12-2020

Traffic Time Headway Prediction and Analysis: A Deep Learning Approach

Saumik Sakib Bin Masud
The University of Texas Rio Grande Valley

TRAFFIC TIME HEADWAY PREDICTION AND ANALYSIS:

A DEEP LEARNING APPROACH

A Thesis

by

SAUMIK SAKIB BIN MASUD

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2020

Major Subject: Civil Engineering

TRAFFIC TIME HEADWAY PREDICTION AND ANALYSIS:
A DEEP LEARNING APPROACH

A Thesis
by

SAUMIK SAKIB BIN MASUD

COMMITTEE MEMBERS

Dr. Mohamadhossein Noruzoliaee
Chair of Committee

Dr. Fatemeh Nazari
Committee Member

Dr. Mohamed Abdel Raheem
Committee Member

December 2020

ABSTRACT

Masud, Saumik Sakib Bin, Traffic time headway prediction, and analysis: a deep learning approach. Master of Science (MS), December 2020, 112 pp., 7 tables, 35 figures, 74 references.

In the modern world of Intelligent Transportation System (ITS), *time headway* is a key traffic flow parameter affecting ITS operations and planning. Defined as "the time difference between any two successive vehicles when they cross a given point", time headway is used in various traffic and transportation engineering research domains, such as capacity analysis, safety studies, car-following, and lane-changing behavior modeling, and level of service evaluation describing stochastic features of traffic flow. Advanced travel and headway information can also help road users avoid traffic congestion through dynamic route planning, for instance. Hence, it is crucial to accurately model headway distribution patterns for the purpose of analyzing traffic operations and making subsequent infrastructure-related decisions. Previous studies have applied a variety of probabilistic models, machine learning algorithms (for example, support vector machine, relevance vector machine, etc.), and neural networks for short-term headway prediction. Recently, *deep learning* has become increasingly popular following a surge of traffic big data with high resolution, thriving algorithms, and evolved computational capacity. However, only a few studies have exploited this emerging technology for headway prediction applications. This is largely due to the difficulty in capturing the random, seasonal, nonlinear, and spatiotemporal correlated nature of traffic data and asymmetric human driving behavior

which has a significant impact on headway. This study employs a novel architecture of deep neural networks, *Long Short-Term Neural Network (LSTM NN)*, to capture nonlinear traffic dynamics effectively to predict vehicle headway. LSTM NN can overcome the issue of back-propagated error decay (that is, vanishing gradient problem) existing in regular Recurrent Neural Network (RNN) through memory blocks which is its special feature, and thus exhibits superior capability for time series prediction with long temporal dependency.

There is no existing appropriate model for long term prediction of traffic headway, as existing models lack using big dataset and solving the vanishing gradient problem because of not having a memory block. To overcome these critics and fill the gaps in previous works, multiple LSTM layers are stacked to incorporate temporal information. For model training and validation, this study used the USDOT's Next Generation Simulation (NGSIM) dataset, which contains historical data of some important features to describe the headway distribution such as lane numbers, microscopic traffic flow parameters, vehicle and road shape, vehicle type, and velocity. LSTM NN can capture the historical relationships between these variables and save them using its unique memory block. At the headway prediction stage, the related spatiotemporal features from the dataset (HighwayI-80) were fed into a fully connected layer and again tested with testing data for validation (both highway I-80 & US 101). The predicted accuracy outperforms previous time headway predictions.

DEDICATION

The completion of my master's studies would not have been possible without the love and support of my family and friends. This study is dedicated to my dear father MOHAMMAD MASUDUL HAQUE, my mother HOSNE JAHAN, my sister SAMIRA MEHNAZ, and my loving wife NAZIFA AKTER who wholeheartedly inspired, motivated and supported me by all means to accomplish this degree. Without their support I could not achieve any of my accomplishments. Thank you for your love and patience. Most importantly, all the praises to the almighty Who made everything possible and easy for me.

ACKNOWLEDGEMENTS

I will be always grateful to my supervisor/advisor Dr. Mohamadhossein Noruzoliaee for his invaluable supervision in scientific research and encouraging me with infinite patience and relentless support instead of his limitations to complete my research work. He has been a great mentor for me. It would be wrong if I do not mention respected and well known data scientists, Mr. Andrew Ng and Mr. Lex Fridmen because their works and guidelines motivated me to dig deeper into Artificial Intelligence. Being one of the advanced and latest algorithms, deep learning was not that easy to cope up with. But my advisor was patient and had believe in me which motivated me more to learn and apply new technologies into transportation engineering quickly. I am also thankful to my friend Sabeel Al Ferdous for helping me out when no one was there. I cannot defy the help from all Bangladeshi Students and Community living in the Rio Grande Valley during my research period. I should mention especially my roommates Md. Wasif Zaman, Apu Deb, Md Mashfiqur Rahman and Mohammad Anis who made me feel a new place like home, living 8,500 miles away from family. Without their constant support and motivation I could not finish my work smoothly. I would also like to mention two of my closest respected senior brothers, Abdullah Al Masum & Aminur Rashid Chowdhury who guided and motivated me to not only do research but also to enrich my personality and goals. I should particularly express my gratitude to all the respected professors of the Civil Engineering Department of UTRGV, especially to Dr. Ho and Dr. Kim for their immense support and guidance throughout the last two years. My research motivation mostly came from my supervisor Dr. Noruzoliaee

who was able to draw a big picture of future transportation engineering. His given flexibility made me think out of the box and utilize most of my potential.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Headway is one of the most significant traffic parameters, used in different transportation aspects like traffic flow theory and infrastructure decision-making policies. Headway can be represented in two ways. One is headway distance (length) and another one is time headway (time, usually in seconds). Time headway refers to the time needed for a vehicle to pass its front vehicle completely where both vehicles are moving in the same lane of a road segment. To be precise, time headway can be defined as ‘‘the time, in seconds, between two successive vehicles as they pass a point on the roadway, measured from the same common feature of both vehicles.’’[73]. The same common feature refers to the rear to rear bump or front to front bump calculation. The time headway of a subject vehicle can be calculated by taking the distance between the subject and the front vehicle and dividing it by the instantaneous velocity of the subject vehicle.

