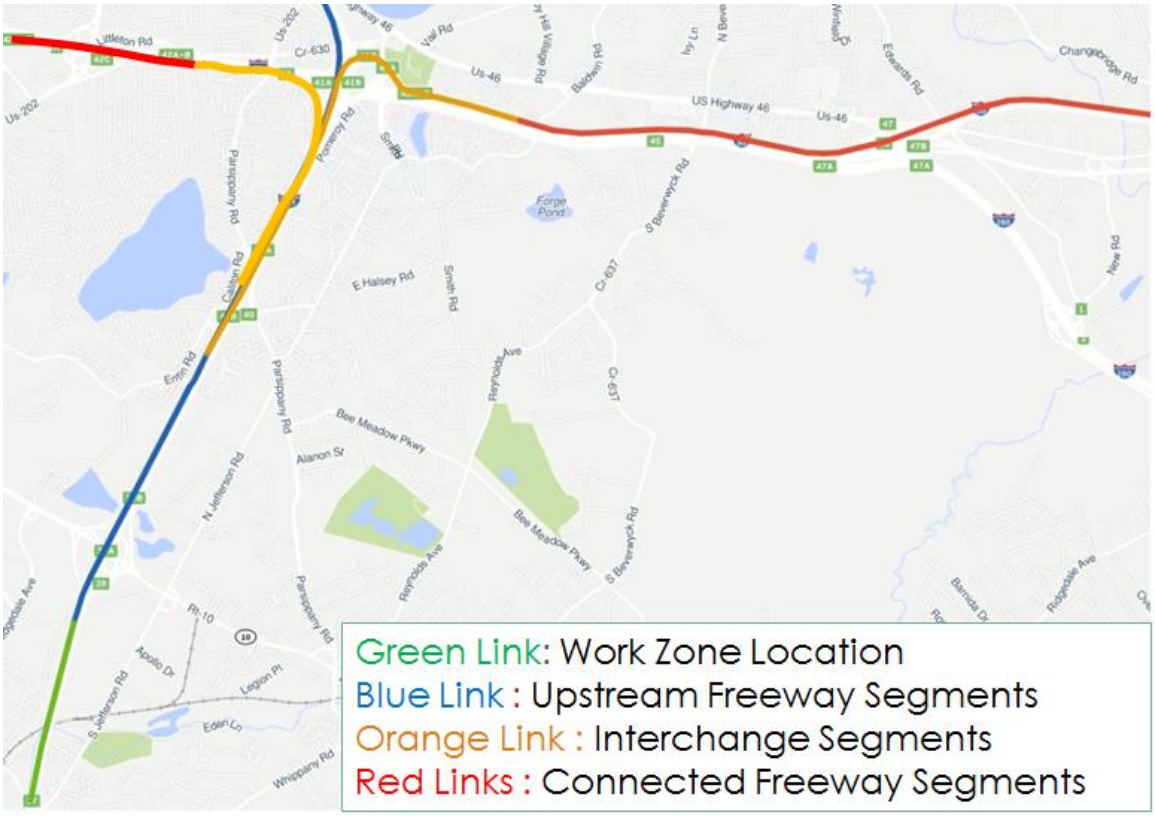


Figure 5.1 Work zone on Interstate-287 and adjacent road network.

5.2 Results

The analysis is conducted at 0.5 miles spatial intervals and 15-minute temporal intervals. The analysis uses a heat map to visualize the model results. The heat map, in its horizontal axis, represents the temporal changes (i.e., time of the day), starting from the work zone starting date and time and ending two hours post the work zone ending date and time. On the other hand, the vertical axis, in the developed heat map, shows the spatial changes (i.e., mileposts upstream work the zone). The spatial changes are illustrated at 0.5-mile intervals starting from the work zone segment link and ending 10-miles upstream the work zone.

Heat maps showing traffic volume change over time and space are demonstrated in Figure 5.2. The heat maps in Figure 5.2 show (a) the passenger car volume and (b) truck volume. The results indicate high traffic volume in heat maps representing both the passenger car and truck volumes between 16:00 and 19:00, for the 3-miles upstream work zone segments. Consequently, high traffic volumes is expected to correspond to any potential congestion during work zone conditions.

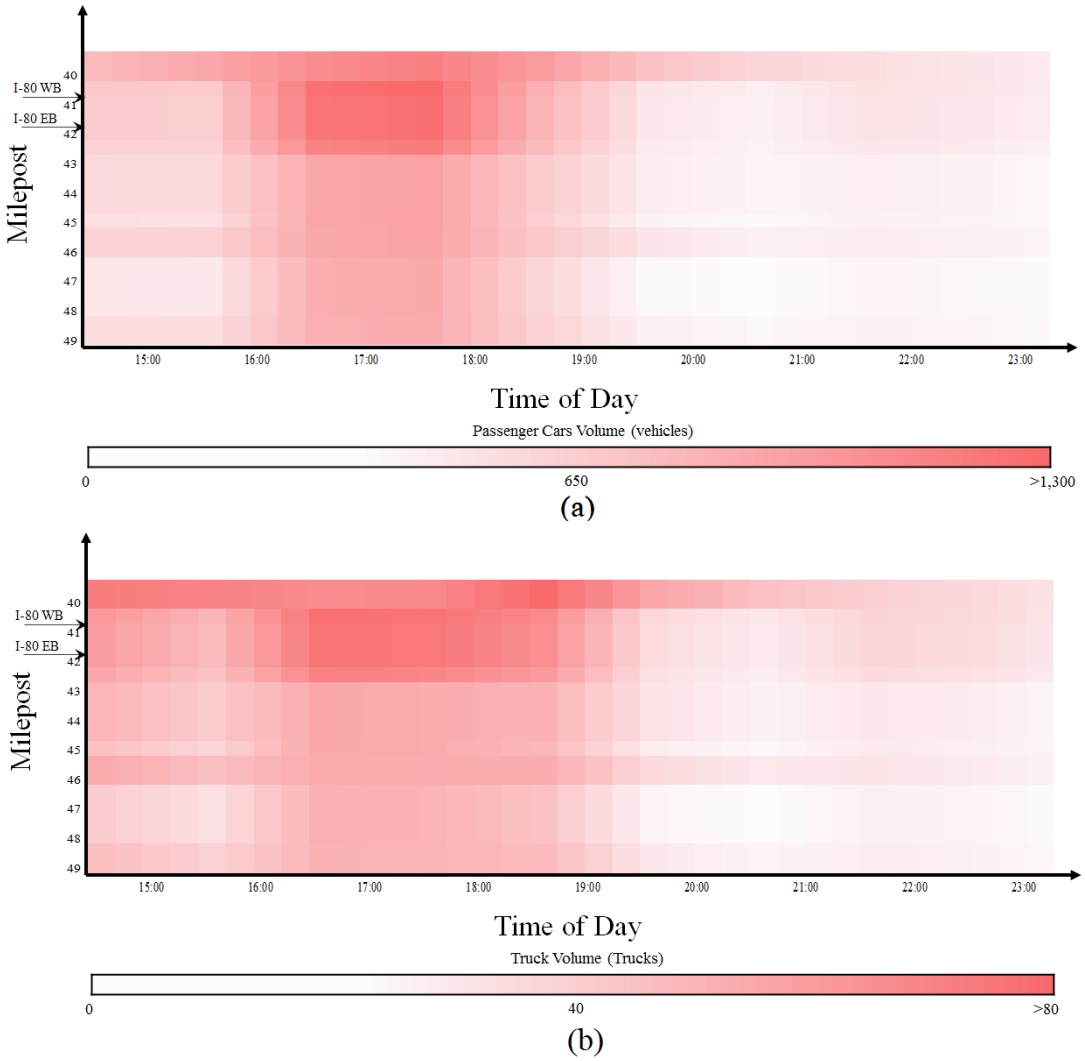


Figure 4.2 Heat map of (a) passenger cars and (b) trucks distribution for I-287 SB.
Source: New Jersey Congestion Management Systems

To show the traffic volume on the connector segments, heat maps are demonstrated in Figure 5.3. The heat maps in Figures 5.3 and 5.4 show (a) the passenger car traffic volume and (b) the truck volume on the upstream ramp and the connected freeway for I-80 Westbound and I-80 Eastbound, respectively. The heat maps show high traffic volume, especially between 15:00 and 18:30, compared to the mainline freeway segments.

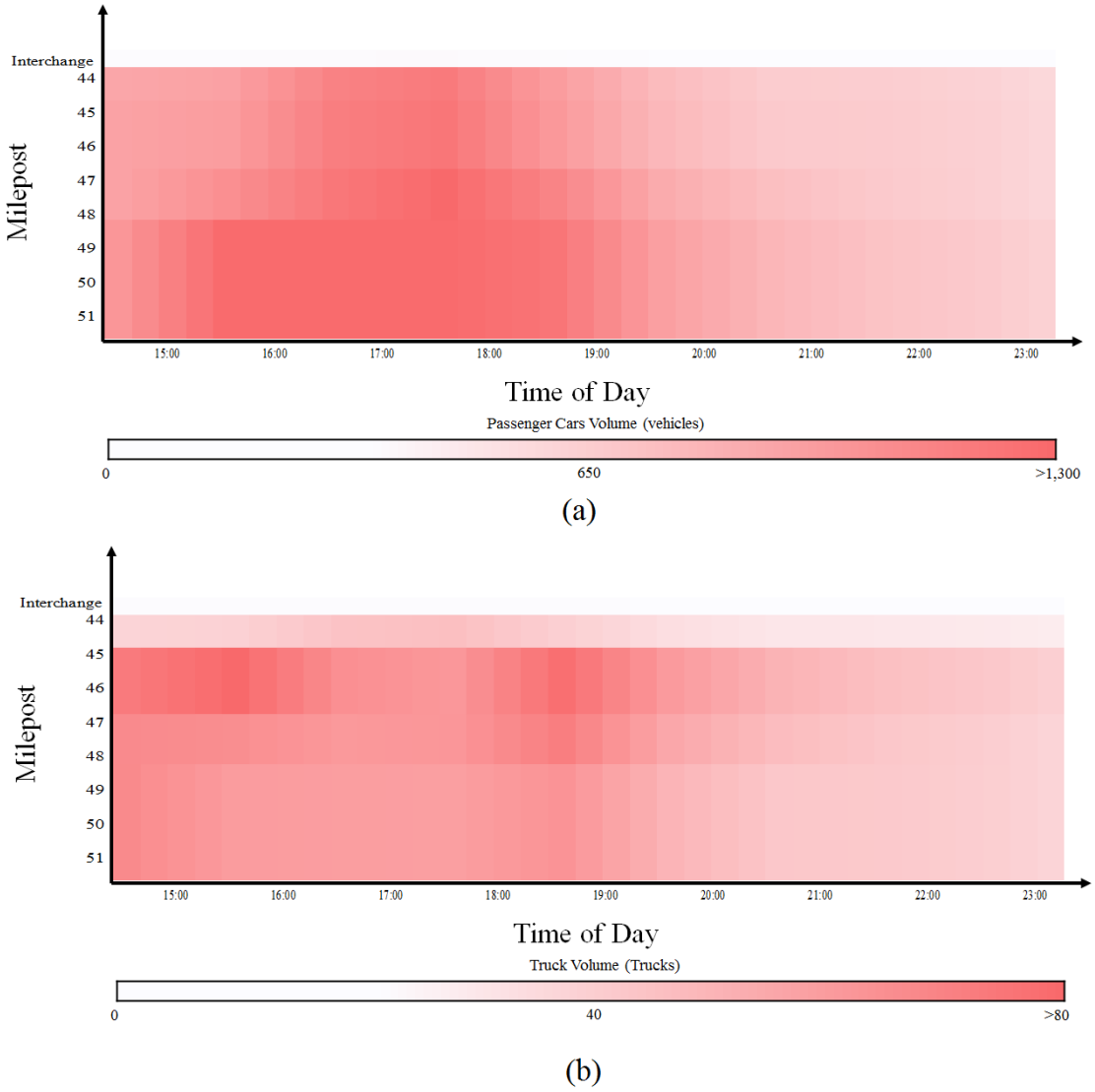


Figure 5.3 Heat map of (a) passenger cars and (b) truck volumes of I-80 Westbound.
Source: New Jersey Congestion Management Systems.