The distribution of the vehicle time headway also represents the overall scenario and the positions of the vehicles explaining whether the road is congested or a freeway. Thus time headway has a remarkable effect on traffic flow and vice-versa. Traffic flow is a macroscopic flow parameter. Being an influence on velocity, density, and traffic flow, time headway is considered as a key parameter for calculating highway capacity, safety, and level of service. Though time headway has effects on macroscopic parameters, it has vast microscopic features that explains the microscopic vehicle to vehicle relationships as well. When congestion arises, the inter-vehicle distance gets shorter which occurs heterogeneity in time headway distribution. At congestion drivers mostly tend to follow the front vehicle. This following behavior can be different for different types of vehicles as well. This kind of model is called the ‘car-following’ model where the vehicle to vehicle relationships are explored. Car-following models are the link between the macroscopic and microscopic models which can explain both micro and macro effects. The driver’s behavior is thus very necessary to analyze for a complete headway distribution model because time headway is stochastic and changes frequently.

Time headway is not constant for many reasons including stochastic actions of the drivers, difference in vehicle performances, lane changing behaviors of the drivers, etc. Evaluating the behavior of the drivers can be a good feature for making planning and safety policies. The main causes of rear-end crashes, which typically constitute approximately 30% of all police-reported crashes (National Highway Traffic Safety Administration, 1999), are driver inattention and following too closely, alone or in combination [73]. This reflects the driver’s inadequacy to maintain a safer headway and misjudgment. Appropriate time headway prediction can also make the driver aware of the surroundings through the vehicle to vehicle and vehicle to infrastructure communications. Thus time headway is a key parameter for predicting crashes and making safety protocols.

When vehicles are in a platoon maintaining a decreased time headway can increase the capacity of the road. Researchers are using headway models to predict the minimum safety distance within the vehicles which can be also reproductive in researches including connected autonomous vehicles. [74]. Time headway has been studied over a long period. As traffic scenarios are constantly changing through rapid changes in the economy, population, mobility, technologies, etc, the distribution and modeling of time headway have also changed over time. The availability of modern technologies and heterogeneity in traffic leads us to a better understanding of time headway. For example, in ancient times there were not too many car owners like today. Infrastructures were much lesser than it is now. Type of the vehicles or the lane changing behaviors were not considered in models of those times. Sara et al. (2014) explained the effect of different types of vehicles on headway distribution analyzing different mathematical models for different sets of vehicles [20]. But an investigation on choosing lanes and lane changing behaviors was not explained. Guo et al. (2016) filled the void by introducing headway distribution models for a multi-lane freeway [16]. They came up with different distribution models for different scenarios. Several studies have implemented separate headway models for different traffic flows.

Researchers are trying to generate a single model that describes and predicts time headway in various traffic conditions regardless of the place and type of road. Recently developed technologies and artificial intelligence implemented models are overtaking the previous probabilistic and mathematical models. Because these models can describe travel behavior and predict very accurately. Different machine learning and deep learning technologies have been already introduced in different aspects of ITS (Intelligent Transportation System), modern transportation planning, and also infrastructure asset management. Deep neural networks can perform behavior analysis more precisely, taking different features into action. *Long Short-Term Memory (LSTM)* is a deep learning algorithm that is a modified version of *Recurrent Neural Network (RNN)* which has an excellent performance in time series forecasting. LSTM has been already implemented in predicting different traffic parameters such as speed, travel time, aircraft boarding time, etc (Table 2). It has been found to be one of the most accurate prediction models for describing car-following and lane-changing behaviors. But it is yet to be implemented on traffic time headway prediction. As time headway is constantly changing and highly correlated with the driver's behavior, to build a single model describing all the scenarios altogether, a deep learning approach has been made in this study.

This study explains the complex relationship of traffic characteristics, vehicle shape & class, lane usage with traffic time headway. Historical time headway and actions of the drivers in different traffic conditions have been trained by the model. The model learns from past trajectory history and predicts accordingly. A safety constraint of 1s in time headway has been taken into account so that the model can be used in safety analysis as well. The NGSIM data of highway I-80 and US-101 have been used in our model. We trained our model with the trajectory dataset of highway I-80 and further validated with the trajectory dataset of US-101. We chose highway I-80 as our training dataset because the data includes a lot of heterogeneity and a high ranged of time headways (including very small and very high headway). Thus the model can learn more sophisticated scenarios in a mixed traffic condition and predict accurately. Chapter 2 describes the literature about headway models and the use of AI in ITS. The methodology behind the model is explained in the 3rd chapter. A vast analysis of data and time headway has been shown in chapter 4 through distributions and graphs. Data analysis include both vehicle type wise and lane wise analysis explaining the effects of other important features on headway. The summary of the model parameters with the detailed implementation and validation process is described in the 5th chapter. Results and further analysis have been visualized in chapter 6 where a comparison of current and past headway models has been also shown. The study is summarized and shown how it can be useful in different applications of ITS as a further recommendation in chapter 7.

CHAPTER II

LITERATURE REVIEW

2.1. Traffic Time Headway

The headway between two successive vehicles is considered as a random variable which is one of the main variables in traffic flow theory. Time headway is the time elapsed between two consecutive vehicles arriving at a measurement point of a road section. As the vehicle, driver, road, and environmental factors vary even in the same roadway network, this arrival of vehicles is considered as a stochastic variable. Headway distribution model being the basis of the traffic flow modeling and microscopic simulation, the distribution of this variable is being used in many studies including qualitative measurement of traffic (free flow, congested, etc.) on a given road under given conditions; quantitative analysis of road capacity (the ability of a road to service vehicles); traffic safety analysis and generating traffic simulation for analysis and prediction. (Tohbi et al., 2018).

There have been a lot of works and findings regarding the modeling of traffic headway using various tools and methods. We have divided these models into two parts based on their methodology (probabilistic models & use of artificial intelligence).