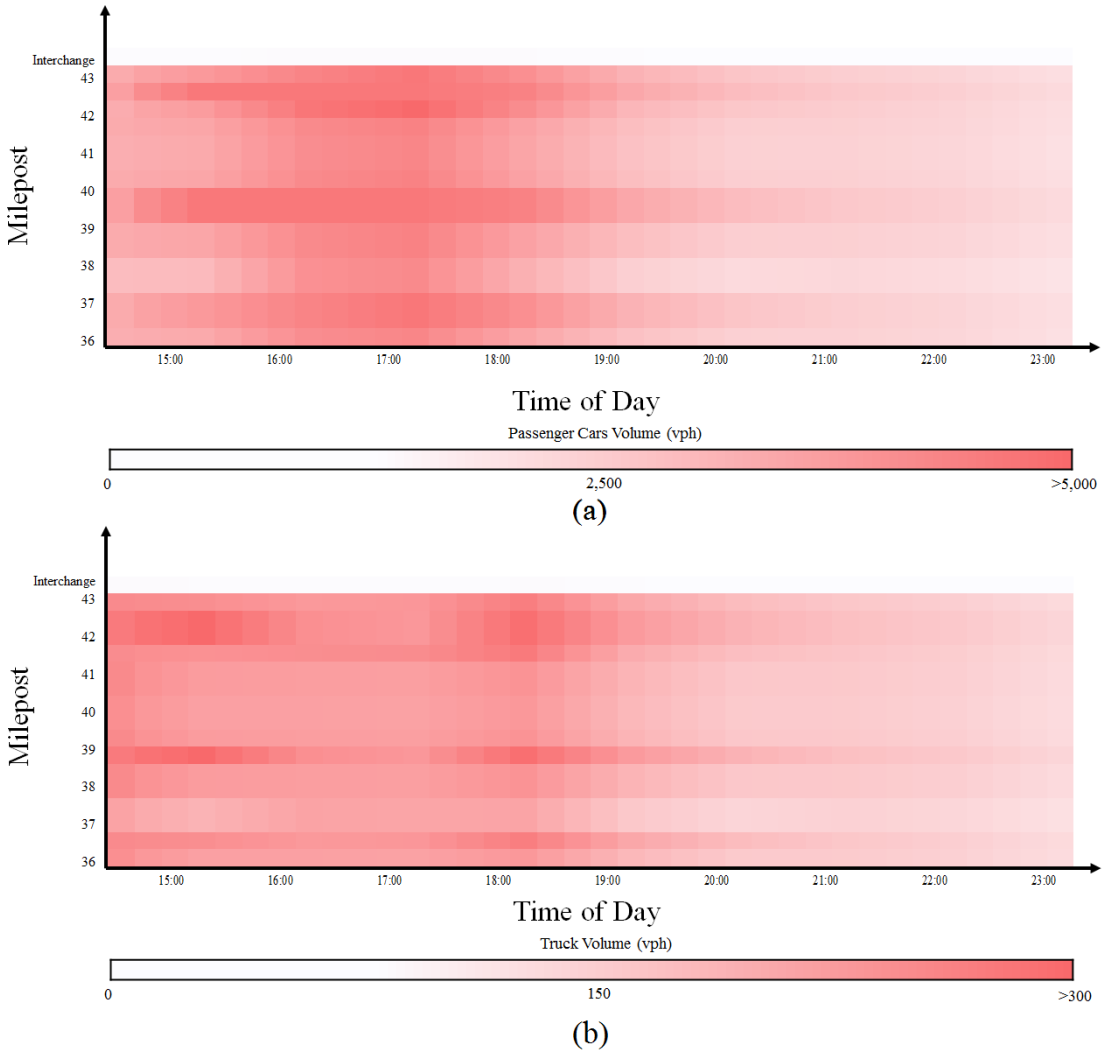


Figure 5.4 Heat map of (a) passenger cars and (b) truck volumes of I-80 Eastbound.
Source: New Jersey Congestion Management Systems.

The heat maps showing normal traffic speed are shown in Figure 5.5 for (a) I-287 SB (b) I-80 WB and (c) I-80 EB. The normal traffic speed is obtained from the average traffic speed of the same month the work zone occurred in during the same day and time, excluding the periods in which accidents occurred. These heat maps show no major congestion during non-work zone conditions. It is worth noting that the normal traffic speed on the ramp is low compared to the other freeway segments due to speed limit of the ramps.

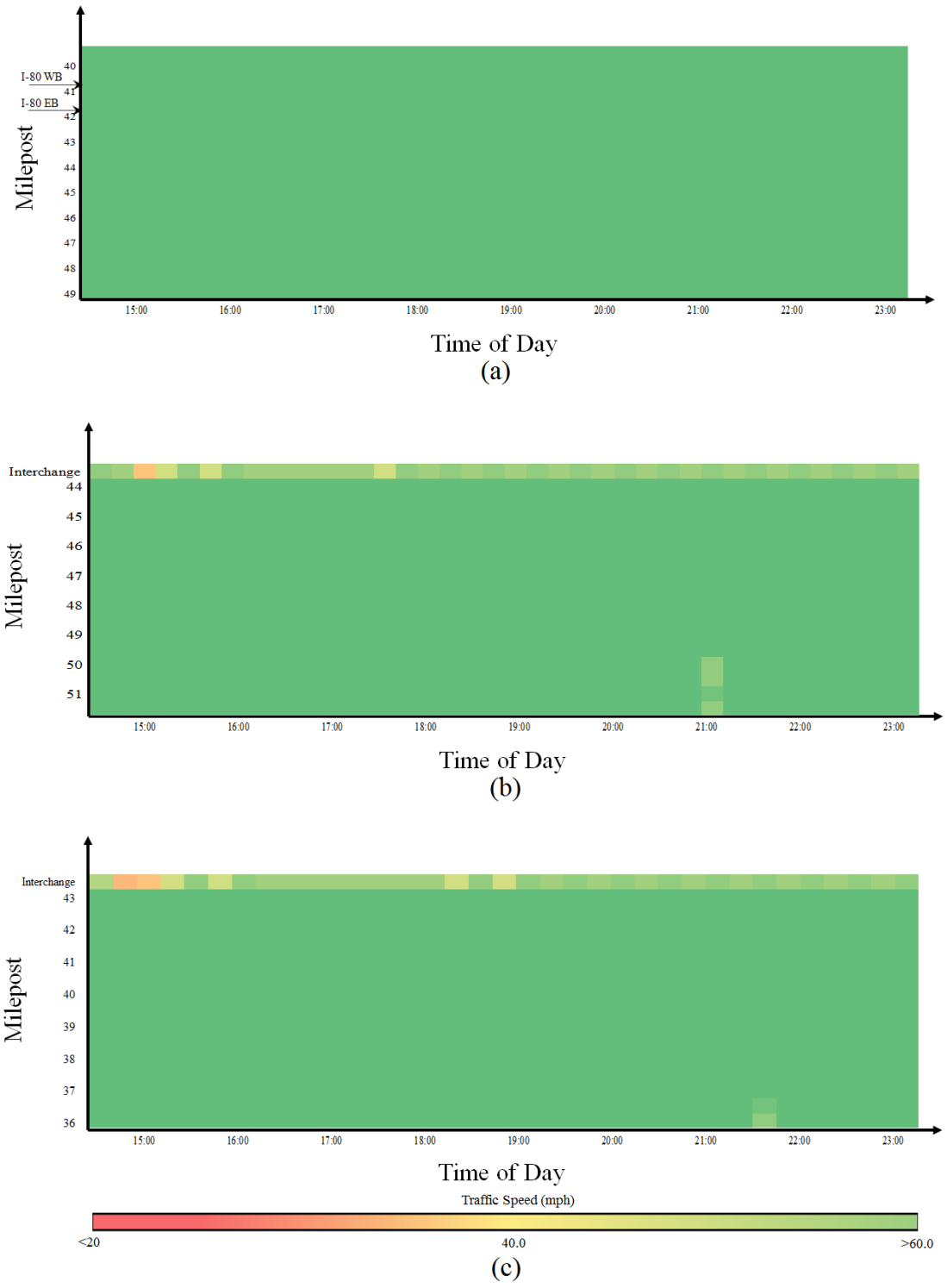


Figure 5.5 Heat map of traffic speed without work zone conditions for (a) I-287 SB (b) I-80 WB and (c) I-80 EB.

The heat map showing the traffic speed of the mainline (i.e., I-287 SB) is shown in Figure 5.5 for (a) actual traffic speed reported from INRIX (b) traffic speed predicted from the CNN model and (c) predicted speed from the model of WIMAP-P. On The results, in both the actual and the predicted traffic speed heat map results, show that there is a congestion for around 3-miles upstream the work zone. The congestion mainly occurred between 15:00 and 19:00, especially in the first three hours. Additionally, the model of WIMAP-P overestimates congestion 4-miles upstream the work zone between 16:00 and 19:00. The results indicate that the model of WIMAP-P underestimates the congestion between 15:00 and 17:00 and overestimates the congestion between 17:00 and 19:00. On another note, WIMAP-P does not provide any indication of congestion spillback to other freeways. Therefore, there is no further analysis for other connected freeway using the model of WIMAP-P.