2.2. Headway models

2.2.1. Mathematical/ probabilistic models

Several authors have proposed many sophisticated models for different distributions which were accumulated by Faderico Pascucci in his work [9]. In fact, the very first study of headway distribution for low traffic (uninterrupted flows) is dated back to the 1930s (Adams, 1936) [1] where *Gaussian distribution* and *Negative exponential distribution* methods were introduced. In 1966, Greenburg described time headways applying *Shifted exponential distribution* and *Shifted log-normal distribution* which described the right-skewness of headway distribution [2]. The skewness of distribution represents the preference of drivers to maintain a short headway than a large headway. As traffic flow was very low, vehicles were considered to be running at a free-flow speed which eliminated the fact of interaction between vehicles in earlier theories. Vehicle arriving rate for a road segment was considered as an important feature in those previous studies which was strongly correlated with the headway distribution.

As roads were becoming congested day by day, researchers began to modify their models to complex models, trying to fit the headway distribution precisely. In 1976, Tolle et al. came up with composite exponential, Pearson Type III, and log-normal distributions which represented the best results for headway distribution in a wide range of traffic volume [3]. Pearson type III model is the most suitable time headway mathematical distribution for the intermediate headway state, including both random and constant boundary conditions. Some common Pearson's distributions like *Gamma, Erlang, negative exponential, and shift-negative exponential probability distributions* are used in some studies (Luttinen, 1999) [4]. In 1975, Cowan developed an important new model named the *M3 headway distribution model* which was used later in different traffic applications (for example, modeling roundabout flow by Hagring in 1996) [5]. Many researchers thought that the lognormal model should be used in analyzing the traffic flow in the freeway, while Mr. Luttinen in 1999 found the M3 distribution model as a better choice for analyzing traffic practically, especially for larger vehicle headways G[4].

Many factors such as traffic volume, the proportion of heavy vehicles, lane position, road structure, time of the day, and weather conditions were not considered in earlier studies which are key features for analyzing and modeling headway distribution [6]. Making lane numbers an important factor, Mei and Bullen measured the headway for high traffic flow in a four-lane highway using different statistical distribution and found the *lognormal distribution* (with a shift of 0.3 or 0.4 seconds) was the best fit for the time headways in high traffic volumes. (Mei M and Bullen, 1993) [7]. In the year 2000, Al-Ghamdi conducted research on different vehicle distribution under different flow rates and ended up finding *negative-exponential, shift-exponential, and Erlang distributions* are the optimal models for low, medium, and high levels of flow, respectively [8]. Before earlier 1990, researchers proposed their models without considering vehicle speed which was later proved to be one of the most influencing factors for headway distribution (Krbálek et al., 2001; Nishinari et al. 2003) [11, 12].

Another study of the headway distribution was investigated for a four-lane divided urban arterial in Chennai City in India [10] by Arasan in 2003 where *negative exponential distribution* was found to be best fitted for modeling headways at different lanes and over the entire range of traffic flows. To investigate headway distribution in different types of sections of a road, In 2006, Bham and Ancha analyzed the headway distribution in three sections where (i) a ramp merges, (ii) a lane drops, and (iii) a ramp weaving section and found *shifted lognormal distribution* accurately fits for all studied areas [13]. For blessings of technology, road-side video cameras, drones, on-vehicle radars, laser scanners, etc. have been used in the data collection process nowadays which enables us to further investigate both microscopic and macroscopic analysis quite accurately. Chen et al. (2010) carefully studying the headway data retrieved from Next-Generation Simulation (NGSIM) trajectory datasets (NGSIM, 2006) explained the relationship between headway and speed as well as traffic flow and speed in both macroscopic and microscopic point of view by a Markov model [19]. It certainly became important to investigate the headway distribution in different lanes. So, Zwahlen et al. analyzed the cumulative headway distributions at different traffic flows at each lane in Ohio freeways in the U.S [14] and interestingly found that for hourly traffic flow, the distributions at each lane are almost the same in shape.

As the world changes with time, there has been an increase in different types of vehicles on a road. Ye and Zhang analyzed the headway distribution between two different types of leading and following vehicles [15]. To dig deeper, in 2016 Kong and Guo found the optimal distribution considering six common distribution models (lognormal, gamma, exponential, normal, inverse Gaussian, and Erlang) to fit the Car-Car (C-C), Car-Truck (C-T), and Truck-Truck (T-T) headway types which resulted that, *the lognormal mode*l is suitable for C-C and T-T type headway distribution and *inverse Gaussian* is accepted by the C-T headway type. Also found different traffic flow rates, percentage of trucks, and lane positions have a great impact on each distribution model [16]. The time headways of preceding and following vehicles for different types of vehicles (heavy & passenger) were analyzed for an urban highway at different traffic flow rates during the congestion period using the trajectory data for a highway section in California: Berkeley Highway, I-80. The *log-normal distribution model* with different shifting values was found well fitted for different types of vehicles (both preceding and following) [20].

For being slow in movement, greater in length, and safety concerns, both the preceding and following headways of the heavy vehicles were greater than passenger cars. A study on mixed traffic, however, indicates that headway between two vehicles also depends on the length of the lead vehicle [17]. R. Roy and P. Saha. conducted a field study on two-way lane highways in India having a heterogenic traffic condition in 2018 [18]. They evaluated four distribution functions (*log-logistic, lognormal, Pearson 5, and Pearson 6*) while modeling the headway data and proposed that, *Log-logistic distribution* is an appropriate model for moderate flow whereas, *Pearson 5* satisfies well at congested flow.

2.2.2. Artificial Intelligence

In recent years, there has been a huge usage of Artificial Intelligence in several sectors both in engineering and the general field. For the blessing of technology, big data can be collected through several techs. One of the promising facts of AI methods is, it enables computers to sort through large datasets which makes the predictions more accurate and faster. Though there has been the usage of AI methods in traffic forecasting in some studies, limited work has been done on predicting traffic headway.