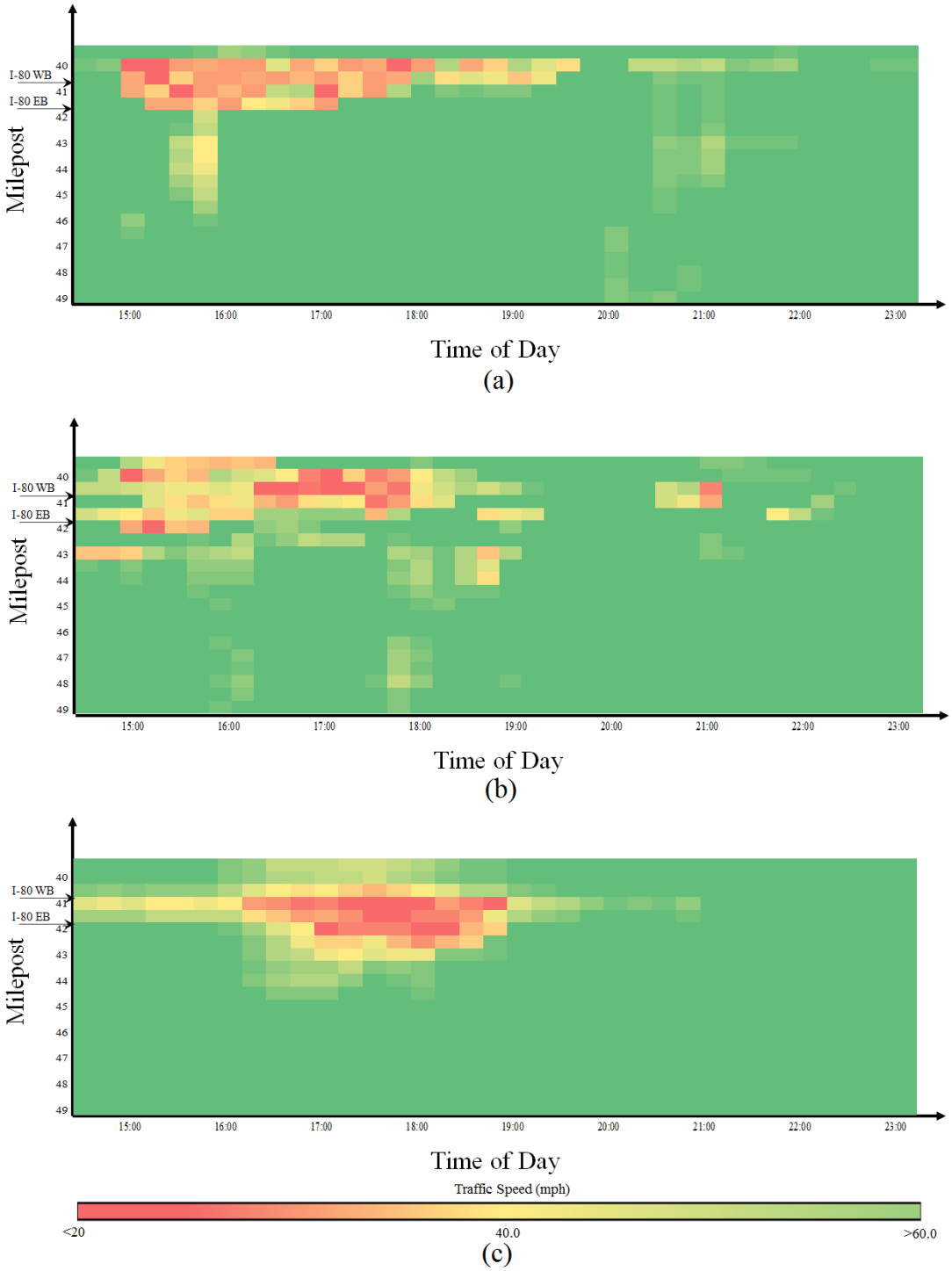


Figure 5.6 Heat map of I-287 SB of (a) Actual speed reported from INRIX (b) predicted speed from the CNN Model (c) predicted speed from the model of WIMAP-P.

This study provides a model that predicts traffic speed on connected freeways under weather conditions. The connected freeways are merged into the mainline freeway segments through ramp segments. In the case study the ramp merges into the mainline freeway on milepost 38.5 (i.e., 1.5-mile upstream work zone). The heat maps, illustrated in Figures 5.8 and 5.9, show the comparison between (a) the CNN predicted upstream connected freeway traffic speed and the (b) actual traffic speed on the same connector freeway. The actual traffic speed shows a higher congestion between 15:30 and 16:30 compared to the CNN predicted values. Hence, the CNN model underestimated the congestion during these periods.

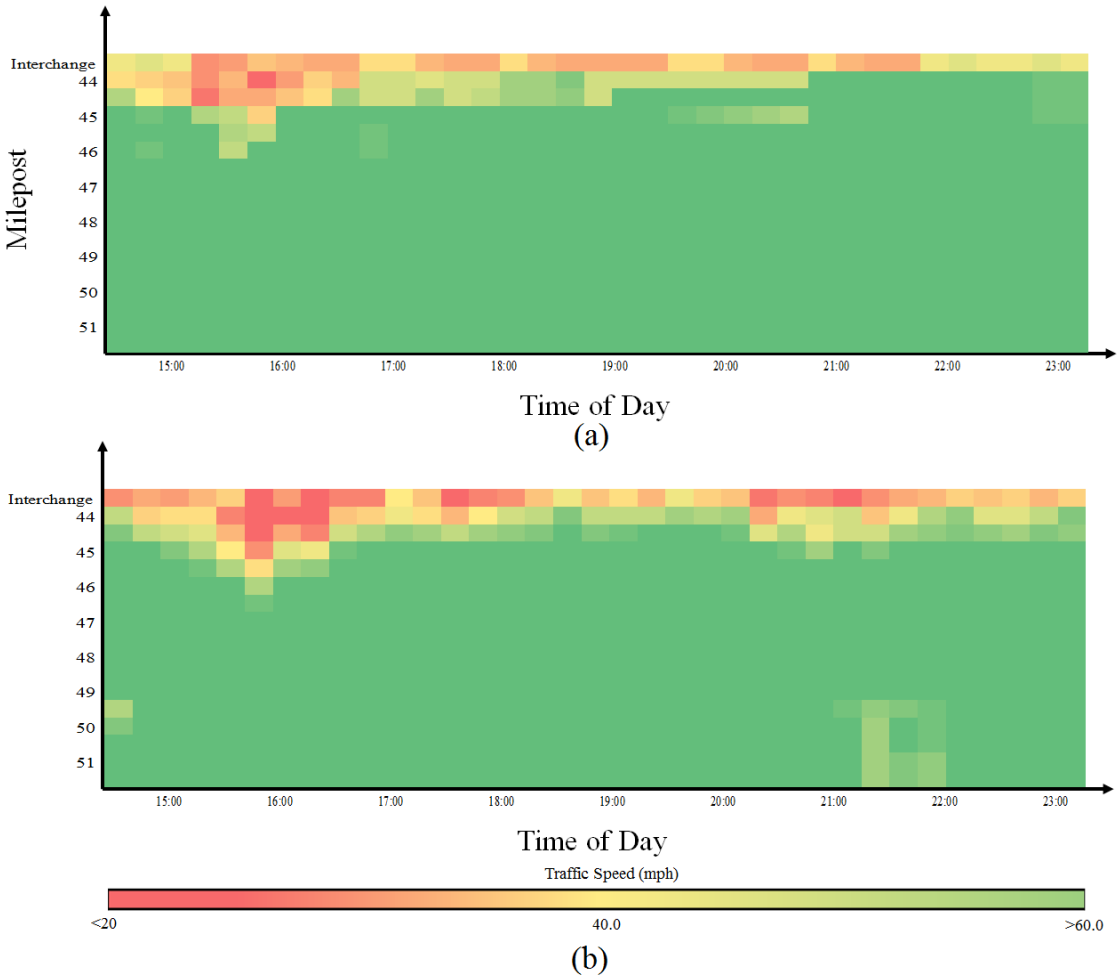


Figure 5.7 Heat map of traffic speed on I-80 WB from (a) the CNN prediction model (b) the actual traffic speed reported from INRIX.

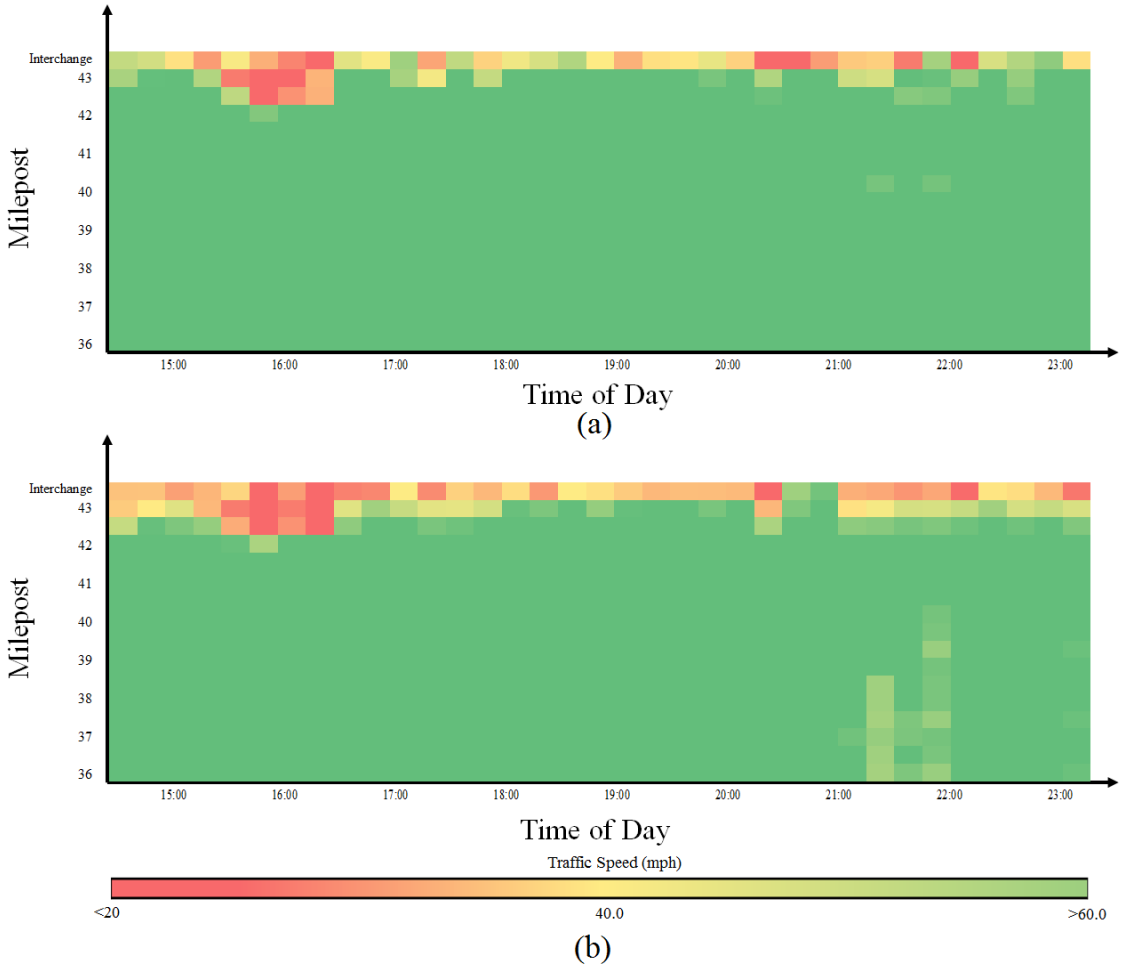


Figure 5.8 Heat map of traffic speed on I-80 EB from (a) the CNN prediction model (b) the actual traffic speed reported from INRIX.

The absolute error between the CNN model results and actual traffic speed heat maps are illustrated in Figure 5.9 for (a) I-287 SB (b) I-80 WB and (c) I-80 EB. The results indicate higher absolute errors around the 3-miles upstream area in I-287 SB. Additionally, the closer segments to the work zone tend to have higher absolute errors compared to further segments.

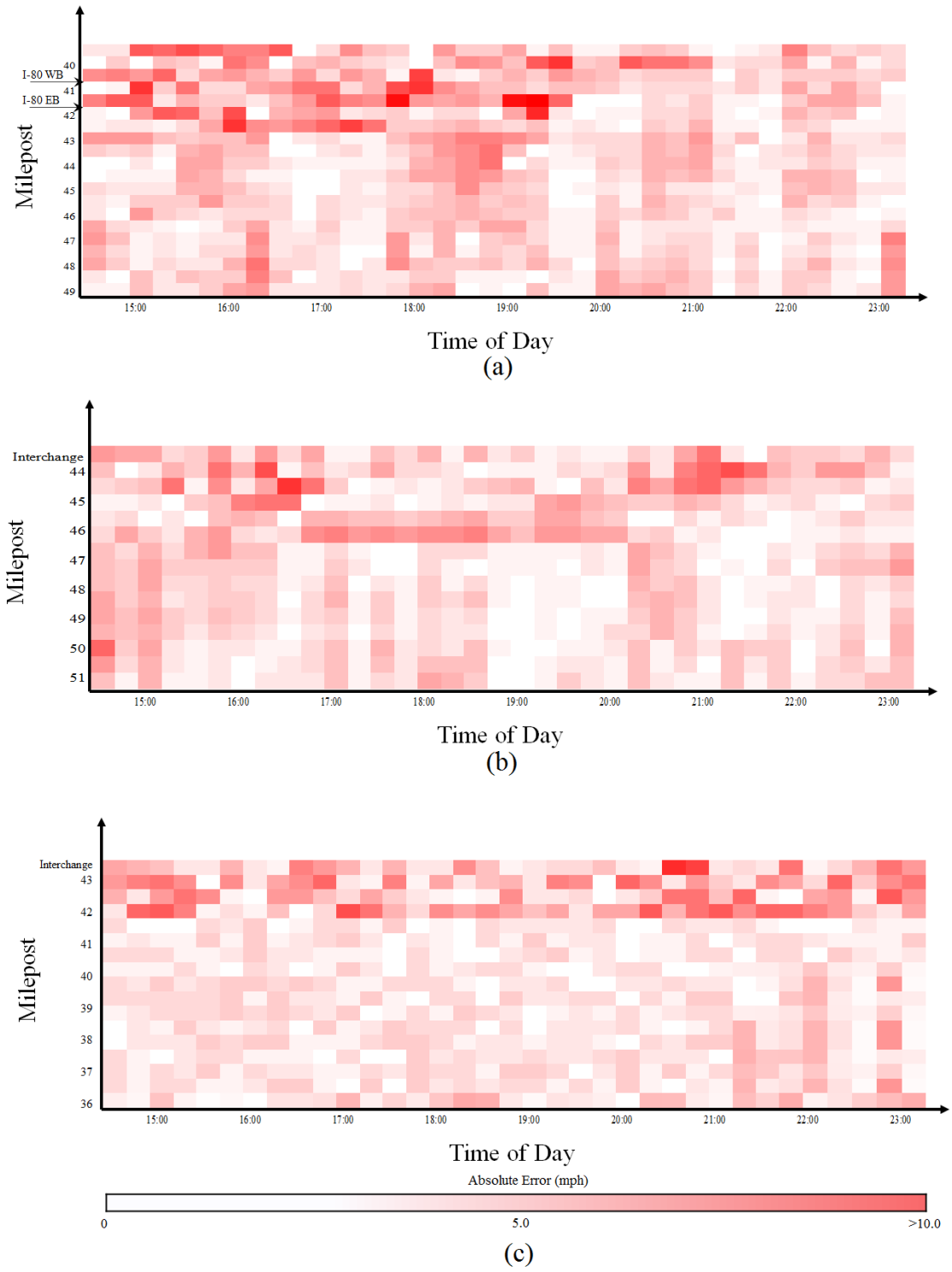


Figure 5.9 Heat map of absolute error of the CNN results again the actual speed for (a) I-287 SB (b) I-80 WB and (c) I-80 EB.

Traffic Delay is the additional delay caused by work zone congestion due to traffic speed reduction from traffic speed of segment i at time j during normal conditions s_{ij} to traffic speed under work zone condition $\hat{y}_{ij}$. Given a traffic volume V_{ij} for segment i at time j and l_i as a length of segment i, a total queue delay caused by work zone D is denoted in Equation (5.1).

$$D = \sum_{i=1}^{m}\sum_{j=1}^{n} \max\{l_i[\frac{1}{\hat{y}_{ij}} - \frac{1}{s_{ij}}]\tau_{ij}V_{ij}, 0\} \qquad \forall \tau_{ij} = 1 \tag{5.1}$$

Where, τ_{ij} represents a congestion status of segment i at time j. As denoted in Equation (5.2), τ_{ij} is 1 when it is congested and 0 otherwise

$$\tau_{ij}\begin{cases}1 & if\ \hat{y}_{ij} \leq 0.75 * s_{ij} \\ 0 & Otherwise\end{cases} \tag{5.2}$$

Delay cost C_d , as denoted by Equation 5.3, is calculated based on the percentage of passenger cars P_c of the overall traffic volume and the percentage of trucks P_t of the overall traffic volume.

$$C_d = P_c\mu_c + P_t\mu_t \tag{5.3}$$

where:

μ_c is the value of travel time delay for passenger cars ($/veh-hr)

μ_t is the value of travel time delay for trucks ($/veh-hr)

The queue length L_j at time j, which is defined in Equation (5.4) is the total length of the congested segments, affected by the work zone.

$$L_j = \sum_{i=1}^{m} \tau_{ij} * l_i \qquad \forall \tau_{ij} = 1 \tag{5.4}$$

This research investigates the accuracy of the model in relation to the distance from the work zone. Figure 5.10 demonstrate the RMSE values with a variation of the distance from the work zone. The mainline freeway is I-287 Southbound and there are two connected freeways 1.5-mile upstream work zone (i.e., I-80 Eastbound, and I-80 Westbound). The results show that the RMSE for I-287 Southbound is lower than the RMSE for both connected freeways at locations that have a distance to work zone greater than 4-miles.

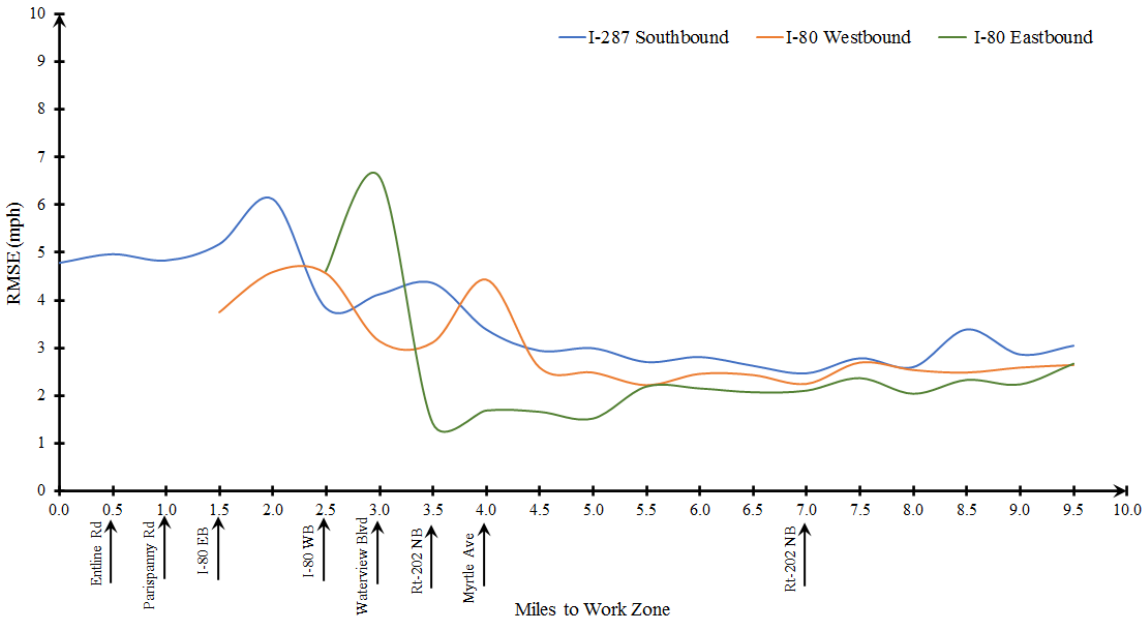


Figure 5.10 The RMSE values in variation of distance to work zone.

Predicting work zone congestion before they happen is one of the prime concerns of transportation agencies. Traffic congestion leads to users delay due to work zone. This study develops a CNN model to predict traffic speed on the upstream mainline segments and the connected freeway segments.

Traffic Delay is the additional delay caused by work zone congestion due to traffic speed reduction compared to traffic speed under normal conditions. In this case study

traffic delay is calculated for the mainline segments and the connected freeways segments. Figure 5.11 shows the change of users delay on I-287 due to work zone lane closure, varying over time. The results indicate that for the selected work zone, the model of WIMAP-P underestimates the work zone delay during the start of the work zone, but overestimates the results around the 17:30 time period. Figure 5.12 demonstrates the users delay variation over time on the connected freeway segments (i.e., I-80 Eastbound and I-80 Westbound) for the CNN model and the actual data. The results show that the CNN model overestimates the user delay between 15:30 and 16:30 of the work zone. Based on the results of Figures 5.11 and 5.12, more errors can occur during peak-hour periods, in which traffic volume tend to be high. Thus, an evaluation in section 5.3 is conducted to evaluate the models under various V/C ratios.

It is worth noting that previous models (e.g., the model of WIMAP-P) cannot capture the user delay of these connected freeways, resulting in less reported user delays. The results show that the CNN users delay, on the connector segments, is underestimated between 15:30 and 16:00, when compared with actual users' delay. On the other hand, the model of WIMAP-P results overestimates the work zone delays between 17:00 and 19:00. WIMAP-P does not provide delay prediction on the connectors. Thus, only mainline segments are included in the analysis. From the analysis, it can be surmised that the predicted and actual values are almost the same after 17:30 PM.