There is a complex inter-relationship between individual driver behaviors, vehicle characteristics, and traffic conditions which makes the headway prediction for an individual vehicle more challenging (Lee and Chen, 1986). Tong and Hung used feed-forward *neural network* to model vehicle discharge headway at the signalized intersection where they used eight variables concluding lane width and position, queue position & vehicle type of subject and preceding vehicles, etc. in 2000 [22]. Also, many researchers proposed the arriving time prediction models of the bus transit system, by modeling the bus headway using different machine learning algorithms (Support Vector Machine, K-Nearest-Neighbor, Kalman Filtering, Artificial Neural Network, Time Series and Hybrid model, etc.). For example, Chien et al. [25] proposed a headway-based microscopic simulation model, both link-based and stop-based *ANN models* for bus arrival time prediction. They used historical arrival and departure times, travel speed, traffic volume, as well as dwell time as inputs showing a superior prediction performance. Yu et al. [23] proposed an *SVM (Support Vector Machine)* model to predict bus arrival time considering segment-level travel time (current and next segment), weather condition, and headway based on GPS data as inputs. Later in 2011, they enhanced the prediction model by considering data from multiple routes at the same time, which results in higher accuracy than using single-route information [24]. To reduce operating cost and waiting time for passengers, Recently, Yu et al. (2016) established a *Least Squares Support Vector Machine regression* to predict the headway irregularity to detect bus bunching. They used transit smart cards as their data which included historical headway, travel time, and passenger demands. [26].

2.3 Artificial Intelligence (AI) in ITS

AI reflects the natural evolution of technology as increased computing power enables computers to sort through large data sets to identify patterns and predict more accurately and time efficiently. Machine learning and deep learning are the subsets of Artificial Intelligence. The goal of machine learning generally is not only to learn from the experience for specific tasks but also to further analyze and predict that can be understood and utilized by people. The performance is considered to be increased with experience. Two of the most widely adopted machine learning methods are *supervised learning* and *unsupervised learning*. Supervised learning deals with labeled data (for example, regression analysis) whereas unsupervised learning is used in unlabeled data to find the shape and structure of the data (classification problems). [27]. Being a subset of machine learning, deep learning performs the same task using the training data that should not necessarily be structured or labeled. Deep learning enables computers to learn on their own by training historical data.

2.3.1. Machine learning

Sababa et al. in 2016 estimated average daily traffic (AADT) and missing hourly volume using machine learning algorithms where a comparison between *artificial neural network (ANN)* and *support vector regression (SVR)* revealed that SVR functions better than ANN in AADT estimation for different functional classes of roadways [28]. Laha et al. used machine learning algorithms in multivariate multiple regression, spherical-spherical regression, and randomized *spherical K-NN regression* to predict real-time location with taxi-GPS data streams. The Multivariate multiple regression method has the best performance in terms of prediction accuracy [29]. Another prediction model (supervised learning model) with added behavioral and physiological features was presented by Ba et al. in 2017 to predict vehicle crashes which is the key component of the Vehicle Collision Avoidance System (VCAS) [30].

Besides predicting using regression, there have been some works on classification and clustering as well as using machine learning. To further improve the accuracy of traffic classifiers and reduce the cost, Zhao et al. (2019) proposed a feature selection algorithm based on Machine Learning which provided internet traffic classification and application identification associated with network traffic. Hasnat et al. (2018) came up with an ensemble machine learning classification technique to identify tourists and analyze the patterns of destination choices of tourists from location-based social media data [31].

Machine learning has been widely used to predict traffic flow, volume, and vehicle velocity over recent years. Cheng et al. (2017) used *chaos theory* and *support vector machine* for traffic flow prediction where the stochastic characteristics of traffic flow associated with the speed, occupancy, and flow are identified. The support vector regression (SVR) model was designed to predict the traffic flow which had better performance for the short-term traffic flow prediction in terms of accuracy and timeliness [32]. Another short-term traffic forecasting model was introduced by Guo et al. (2018) which had a fusion-based framework (which combines individual predictors like, Machine Learning, Neural Networks (NN), Support Vector Regression (SVR), and Random Forests (RF)) under different traffic conditions [33]. A tailored machine learning approach for urban transport network flow estimation was proposed by Liu et al. (2019) which aims to estimate traffic flow on a single link extracting Spatio-temporal traffic features [34]. Li et al. (2019) predicted traffic speed based on a deep feature fusion model where different AI methods were implemented and compared. The most validated result was obtained when the *deep feature fusion model* and *support vector regression* were jointly applied [35]. A *hybrid genetic algorithm* (a new soft unsupervised classification method) and machine learning (backpropagation network (BPN)) were combined to predict the rental demand for a bicycle-sharing transportation system to increase profits and also to improve user satisfaction. (Guo et al., 2019) [36].

2.3.2. Deep learning

Deep learning has been successfully employed in computer vision, speech & audio recognition, inspection process, and natural language processing which eventually provoked a storm in ITS (Intelligent Transportation Systems). This new learning technique replaced traditional ML models in many applications and the outlook of ITS is being reshaped. DL has been used in different transportation engineering applications; for example, visual recognition tasks, traffic flow prediction (TFP), traffic speed prediction (TSP), travel time prediction (TTP), Car-following models, and Miscellaneous tasks by introducing new learning models like *Deep Neural Network (MLP, DBN, SAE), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Graph Convolutional Network (GraphCNN), Long Short-Term Memory (LSTM)*, etc.

There have been a lot of works regarding traffic flow prediction using various deep learning techniques achieving promising results. Some useful recent works have been accumulated in table 1.

Table 1. Recent deep learning techniques used in ITS are listed in this table.