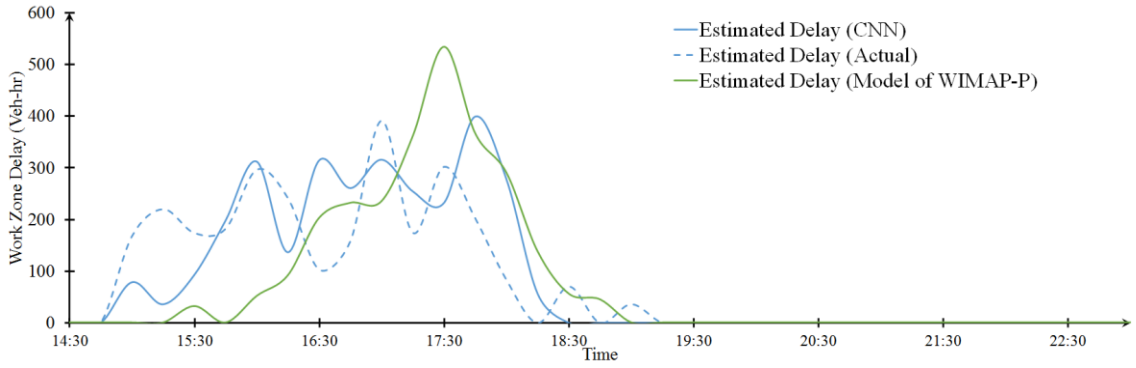


Figure 5.11 Delay varying over time for I-287 freeway segments using CNN, WIMAP-P, and the actual results.

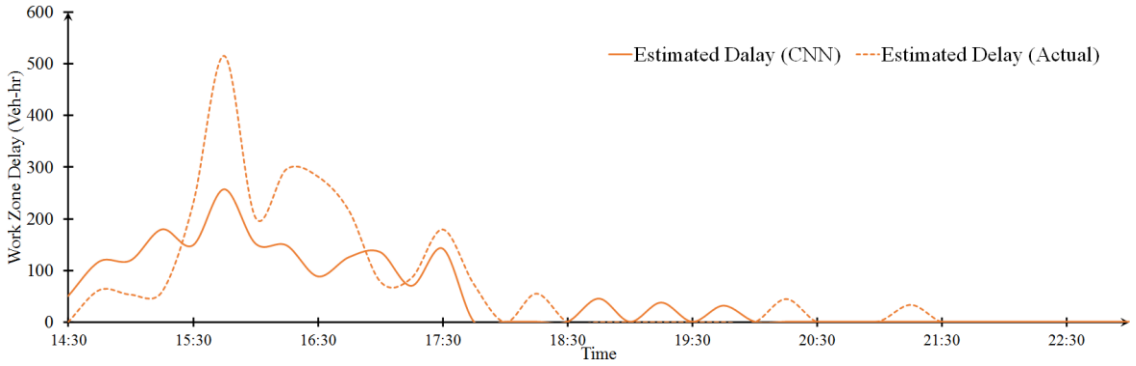


Figure 5.12 Estimated Delay varying over time for the connected freeway segments (i.e., I-80 EB and I-80 WB) using the CNN model and the actual results.

To summarize the results of the analysis, a total delay cost is conducted for each method and for both of the mainline and the connected freeways. Figure 5.13 demonstrates the comparison between the mainline and the connected freeway segments for CNN, the model of WIMAP-P, and actual delay cost results. The mode of WIMAP-P is unable to predict the work zone delay for connecting freeways, providing in accurate final results. The comparison indicates that the CNN model overestimates the work zone delay cost compared to the actual delay whereas the WIMAP-P model underestimate the delay. On the other hand, the CNN model underestimates the work zone delay costs for the connector segments.

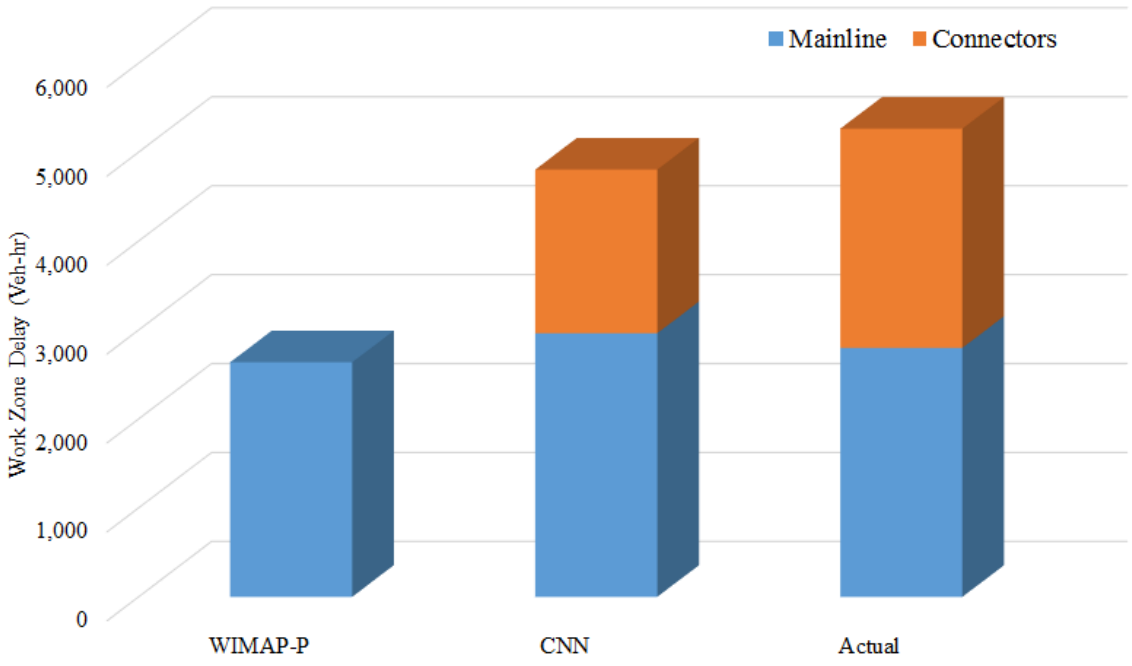


Figure 5.13 Comparison of total delay cost for both the mainline (i.e., I-287 SB) and the connectors segments (i.e., I-80 EB and I-80 WB) to the actual work zone delay.

In Figure 5.14, the queue length is demonstrated for the mainline segments using WIMAP-P, the CNN, and the actual queue length. Figure 5.15 shows the queue length on the connected freeways obtained from the CNN model, and the actual estimated queue length. The results indicate that WIMAP-P overestimates the queue length at the start of the work zone but overestimates the results around 17:30 time period. Moreover, the results indicate higher errors at the peak-hour periods (i.e., traffic volume is high). Additionally, the connector queue length is compared between the CNN predicted results and the actual queue length. The results indicate that WIMAP-P underestimates the queue length between 15:00 and 16:30 whereas it overestimates the queue length between 17:00 and 19:00. The CNN model, on the other hand, overestimates the queue length between 21:15 and 21:45.

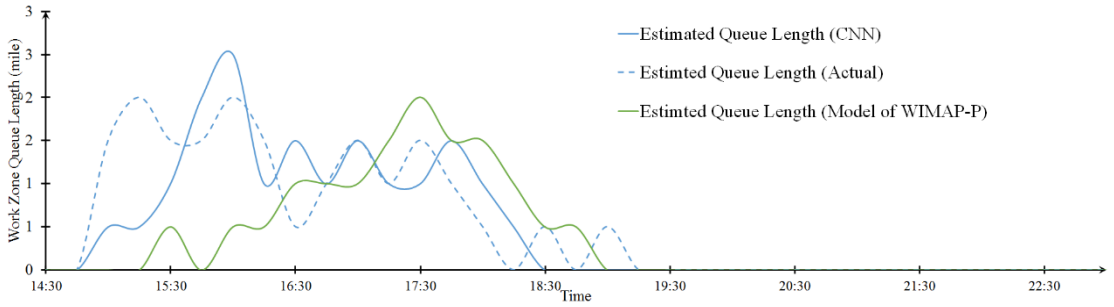


Figure 5.14 Queue length varying over time for the I-287 SB route using the CNN model, the WIMAP-P, and the actual results.

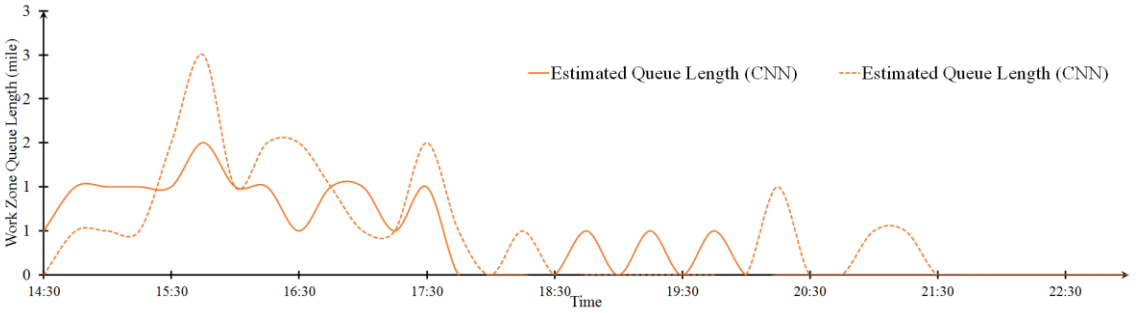


Figure 5.15 Queue length varying over time the connected freeways (i.e., I-80 EB and I-80 WB) using the CNN model and the actual results.

5.3 Models Comparison

WIMAP-P does not cover congestion spillback on freeways. It only predicts traffic speed on the same freeway that work zone occurs. Therefore, for comparison reasons, the connected and ramp segments are emitted from the data to include only mainline segments. Therefore, a comparison analysis between the Deep ANN model, the CNN model, and the WIMAP-P model is shown in Table 5.1, in which only mainline freeway segments are considered, and based on testing sample size indicated in Table 4.2. The results show high accuracy for Deep ANN and CNN models compared to WIMAP-P model, and slight improvement in terms of the accuracy for the CNN model. Additionally, the results indicate that with lower sample size for each category, the accuracy of the prediction model becomes less (e.g., two-lane closure scenarios).

Table 5.1 RMSE with Deep ANN, CNN, and WIMAP-P under different lane-closure

Model	Number of Lanes	RMSE (mph) (% of testing data)		
		Shoulder Closure	One Lane Closure	Two Lane Closure
Deep ANN	2	6.2 (2%)	6.2 (6%)	NA (0%)
	3	6.1 (8%)	5.8 (20%)	7.4 (1%)
	4	7.0 (2%)	6.2 (5%)	7.6 (1%)
CNN	2	6.0 (2%)	5.9 (6%)	NA (0%)
	3	5.8 (8%)	5.2 (20%)	7.2 (1%)
	4	6.4 (2%)	5.8 (5%)	7.3 (1%)
WIMAP-P	2	7.7 (2%)	8.8 (6%)	NA (0%
	3	9.1 (8%)	9.4 (20%)	10.4 (1%)
	4	9.9 (2%)	9.6 (5%)	10.8 (1%)

Since the Deep ANN and CNN models outperforms the WIMAP-P, a more inclusive comparison analysis including both the mainline and the connected freeway segments is conducted in Table 5.2. The results in Table 5.1 show higher accuracy for CNN model in addition to less accurate results when compared to Table 5.2 in general. Therefore, the results indicate that connected freeway prediction results have less accurate results than mainline segment predictions in most of lane closure types. Furthermore, the low sample size, shown in Table 4.2, leads to greater RMSE values (e.g., two-lane closure scenarios).

Table 5.2 RMSE with Deep ANN and CNN, considering other freeway segments.

Model	Number of Lanes	RMSE (mph) (% of testing data)		
		Shoulder Closure	One Lane Closure	Two Lane Closure
Deep ANN	2	6.6 (3 %)	6.0 (13%)	NA (0%)
	3	6.4 (20%)	5.8 (44%)	7.5 (2%)
	4	7.3 (3%)	6.2 (14%)	7.8 (1%)
CNN	2	6.2 (3%)	5.8 (13%)	NA (0%)
	3	5.9 (20%)	5.2 (44%)	7.4 (2%)
	4	6.7 (3%)	6.1 (14%)	7.6 (1%)

To further evaluate the accuracy of the CNN and the Deep ANN models, the testing results are analyzed in terms of weather conditions. Table 5.3 shows the comparison results of rain and no rain results for Deep ANN and CNN. The results indicate that in rain conditions, the accuracy of both the CNN model and the Deep ANN model is less, compared to no-rain conditions. Additionally, the accuracy of the models is assessed based on the distance of the segment to the work zone.

Table 5.3 RMSE Values of Deep ANN and CNN Models in Terms of Weather Conditions

Weather Condition	Distance to Work Zone	Model	Number of Lanes	RMSE (mph) (% of testing data)		
				Shoulder Closure	One Lane Closure	Two Lane Closure
Rain	Less than 5 miles	Deep ANN	2	7.4 (13%)	6.0 (9 %)	NA (0%)
			3	6.5 (6%)	6.0 (5 %)	NA (0 %)
			4	NA (0%)	7.9 (3%)	NA (0 %)
		CNN	2	7.0 (13%)	5.6 (9%)	NA (0%)
			3	6.0 (6%)	5.5 (5%)	NA (0 %)
			4	NA (0%)	7.7 (3%)	NA (0 %)
	Greater than 5 miles	Deep ANN	2	5.1 (20%)	6.1 (14 %)	NA (0%)
			3	7.5 (9%)	6.4 (8 %)	NA (0 %)
			4	NA (0%)	6.5 (7%)	NA (0 %)
		CNN	2	4.7 (20%)	5.9 (14%)	NA (0%)
			3	7.2 (9%)	5.2 (8%)	NA (0 %)
			4	NA (0%)	6.2 (7%)	NA (0 %)
No Rain	Less than 5 miles	Deep ANN	2	7.4 (25%)	6.3 (26%)	NA (0%)
			3	7.5 (30%)	5.3 (28%)	8.8 (32%)
			4	7.7 (37%)	6.3 (32%)	8.5 (30%)
		CNN	2	6.9 (25%)	5.9 (26%)	NA (0%)
			3	7.1 (30%)	5.0 (28%)	8.5 (32%)
			4	7.4 (37%)	5.9 (32%)	8.3 (30%)
	Greater than 5 miles	Deep ANN	2	6.9 (42%)	6.0 (51%)	NA (0%)
			3	6.8 (55%)	5.6 (59%)	7.3 (68 %)
			4	7.0 (63%)	6.6 (58%)	7.6 (70 %)
		CNN	2	6.2 (42%)	5.6 (51%)	NA (0%)
			3	6.4 (55%)	5.2 (59%)	6.9 (68 %)
			4	6.2 (63%)	6.1 (58%)	7.3 (70 %)

For further evaluation of the models, Table 5.4 shows the comparison results between the models in terms of volume approaching work zone over work zone capacity denoted as (V/C_w) ratio and the segments' distance to the work zone. The results indicate that high V/C_w ratios corresponds to higher RMSE values.

Table 5.4 RMSE Values of Deep ANN and CNN Models for two various V/C_w ratios categories.

V/C_w Ratio	Distance to Work Zone	Model	Number of Lanes	RMSE (mph) (% of testing data)		
				Shoulder Closure	One Lane Closure	Two Lane Closure
$V/C_w \geq 0.5$	Less than 5 miles	Deep ANN	2	NA (0%)	6.3 (18 %)	NA (0%)
			3	5.9 (5%)	5.6 (11 %)	NA (0 %)
			4	7.5 (3%)	6.7 (15%)	NA (0 %)
		CNN	2	NA (0%)	5.9 (18%)	NA (0%)
			3	5.7 (5%)	5.1 (11%)	NA (0 %)
			4	7.1 (3%)	6.3(15%)	NA (0 %)
	Greater than 5 miles	Deep ANN	2	NA (0%)	6.2 (30 %)	NA (0%)
			3	6.2 (11%)	5.7 (21%)	NA (0 %)
			4	7.1 (7%)	6.7 (29%)	NA (0 %)
		CNN	2	NA (0%)	5.9 (30%)	NA (0%)
			3	5.9 (11%)	5.3 (21%)	NA (0 %)
			4	6.7 (7%)	6.3 (29%)	NA (0 %)
$V/C_w < 0.5$	Less than 5 miles	Deep ANN	2	7.4 (35%)	5.9 (17%)	NA (0%)
			3	10.3 (27%)	5.5 (23%)	10.5 (31%)
			4	6.9 (30%)	6.4 (20%)	10.8 (32%)
		CNN	2	7.0 (35%)	5.7 (17%)	NA (0%)
			3	9.9 (27%)	5.1 (23%)	9.9 (31%)
			4	6.5 (30%)	6.1 (20%)	10.3 (32%)
	Greater than 5 miles	Deep ANN	2	6.1 (65%)	5.8 (35%)	NA (0%)
			3	4.8 (57%)	5.8 (45%)	6.5 (69 %)
			4	6.2 (60%)	6.1 (36%)	6.5 (68 %)
		CNN	2	5.8 (65%)	5.7 (35%)	NA (0%)
			3	4.2 (57%)	5.4 (45%)	6.2 (69 %)
			4	5.9 (60%)	5.8 (36%)	6.3 (68 %)

The developed deep learning models are assessed based on the location of the TMC segment. Two main categories of TMC segments are distinguished: Type 1, which

is the TMC segments on the mainline only immediate upstream to the on-ramp and Type 2, which is all the other TMC segments on the mainline and connected freeway segments. Table 5.5 provides the RMSE values of both the Deep ANN and CNN models for both Type 1 and Type 2 TMC segments. It is worth noting that the developed models of Deep ANN and CNN have higher error for the Type 1 TMC segments compared to Type 2.

Table 5.5 RMSE Values of Deep ANN and CNN Models for Two TMC Categories.

Type of TMC segment	Model	Number of Lanes	RMSE (mph) (% of testing data)		
			Shoulder Closure	One Lane Closure	Two Lane Closure
Type 1	Deep ANN	2	11.2 (5%)	9.5 (13 %)	NA (0 %)
		3	12.3 (8%)	9.1 (12 %)	10.5 (6 %)
		4	14.9 (4%)	11.0 (10 %)	11.3 (3 %)
	CNN	2	10.0 (5%)	9.2 (13%)	NA (0%)
		3	11.6 (8%)	8.2 (12%)	9.9 (6 %)
		4	14.1 (4%)	10.3 (10%)	10.6 (3 %)
Type 2	Deep ANN	2	6.4 (95%)	5.5 (87%)	NA (0%)
		3	5.9 (92%)	5.4 (88%)	7.3 (94 %)
		4	7.0 (96%)	5.7 (90%)	7.7 (97 %)
	CNN	2	6.0 (95%)	5.3 (87%)	NA (0%)
		3	5.4 (92%)	4.8 (88%)	7.2 (94 %)
		4	6.4 (96%)	5.8 (90%)	7.5 (97 %)

Based on the previous analysis, the results indicate higher accuracy for the developed CNN model compared to the Deep ANN and the WIMAP-P models. Adding the dropout decreases the overfitting problem. However, the CNN model further mitigates the overfitting problem through reducing the matrix size in the hidden layers to include only the important features. Hence, for the available work zone data in New Jersey, the CNN model shows higher fidelity compared to Deep ANN and WIMAP-P, when conducting the model comparison in Chapter 5.

5.4 Applications

The proposed CNN model is developed to predict delay cost on both the mainline and the connector freeway segments. The model and the developed database can be used to support state, local TOC, the planning agencies, and work zone contractors to:

- Quantify the congestion costs (i.e., spatio-temporal), due to work zone activities (e.g., shoulder closure, lane closures) in the freeway system of the State of New Jersey. When the transportation agencies develop a congestion mitigation plans, the congestion on the connected freeways may be included by using the developed model of this research.
- Identify the user delay costs for each roadway connected to the freeway. Therefore, various road agencies can collaborate in mitigating the effect of work zone congestion. For instance, a connected freeway may be in the jurisdiction of another agency, affecting the user delay costs. If user delays are increased on the connected freeway segments, users might not use this route, impacting any existing toll revenues, if there is any.
- Conduct a sensitivity analysis between user delay costs vis-à-vis agencies costs when planning for the work zone activities. When the transportation agencies schedule a work zone, they outline the different options to start the work zone, depending on the minimum value of user and agency costs combined. However, by not including other connected freeways, the reported user delay costs may be less than the actual one. Consequently, the developed model may be used to aid the agencies in reflecting more accurate total user delay costs.
- Assess queue warning systems on the connected freeways that are predicted to have a congestion spillback, by predicting the locations of potential congestion spillbacks. The transportation agencies usually distribute queue warning systems upstream work zone. However, with a limited resources environment, the agencies can prioritize the locations of their queue warning systems depending on the predicted traffic speed of the developed model and the level of congestion of the connected freeways upstream work zone.

One example of how the developed CNN model can be deployed is by deploying the work zones that are scheduled during the day and predicted to have congestion spillback, to be during the night periods. Depending on the agency costs for deploying work zones during the night and the user delay costs, the work zone schedule is decided. The developed model can be useful for congestion mitigation plans. One of these

mitigation plans can include rerouting traffic upstream work zones, including rerouting upstream connected freeway segments that are predicted to have a congestion.