Author and Year	**Model used**	**Output**
Lv et al. (2015)	Stacked Autoencoder (*SAE Model*)	Traffic Flow Prediction
Miyajima et al. (2016)	Hidden Markov models (*HMM*s) and deep learning	Driver-Behavior Modeling
Li et al. (2019)	Deep feature fusion model and support vector regression is jointly applied	Traffic speed prediction
Wu et al. (2018)	A Deep Neural Network (DNN)-Bidirectional Texture Function (*DNN-BTF*)	Traffic flow prediction
Duan et al. (2016)	LSTM	Travel time prediction
Liu et al. (2019)	Deep Passenger Flow (DeepPF)	Metro passenger flow prediction
Hao et al. (2019)	End-to-end deep learning framework	Short-term passenger flow prediction
Zhang et al. (2019)	Graph convolutional sequence-to-sequence model (AGC-Seq2Seq).	Multistep speed prediction
Gu et al. (2019)	Combination of long short-term memory (LSTM) neural network and the gated recurrent unit (GRU) neural network	Lane-level traffic speeds predictions
Yang et al. (2019)	Graph-Convolutional Neural Networks (GCNN)	Predicting block-level parking occupancy in real-time (30 mins advance)
Lee et al. (2019)	Deep learning	Modeling car-following behaviors on a multi-lane motorway
Wang et al. (2019)	Bidirectional long short-term memory neural	Better traffic speed prediction at city
Xu et al. (2018)	A deep learning method	Dynamic demand forecasting model for station-free bike sharing

Zhang et al. (2018)	Deep Belief Network (DBN) and Long Short-Term Memory (LSTM)	Detects traffic accidents with social media.
Polson et al. (2017)	Deep learning	Short-time traffic flows prediction.
Liu et al. (2017)	A hybrid deep network of unsupervised SAE and supervised DNN	Passenger flow prediction
Ke et al. (2017)	The fusion convolutional long short-term memory network (FCL-Net)	Short-term forecasting of passenger demand
Do et al. (2019)	Deep learning-based traffic flow predictor with spatial and temporal attention (STANN)	Traffic flow prediction
Dabiri et al. (2018)	Convolutional Neural Network (CNN)	Predict travel modes based on raw GPS trajectories
Simoncini et al (2018)	Long Short-Term Memory (LSTM) recurrent neural networks	Vehicle classification

Some works may focus on predicting the traffic flow of the next several time intervals from (t + 1) to (t + n) as well where 't' denotes the current time. In practice, traffic flow forecasts are divided into three groups according to the length of projection time, such as short-term (5–30 min), medium-term (30–60 min), and long-term (over an hour) (Yu et al., 2017).

Lv et al. (2015) predicted traffic flow using a deep learning approach named *Sparse Autoencoder (SAE)* which proved to be superior in performance to probabilistic and machine learning models [37]. Traffic flow being nonlinear in nature, Polson and Sokolov (2017) used deep learning architecture to capture these nonlinear Spatio-temporal effects which outperform the linear and single-layer neural network models and predicted short-time traffic flow. They also explained this non-linearity because of anonymous transitions between free flow and breakdown [38]. A hybrid deep learning-based traffic flow prediction method (*DNN-BTF*) has been proposed by Wu et al. (2018) which has been used in several studies. *Recurrent Neural Network (RNN)* was implemented to capture the temporal feature and *Convolutional Neural Network (CNN)* was used to extract the spatial feature of the historical data of traffic flow and velocity. [39].

Usually, limited historical data (for example, 15 mins, 30 mins, 1 hour) is being used for training in deep learning models. It may fail to capture the trending nature of that particular road segment. Dai et al. (2018) introduced a detrending-based many-to-many traffic prediction model called *DeepTrend 2.0* where *graph convolutional neural networks (GCNN)* were used to extract spatial dependency [40]. Deep learning techniques are not only being used to predict traffic flow but also in passenger flow prediction. Liu et al. (2017) proposed a hybrid deep network of *Stack Autoencoder and supervised DNN* to predict the passenger flow. The input features included real-time passenger flow and average historical passenger flow. These features were stacked together after encoding. Autoencoder extracts the non-linear relationship. *DNN* was then further used for prediction. The proposed model provided high accuracy predicting in different BRT stations in different traffic conditions. [41]. There had been a lack of multistep traffic prediction. Recently, Zhang et al. (2019) implemented the attention graph convolutional network in a sequence-to-sequence framework (*AGC-Seq2Seq*) to not only capture the stochastic Spatio-temporal characteristics of traffic flow but also produce multistep traffic prediction. [42]. *Convolutional Neural Networks* (CNNs) have been successfully applied to exploit the spatial correlations of the road network in several studies as it has a great visual ability. [43, 44]. A combination of *Convolutional Neural Network (CNN)* and *Recurrent Neural Network (RNN)* has been used in many studies as these networks work well in capturing different aspects (spatial or temporal) of the data. Wang et al. [45] and Wu et al. [39], proposed the combination model to predict the traffic speed and volume, respectively. CNN has also been used in predicting travel modes used by passengers based on only raw GPS trajectories. Five travel modes (by walk, bike, bus, driving, and train) were taken into consideration where CNN detected the characteristics of moving vehicles [46]. *Deep reinforcement learning* (RL) models are good at decision making. It has been used in various fields of ITS. For example, in route prediction problems in stochastic and time-dependent (STD) network (Mao et al. (2018) [47]), also in controlling passenger inflow of urban rail transit system in congested and peak hours [48].

Recurrent neural network (RNN) is one of the best network models to process the time-sequence data themselves which were structured to learn temporal correlations from sequences of data and have been applied to predict traffic flow in several studies (49–53). *Long Short-Term Memory (LSTM)* units are an element of RNNs that can correlate both spatial and temporal features of historical time series data across different timesteps. LSTM has been widely used to model car following and lane changing behavior in several studies recently [54-58]. Duan et al. (2016) used LSTM in travel time prediction [60]. LSTM is also being used to predict traffic speed in some studies [60, 61] including lane-level speed prediction [61]. Providing greater accuracy in time series prediction, LSTM has been used widely in other different fields of ITS, such as traffic flow [63], bike-sharing demand model [64], route prediction, boarding time prediction[65], etc. Some of the recent works implementing LSTM in ITS are listed in table-2.

Table 2. Recent works implementing LSTM in different studies of ITS

Year	**Author**	**Output**
2016	Duan et al.	Travel time prediction
2017	Ke et al.	Short-term forecasting of passenger demand under an on-demand ride service platform.
2018	LI et al.	Missing value imputation
2018	Xu et al.	Demand forecasting model for station-free bike sharing
2018	Zhang et al.	Detecting traffic accidents with social media.
2018	Simoncini et al.	The categorization of the type of vehicles on a road network
2018	Huang et al.	A car-following model considering asymmetric driving behavior
2018	Rosberg et al.	Route and travel time prediction
2019	Gu et al.	Lane-level traffic speed prediction

2019	Zhang et al.	Modeling the Car following (CF) and Lane Changing (LC) behaviors.
2019	Yang et al.	Predicting block-level parking occupancy in real-time (30 mins advance)
2019	Wang et al.	Traffic speed prediction for urban transportation
2019	Schultz et al.	Aircraft boarding time prediction

Though LSTM has touched on various aspects of ITS, it has not been used to predict traffic time headway. In this study, the novel *Long-Short-Term Memory (LSTM)* has been introduced for time headway prediction with microscopic and mesoscopic analysis.