5.5 Summary

This study illustrates the functionalities of the developed model in a case study. The case study is chosen in which an upstream connected freeway is located upstream the work zone. The results are demonstrated in a heat map method to show the spatio-temporal variations in predicted traffic speed under work zone conditions.

This research compares the results between Deep ANN, CNN, and the model of WIMAP-P. WIMAP-P is developed to predict traffic speed on the mainline segments only. Thus, a subset of the data is used for the comparison between the three models. It is found that the CNN model outperforms both the Deep ANN and the model of WIMAP-P. The CNN and Deep ANN models are evaluated and discussed under various scenarios (i.e., weather conditions, distance to work zone, and V/C ratio).

CHAPTER 6
CONCLUSIONS AND FUTURE RESEARCH

Predicting work zone congestion before they happen is one of the prime concerns of transportation agencies. With the increase of infrastructures ages, roadway rehabilitation and construction activities are becoming necessary. With the increase of work zone activities, transportation agencies need to plan their work zone activities ahead of time. Therefore, predicting work zone activities precisely is becoming increasingly critical. Additionally, work zone congestion may spillback on other freeways leading to more congestion. In response to this challenge, two models, the Deep ANN and the CNN models, for predicting work zone delay were developed using big data in this study. In the Deep ANN model, multiple layers are considered in addition to integrating the dropout technique to mitigate the overfitting problems traditional ANN model suffer. In the CNN model, convolutional layers are added to mitigate the overfitting by extracting the important features from the previous layers. The CNN model shows higher accuracy compared to the Deep ANN model and the ANN model used in WIMAP-P.

6.1 Conclusions

The developed CNN model for predicting traffic speed and delay cost under work zone conditions faced various challenges and improvements in the areas of data collection and the performance measurements, which is listed in the next sections.

6.1.1 Spatio-temporal Work Zone Delay Prediction

The proposed CNN model uses various parameters (e.g., filter height, filter stride, number of layers, and number of neurons at each layer). The CNN parameters are chosen through sensitivity analysis, based on the freeway network in New Jersey. When evaluating the CNN model, the results indicate that the CNN model outperforms the Deep ANN model and the model used in WIMAP-P. Consequently, the CNN model is the least affected by the overfitting problem, especially when dropout is integrated into the model.

The developed CNN model can be deployed to help aiding transportation agencies in predicting traffic congestion upstream work zone, including plans for connected freeways. The model can be helpful for planning purposes, including determining the start and end timing of work zone, including the connected freeways as a decision variable. Moreover, contractor penalties can be assessed to reflect more accurate user delay costs. On the other hand, contractors may have reward incentives that are more precise.

6.1.2 Big Data Analysis in Work Zone Impact Studies

With the technological advancement in collecting data, data analysis has become a focal point in any modeling. The increase of the amount of the collected data over the recent years has led to big data analysis that is able to uncover hidden information in the datasets. Transportation agencies can analyze enormous information and make decisions according to the data insights.

In the freeway work zone analysis, the available datasets include various inconsistent data. Therefore, a big data analysis is required to extract the accurate information. Deep ANN and CNN models require data for training and validation purposes. Unlike deterministic approaches and other traditional ANN approaches, the CNN model

can predict the work zone impact more accurately than the other ones. Big data analysis offers flexibilities in managing and transforming the data between different models. Hence, transportation agencies can invest in these new technologies to enhance and improve the work zone operation management aspects and reduce the users delay costs.

6.1.3 Research Findings

This research develops a CNN model to predict traffic speed under work zone conditions for mainline and connector freeway segments. The major findings of the research is summarized:

- The developed model is affected by the closeness to the work zone and the by the proximity to the mainline links immediate upstream on the on-ramp.
- Traffic speed is collected from INRIX database, which reports the speed using longitude and latitude systems. However, the freeway geometric information is defined using the milepost systems. Thus, matching the INRIX database with the milepost system requires substantial amount of time, in which some of the segments are matched manually.
- The traffic volume information is vital for predicting traffic congestion on the upstream segments, which is not available in most roadways. Therefore, NJCMS dataset is used for model development.
- Weather data is used to evaluate the models. However, when more work zone is available under adverse weather conditions, weather data can be considered in the inputs of the model in the development processes.
- Driver behavior is not considered as an input in the model. For instance, delay can vary between commuter routes and recreational routes.
- The ramps are not illustrated in NJSLD, making the identification of the ramp segment names in INRIX a difficult task. Additionally, the ramps intersection points with the freeway segments needs to be identified. In this study, after the identification the intersection points and the ramp segments, each freeway would have a new developed network identifying all the connected segments and the intersection points. However, as INRIX is adding new segments, the mapping needs to be done again.

6.2 Future Research

The developed CNN model for predicting traffic speed under work zone conditions can be enhanced in the following areas:

The traffic volume is calculated using NJCMS, identifying new sources for updating the traffic volume to reflect the actual traffic volume would enhance the accuracy of the CNN model. Moreover, incorporating a simulated model to estimate the work zone capacity can advance the developed CNN model. The improvement in the OpenReach data, by incorporating more agencies in addition to the precise location would enhance the training of the CNN model and the quality of the data. Thus, transportation agencies would be able to identify the starting and ending time of the work zone more accurately.

The databases are growing, making downloading big data using traditional techniques burdensome. Consequently, automating the databases through repositories would ease the data analysis for new products and developments. Additionally, the privacy issues for data sharing can be excluded for research purposes, allowing assessment of the databases in terms of accuracy. Databases can be open-sourced, for research purposes, to ease the collaboration between different agencies.

INRIX XD database can be used to enhance the accuracy of the model, as the granularity of this database can reach to 0.1-mile TMC segment length. Furthermore, integrating traffic volume and other geometric information would be useful for matching the databases. Other crowdsourcing datasets can be considered to enhance the accuracy of the model (e.g., WAZE). A comparison analysis between various datasets can enrich the understanding of the advantages and disadvantages of the datasets. Additionally, with the availability of high granularity data, the model can be enhanced to predict traffic congestion on the opposite traffic direction (i.e., opposite bound).

The CNN model, in this study, is used to predict traffic speed on the upstream mainline and connected freeways, under work zone conditions. The proposed CNN model can be extended to include the following functionalities: (a) an optimal work zone scheduling with rerouting plans (b) work zone staging optimization (c) combination of work zone and accidents prediction modulus.

REFERENCES

511NJ. (2020). 511NJ. Retrieved on February 2020, from https://511nj.org/trafficmap

Ahsani, V., Amin-Naseri, M., Knickerbocker, S., & Sharma, A. (2019). Quantitative analysis of probe data characteristics: Coverage, speed bias and congestion detection precision. *Journal of Intelligent Transportation Systems*, *23*(2), 103-119.

Bloomberg, L., & Dale, J. (2000). Comparison of VISSIM and CORSIM traffic simulation models on a congested network. *Transportation Research Record*, *1727*(1), 52-60.

Bottou, L. (2012). Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade* (pp. 421-436). Springer, Berlin, Heidelberg.

Bochinski, E., Eiselein, V., & Sikora, T. (2016, August). Training a convolutional neural network for multi-class object detection using solely virtual world data. In *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 278-285). IEEE.

Boukhechba, M., Bouzouane, A., Gaboury, S., Gouin-Vallerand, C., Giroux, S., & Bouchard, B. (2017). A novel Bluetooth low energy based system for spatial exploration in smart cities. *Expert Systems with Applications*, *77*, 71-82.

Bronzi, W. (2017). *Enhancing Mobility Applications Through Bluetooth Communications* (Doctoral dissertation, University of Luxembourg, Luxembourg).

Burt, M., Cuddy, M., & Razo, M. (2014). *Big data's implications for transportation operations: an exploration* (No. FHWA-JPO-14-157; DOT-VNTSC-FHWA-14-13). United States. Department of Transportation. Intelligent Transportation Systems Joint Program Office.

Cathey, F. W., & Dailey, D. J. (2005, June). A novel technique to dynamically measure vehicle speed using uncalibrated roadway cameras. In *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.* (pp. 777-782). IEEE.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, *7*(3), 1247-1250.

Chen, H., & Rakha, H. (2014). *Agent-based modeling approach to predict experienced travel times* (No. 14-3851).

Chien, S., & Schonfeld, P. (2001). Optimal work zone lengths for four-lane highways. *Journal of Transportation Engineering*, *127*(2), 124-131.

Chung, Y., Kim, H., & Park, M. (2012). Quantifying non-recurrent traffic congestion caused by freeway work zones using archived work zone and ITS traffic data. *Transportmetrica*, *8*(4), 307-320.

Chung, J., & Sohn, K. (2017). Image-based learning to measure traffic density using a deep convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, *19*(5), 1670-1675.

CoVal Systems. (2016). Retrieved on (January 2020) from http://www.covalsystems.com/latest/openreach/openreach.html

Dogru, N., & Subasi, A. (2018, February). Traffic accident detection using random forest classifier. In *2018 15th learning and technology conference (L&T)* (pp. 40-45). IEEE.

Dorai, Y., Gazzah, S., Chausse, F., & Ben Amara, N. E. (2016, September). Tracking multi-object using tracklet and Faster R-CNN: PhD Forum. In *Proceedings of the 10th International Conference on Distributed Smart Camera* (pp. 222-223).

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, *12*(Jul), 2121-2159.

Dudek, C. L., & Richards, S. H. (1982). *Traffic capacity through urban freeway work zones in Texas* (No. 869).

Du, B., Lee, J., Chien, S., Dimitrijevic, B., & Kim, K. (2015). *Short-term freeway work zone capacity estimation using support vector machine incorporated with probe-vehicle data* (No. 15-4248).

Du, B., Chien, S., Lee, J., & Spasovic, L. (2017). Predicting Freeway Work Zone Delays and Costs with a Hybrid Machine-Learning Model. *Journal of Advanced Transportation*, *2017*.

Du, B., Steven, I., & Chien, J. (2014). Feasibility of shoulder use for highway work zone optimization. *Journal of Traffic and Transportation Engineering (English edition)*, *1*(4), 235-246.

Edara, P. K., & Cottrell, B. H. (2007, January). Estimation of traffic mobility impacts at work zones: state of the practice. In *Proceedings of the Transportation Research Board 2007 Annual Meeting*.

Elhenawy, M., Chen, H., & Rakha, H. A. (2014). Dynamic travel time prediction using data clustering and genetic programming. *Transportation Research Part C: Emerging Technologies*, *42*, 82-98.

Elhenawy, M., & Rakha, H. A. (2017). *Random Forest-Hidden Markov Transportation Mode Recognition Model Using Smartphone Sensor Data* (No. 17-00213).

Elisseeff, A., & Paugam-Moisy, H. (1997). Size of multilayer networks for exact learning: analytic approach. In *Advances in Neural Information Processing Systems* (pp. 162-168).