2.4. Contributing factors for headway estimation

As described earlier, various probabilistic, machine learning, and deep learning models have been implemented to analyze and predict headway. Each author used different contributing features to estimate headway. Selections of features have been changed over time due to innovative models and updated literature. Table-3 describes the selected features used for predicting headway over time.

Table 3. Contributing features for analyzing vehicle headway of a road network

Name of the Article	Author	Contributing Features
Evaluating the Time Headway Distributions in` Congested Highways	Sara et al., 2014	Traffic volume, Proportion of heavy vehicles, Lane position, Road structure, Time of the day and Weather condition
Drivers' Time Headway Characteristics and Factors Affecting Tailgating Crashes	Hassan et al, 2017, TRB	Drivers' factors (i.e., gender and nationality), Vehicle factor (i.e., vehicle type), Road and environment factors (i.e., road type, number of lanes and road surface condition).
Determinants of the following headway in congested traffic	Mark et. Al, 2009	Type of vehicle, Traffic flow, Road type, A distinct day-to-day variation in individual behavior

Determinants of time headway in staggered car following conditions	Das, et al., 2017	Distance headway, Preceding & subject vehicle speed, Road width
Headway distribution models of two-lane roads under mixed traffic conditions: a case study from India	Roy et al., 2018	Traffic flow, Length of the vehicle
Analysis of vehicle headway distribution on multi-lane freeway considering car–truck interaction	Guo et al., 2016	Traffic flow rate, Percentage of Truck, Traffic speed, Lane number, Vehicle type
Flow–headway Distribution Relationship: A Case Study of Yangon	San et al., 2019	Traffic flow rate, Number of lanes
NN model of vehicle discharge headway	Tong et al., 2000	Lane width, Lane position, Vehicle type of sub and preceding, Headway of preceding vehicle
Scheduling Combination and Headway Optimization of Bus Rapid Transit	Wey et al., 2008	Traffic flow, Traffic speed

From table-3 it can be understood that vehicle time headway depends mainly on three factors.

(i) Vehicle factors (vehicle type, width, length, velocity, etc.)

(ii) Driver's factors (driver behavior, gender, nationality, etc.)

(iii) Road & traffic factors (traffic flow, density, lane number, road width, time of the day, etc.)

In this study, we have selected eight features which include traffic flow rate, the velocity of the subject and preceding vehicle, acceleration of the subject vehicle, vehicle length, width & class, and lane number. These features were devided by the following four groups which includes both microscopic and macroscopi traffic parameters.

Vehicle information	1. Vehicle length 2. Vehicle width 3. Vehicle class
Vehicle movement	4. Vehicle acceleration 5. Velocity of the subject vehicle 6. Velocity of the preceding vehicle
Lane information	7. Lane number
Macroscopic parameter	8. Vehicle Flow rate

The inputs will be the historical data of these features for modeling traffic time headway.

CHAPTER III

METHODOLOGY

As Traffic time headway has both spatial and temporal correlations, traditional Artificial Neural Network (ANN) cannot perform adequately. On the other hand, RNN, especially the LSTM model has unique features to capture the trajectory and Spatio-temporal time series data very well. In this study LSTM neural network (LSTM NN) has been implemented to predict vehicle time headway. This chapter describes the framework of the LSTM network, it's usage and specialty, as well as the proposed model configuration in detail.

3.1. Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM NN)

RNN is a sequential prediction model that is one of the most used and valuable models of Artificial Intelligence. It is designed to take sequential data as inputs that can recognize the pattern and temporal correlations between the features of the data. Remembering the previous step, RNN can predict the next step quite successfully. The model can be visualized by several units (RNNs) connecting through their hidden layer (figure 1). From figure 1, each RNN cell has three units- input, output, and hidden unit. x_t, $h_{t,}$ and x_{t-1}, h_{t-1} are considered as input and output at time t and (t-1) respectfully. Each unit is interconnected by their hidden unit. This hidden unit can pass the information from the previous timestep to the next one. W_{in}, W_{out}, and W_{rec} are the weights of input, output, and hidden unit(s) respectively. So, RNN learns from the previous scenarios and predicts the next scenario h_{t+1} using x_{t+1} as input.

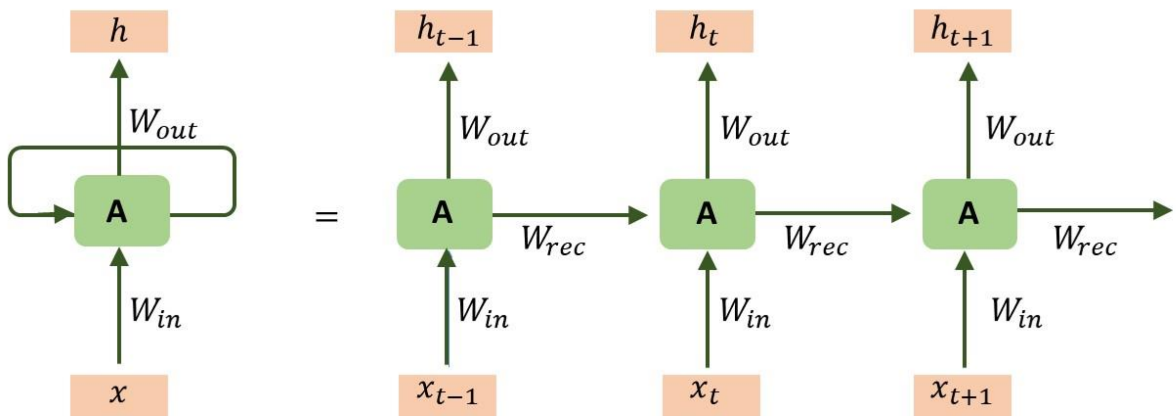


Figure 1. Recurrent Neural Network

Though RNN works well for time series data, it has some limitations. Being a memoryless model, RNN cannot remember much longer historical scenarios that may correlate with recent data. This problem is called a vanishing gradient problem. The lower the gradient, the more difficult and unlikely to find the optimal solution. To eradicate this limitation, a memory-based RNN model named *Long-Short-Time Memory* was introduced by Hochreiter and Schmidhuber (1997) which has the ability to remember and pass information from long periods of time [66]. The basic LSTM cell unit is shown in figure 2. Previous state condition C_{t-1}, hidden units from previous step h_{t-1}, and present data/ feature x_t are the inputs and h_t is the output which is feed forwarded to next LSTM cell.