Eshragh, S., Young, S. E., Sharifi, E., Hamedi, M., & Sadabadi, K. F. (2017). Indirect validation of probe speed data on arterial corridors. *Transportation Research Record*, *2643*(1), 105-111.

FHWA. (2014, January 30). Work Zone Facts and Statistics. Retrieved on (January, 2020) from https://ops.fhwa.dot.gov/wz/resources/facts_stats/mobility.htm

Google. (2018). Google Maps API. Retrieved on (January 2020) from https://cloud.google.com/maps-platform/

Haghani, A., Hamedi, M., & Sadabadi, K. F. (2009). I-95 Corridor coalition vehicle probe project: Validation of INRIX data. *I-95 Corridor Coalition*, *9*.

Hahm, Y., Jun, Y., Kim, K., & Kim, S. (2017). The Types of Road Weather Big Data and the Strategy for Their Use: Case Analysis. *The Korean Journal of Bigdata*, *2*(2), 129-140.

Hanin, B. (2018). Which neural net architectures give rise to exploding and vanishing gradients?. In *Advances in Neural Information Processing Systems* (pp. 582-591).

Hard, E. N., Chigoy, B. T., Songchitruksa, P., Farnsworth, S. P., Borchardt, D. W., & Green, L. L. (2017). *Comparison of cell, GPS, and bluetooth derived external OD data–results from the 2014 Tyler, Texas Study* (No. 17-05678).

Hou, Y., & Edara, P. (2018). Network scale travel time prediction using deep learning. *Transportation Research Record*, *2672*(45), 115-123.

Hutchinson, M. F. (2011). ANUDEM version 5.3, user guide. *Canberra: Fenner School of Environment and Society, Australian National University*.

Im, H., Hong, B., Jeon, S., & Hong, J. (2016, January). Bigdata analytics on CCTV images for collecting traffic information. In *2016 International Conference on Big Data and Smart Computing (BigComp)* (pp. 525-528). IEEE.

INRIX. (2020). INRIX Home Page. Retrieved on (January 2020) from http://inrix.com/

Karim, A., & Adeli, H. (2003). Radial basis function neural network for work zone capacity and queue estimation. *Journal of Transportation Engineering*, *129*(5), 494-503.

Kim, T., Lovell, D. J., & Paracha, J. (2001, January). A new methodology to estimate capacity for freeway work zones. In *80th Annual Meeting of the Transportation Research Board, Washington, DC*.

Kim, Y., Wang, P., Zhu, Y., & Mihaylova, L. (2018, October). A capsule network for traffic speed prediction in complex road networks. In 2018 Sensor Data Fusion: Trends, Solutions, Applications (SDF) (pp. 1-6). IEEE.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krammes, R. A., & Lopez, G. O. (1994). *Updated capacity values for short-term freeway work zone lane closures* (No. 1442).

Lambert, J., Sener, O., & Savarese, S. (2018). Deep learning under privileged information using heteroscedastic dropout. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8886-8895).

Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *229*(1178), 317-345.

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187-197.

Maze, T., & Kamyab, A. (1999). Work zone simulation model.

McCoy, P. T., Pang, L. M., & Post, E. R. (1980). Optimum length of two-lane two-way no-passing traffic operation in construction and maintenance zones on rural four-lane divided highways. *Transportation research record*, *773*, 20-24.

McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E. Y., ... & McHenry, J. (2005). Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004. *Journal of Geophysical Research: Atmospheres*, *110*(D21).

Memmott, J. L., & Dudek, C. L. (1982). *A model to calculate the road user costs at work zones* (No. FHWA-TX-83-20+ 292-1 Intrm Rpt.).

Middleton, D., Rajbhandari, R., Brydia, R., Songchitruksa, P., Kraus, E., Hernandez, S., ... & Turner, S. (2011). Synthesis of TxDOT uses of real-time commercial traffic data.

Moody, J. E. (1992). The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. In *Advances in Neural Information Processing Systems* (pp. 847-854).

Mudge, R., Mahmassani, H. S., Haas, R., Talebpour, A., & Carroll, L. (2013). *Work zone performance measurement using probe data* (No. FHWA-HOP-13-043). United States. Federal Highway Administration. Office of Operations.

Mukkamala, M. C., & Hein, M. (2017, August). Variants of rmsprop and adagrad with logarithmic regret bounds. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (pp. 2545-2553). JMLR. org.

Nguyen, H., Bentley, C., Kieu, L. M., Fu, Y., & Cai, C. (2019). Deep learning system for travel speed predictions on multiple arterial road segments. *Transportation Research Record*, *2673*(4), 145-157.

New Jersey Department of Transportation. (2018). Crash Records. Retrieved on (January 2020) from https://www.state.nj.us/transportation/refdata/accident/

New Jersey Department of Transportation. (2014). Straight Line Diagram. Retrieved on (January 2020) from https://www.state.nj.us/transportation/refdata/sldiag/

New Jersey Department of Transportation. (2015). New Jersey Congestion Management Systems. Retrieved on (January 2020) from https://www.state.nj.us/transportation/refdata/sldiag/

New Jersey Turnpike Authority Website: https://www.njta.com/media/4825/cafr_2016.pdf , accessed on January, 2020.

NJGIN. (2020). NJGIN Open Data. Retrieved on February, 2020, from https://njogis-newjersey.opendata.arcgis.com/datasets

Nyfors, E. (2000). Industrial microwave sensors—A review. *Subsurface Sensing Technologies and Applications*, *1*(1), 23-43.

Oregon Department of Transportation (2010) Retrieved on (January 2020) from https://www.oregon.gov/ODOT/Engineering/Docs_TrafficEng/Work-Zone-Analysis-Manual.pdf

Panchal, G., Ganatra, A., Shah, P., & Panchal, D. (2011). Determination of over-learning and over-fitting problem in back propagation neural network. *International Journal on Soft Computing*, *2*(2), 40-51.

Pu, H., Zhang, H., Li, W., & Wang, L. (2018). *Decision-Making of Railway Maximum Design Gradient in Mountains Terrain Using Deep Learning Algorithms* (No. 18-04161).

Puckett, D. D., & Vickich, M. J. (2010). *Bluetooth-based travel time/speed measuring systems development* (No. UTCM 09-00-17). Texas Transportation Institute. University Transportation Center for Mobility.

Richards, P. I. (1956). Shock waves on the highway. *Operations research*, *4*(1), 42-51.

Satellite Imagery Corporation. (2020). Digital Elevation Models (DEM). Retrieved on February 2020, from https://www.satimagingcorp.com/services/dem/

Savage, N. H., Agnew, P., Davis, L. S., Ordóñez, C., Thorpe, R., Johnson, C. E., ... & Dalvi, M. (2013). Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation. *Geoscientific Model Development*, *6*(2), 353.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85-117.

Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Annals of Statistics*, *48*(4), 1875-1897.

Schroeder, B. J., & Rouphail, N. M. (2010). Estimating operational impacts of freeway work zones on extended facilities. *Transportation Research Record*, *2169*(1), 70-80.

Seymour, E. J., Chaudhary, N. A., Middleton, D., Brydia, R. E., & Miller, K. (2011). *White paper: state of ITS industry and assessment of project types* (No. FHWA/TX-11/0-6672-1). Texas. Dept. of Transportation. Research and Technology Implementation Office.

Shabarek, A., Chien, S., and Hadri, S., "Deep Learning Framework for Freeway Speed Prediction in Adverse Weather" *Transportation Research Record*, August 2020, https://doi.org/10.1177/0361198120947421

Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, *27*(6), 779-802.

Sreekumar, U. K., Devaraj, R., Li, Q., & Liu, K. (2017, August). TPCAM: Real-time traffic pattern collection and analysis model based on deep learning. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (pp. 1-4). IEEE.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929-1958.

Tang, Y., & Chien, S. I. J. (2008). Scheduling work zones for highway maintenance projects: Considering a discrète time-cost relation. *Transportation Research Record*, *2055*(1), 21-30.

Tarnoff, P. J., Bullock, D. M., Young, S. E., Wasson, J., Ganig, N., & Sturdevant, J. R. (2009). *Continuing evolution of travel time data information collection and processing* (No. 09-2030).

Transportation Safety Research Center. (2016). Plan4Safety. Retrieved on (January 2018) from https://cait.rutgers.edu/tsrc/plan4safety

Turner, S., & Qu, T. (2013). *Developing twin cities arterial mobility performance measures using GPS speed data* (No. MN/RC 2013-14). Department of Transportation, Research Services.

Ullman, G. L., & Dudek, C. L. (2003). Theoretical approach to predicting traffic queues at short-term work zones on high-volume roadways in urban areas. *Transportation Research Record*, *1824*(1), 29-36.

Ullman, G. L., Lomax, T. J., & Scriba, T. (2011). *A primer on work zone safety and mobility performance measurement* (No. FHWA-HOP-11-033). United States. Federal Highway Administration. Office of Operations.

United States Geological Survey. (2018). Digital Elevation Model. Retrieved from https://www.usgs.gov/science-explorer-results?es=Digital Elevation Model

Urbancic, J., Pejovic, V., & Mladenic, D. (2018). Transportation mode detection using random forest. *Information Society, Data Mining and Data Warehouses SiKDD, Ljubljana, Slovenia*.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Wang, C., Hao, P., Wu, G., Qi, X., & Barth, M. (2018). *Predicting the Number of Uber Pickups by Deep Learning* (No. 18-06738).

Washburn, S. S., Hiles, T., & Heaslip, K. (2008). *Impact of lane closures on roadway capacity: Development of a two-lane work zone lane closure analysis procedure (Part A)* (No. TRC-FDOT-59056-a-2008).

Weng, J., & Meng, Q. (2013). Estimating capacity and traffic delay in work zones: An overview. *Transportation Research Part C: Emerging Technologies*, *35*, 34-45.

Weston, J., Ratle, F., Mobahi, H., & Collobert, R. (2012). Deep learning via semi-supervised embedding. In *Neural networks: Tricks of the trade* (pp. 639-655). Springer, Berlin, Heidelberg.

Wilson, J. P., & Gallant, J. C. (Eds.). (2000). *Terrain analysis: principles and applications*. John Wiley & Sons.

Wright, J., & Dahlgren, J. (2001). *Using vehicles equipped with toll tags as probes for providing travel times* (No. UCB-ITS-PWP-2001-13).

Yang, H., Wang, Z., Xie, K., Ma, Y., & Zhu, Y. (2018). *A Deep Learning Approach to Predict Severity Levels of Work Zone Crashes* (No. 18-03042).

Yu, B., Wang, Y. T., Yao, J. B., & Wang, J. Y. (2016). A comparison of the performance of ANN and SVM for the prediction of traffic accident duration. *Neural Network World*, *26*(3), 271.

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

Zhang, S., Zhou, L., Chen, X., Zhang, L., Li, L., & Li, M. (2020). Network-wide traffic speed forecasting: 3D convolutional neural network with ensemble empirical mode decomposition. *Computer-Aided Civil and Infrastructure Engineering*.