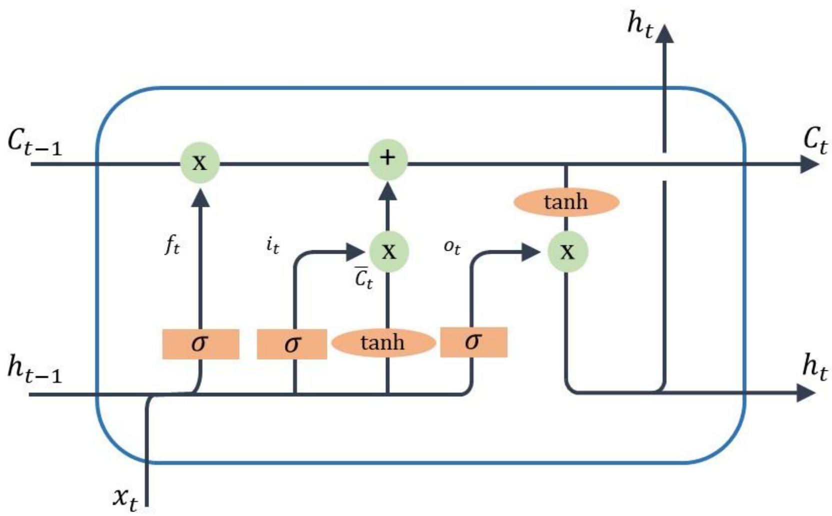


Figure 2. Unit LSTM cell

The main difference between RNN and LSTM is, LSTM includes the memory and forget cell which can consider new important information and ignore less important information. The prediction through LSTM depends on the previous cell state (C_{t-1}), previous hidden state (h_{t-1}), and current information/ input. (x_t).

The unique features of LSTM are:

(i) It can overcome previous critics of RNN resolving the vanishing gradient problem. LSTMs (and GRUs) can model long-term sequential dependencies.

(ii) As it has a forget gate and a memory gate, it can easily forget and remember data that are necessary for a prediction which enables to control the flow of information.

(iii) LSTMs are better for capturing the non-linear relationship between the data.

3.2. Framework of LSTM

Cell state and memory block are some of the unique features of LSTM. It represents the information passing from the previous time step which is connected with three gates and eventually generates a new cell state (C_t) which is again transferred to the next LSTM layer (figure 3) Thus, the state changes over time which has an impact on prediction.

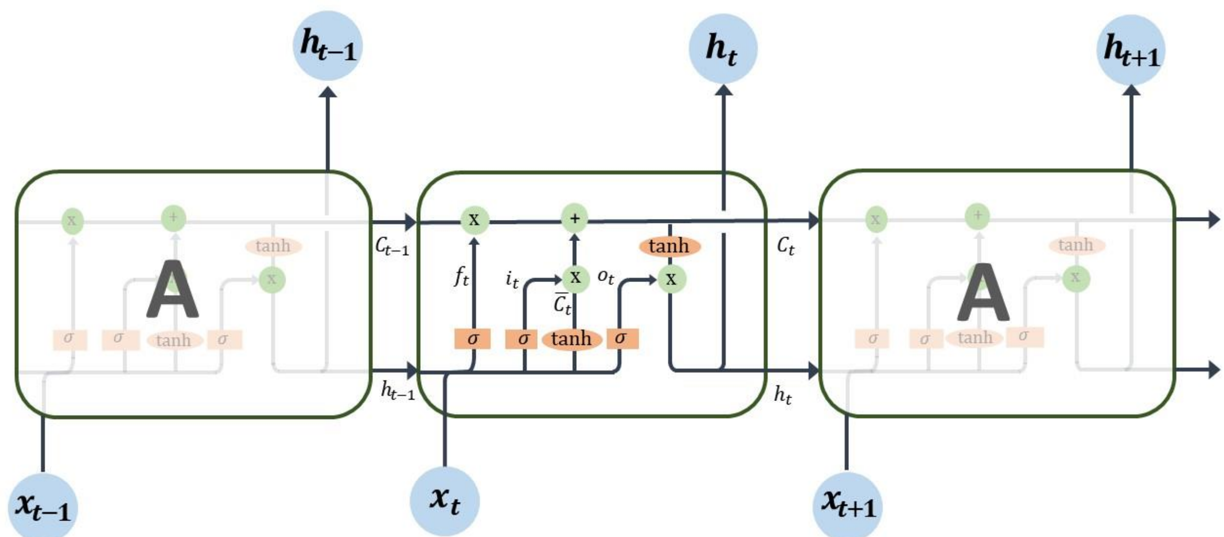


Figure 3. LSTM Neural Networks

3.2.1. Gates

The main framework of the LSTM network lies on three unique gates named forget gate (f_t), input gate (i_t), and output gate (o_t) over time interval t (figure 3). These gates have different purposes and capabilities.

(i) **Forget gate (f_t):** The hidden result from the previous state (h_{t-1}) and current information (input) (x_t) are jointly connected to cell state (C_{t-1}) by scaler product (·). This gate decides which information to pass to the cell state and avoids outdated information by forgetting them. Inputs (x) and hidden output (h) are imported as vectors. They may have different vector shapes from each other.

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f\right) \tag{1}$$

Here, W_{xf}, W_{hf}, W_{cf} are the forget gate weights of input, previous hidden unit, and state respectfully. b_f is the bias vector.

(ii) **Input gate (i_t):** In the previous gate, it was decided what information to forget. By this gate, new information can be added and replaced the outdated one. Previous hidden vector (h_{t-1}) and current input vector (x_t) are first concatenated by scaler product (·). $\bar{C_t}$ refers to the candidate new information vector. This new information is added to the state. State will be then updated to C_t from C_{t-1}.

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i\right) \tag{2}$$

$$\bar{C_t} = tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{3}$$

$$C_t = f_t C_{t-1} + i_t \bar{C}_t \tag{4}$$

W_{xi}, W_{hi}, W_{ci} are the input gate weights of input, previous hidden unit, and state vectors respectfully. b_i is the bias vector of the input gate.

(iii) Output gate (o_t): After updating the state what output is needed for a particular project can be decided by the output gate. Updated state (C_t) and the previous hidden vector (h_{t-1}) will be scaler multiplicated to decide which values to generate in output. h_t will be the final output which may be again connected to another LSTM cell.

$$o_t = \sigma\,(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + b_o) \tag{5}$$

$$h_t = o_t\, tanh(C_t) \tag{6}$$

W_{xo}, W_{ho}, W_{co} are the output gate weights of input, previous hidden unit, and state vectors respectfully. b_o is the bias vector of the output gate. h_t denotes the output of the current LSTM unit.

Sigma $\sigma(\cdot)$ refers to a standard logistic sigmoid function that can transfer the input values into the range of (0,1). Hyperbolic tangent function (*tanh*) squashes input values to output between -1 and 1, not 0. It overcomes the vanishing gradient problem. Both are defined as nonlinear activation functions

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

$$tanh(x) = \frac{e^{x} + e^{-x}}{e^{x} - e^{-x}} \tag{8}$$

3.2.2. Inputs and output

The time headway of a vehicle depends on several traffic parameters. Deep learning algorithms can learn through the past data about how different factors had effected headway and predict according to nature. In this study traffic flow (*q*), the velocity of the preceding (v_p) and subject vehicle (v_s), vehicle length (*l*), vehicle width (*w)*, vehicle acceleration *(a)*, vehicle class *(c)*, lane ID (*i*) are considered as features. The historical values of these features construct the input matrix which has been trained with the LSTM model.

Let x^t be the vector form of mentioned features of tth time. It can be written as follows:

$$x_t = [q^t, v_p{}^t, v_s{}^t, l^t, w^t, a^t, c^t\ i^t] \tag{9}$$

For example, s^t denotes the velocity of the subject vehicle at time t. The input matrix includes the scenario of the features before tth time.

Let Δt and Z respectfully be the timestep and the number of timesteps taken into this study where (Δt × Z) < t. The expression for the input matrix (X_t) is as follows:

$$X_t = [x_{t-Z\Delta t}, x_{t-(Z-1)\Delta t}, \ldots., x_{t-2\Delta t}, x_{t-\Delta t}] \tag{10}$$

The output of the model Y_t denotes the traffic time headway prediction of at the tth time step corresponding the input matrix X_t. The row numbers of both X_t and Y_t matrixes are the same.

3.2.3. The shape of the neural network

The number of layers, total neuron number, number of neurons in each layer combine the structure and shape of the model. Every model has one input and one output layer. It is not necessary to have a hidden layer or more than one hidden layer if the problem is simple and linear. Traffic time headway is a complex parameter that has non-linear relationships with its features. The factors have also a non-linear relationship between them. So, more than one hidden layer must be included.

Different studies used different numbers of neurons and layers which fit accordingly. Stathakis (2009) described how to choose layer and neuron numbers in neural network modeling. It can be seen from past studies that most of the studies performed trials and found the optimal neuron and layer numbers. Yang et al (2019) while implementing the LSTM model for prediction of parking occupancy used both two layers and three layers of structured separated LSTM structures to show the difference in results [67]. They used 1024 & 256 neurons for the two-layer LSTM structure and neuron configuration of 2048-512-128 for three layers. Wang et al. (2019) in his prediction over vehicle velocity used only one hidden layer of 512 neurons [61]. In another study of vehicle velocity prediction, Ma et al. (2015) showed that the prediction accuracy of the LSTM network is independent on the input time lag.

Some authors used different algorithms and analysis as well to automatically discover the optimal shape of the neural network [69]. For example, Xu et al. (2018) performed parameter tuning, and sensitivity analysis to show the relationship between time interval, number of nodes and layers, batch size [64]. They elaborately described how changing these parameters in between them have effects on Mean Average Percentage Error (MAPE) of the model. Huang et al. (2018) came up with the LSTM model having 8 layers and 32 neurons in each layer produces the best accuracy for describing car following and lane changing behavior [70].

In this study, several combinations of trials have been initiated. It is found that 2 hidden layers having 512 & 256 neurons generate the best accuracy for time headway prediction.

CHAPTER IV

DATA ANALYSIS

4.1. Data overview

Applying artificial intelligence using machine learning and deep learning algorithms to establish a microscopic simulation needs lots of versatile data with mixed scenarios so that machine can gather more experiences and predict accurately. The dataset we used, was collected from the Next Generation Simulation (NGSIM) program which was initiated by the United States Department of Transportation (US DOT) Federal Highway Administration (FHWA). They have collected several detailed, high-quality datasets. The proposed model has been implemented on vehicle trajectory data on eastbound I-80 in San Francisco and also validated with training a different set of trajectory data from another highway on southbound US 101 (also known as the Hollywood Freeway), Los Angeles, California. These vehicle trajectory data provide the precise location of each vehicle with detailed lane positions and longitudinal locations relative to other vehicles within the study area every one-tenth of a second. Data not only contains information on multimode and lane choices but also describes the behavior of travelers in different traffic conditions (for example, traffic congestion, freeway, etc.).

Trajectory data of I-80 was collected on April 13th, 2005 which contains both spatial and temporal information. The length of the study area was approximately 500 meters (1,640 feet) consisting of six freeway lanes, including a high-occupancy vehicle (HOV) lane and an on-ramp road. Seven synchronized digital video cameras mounted from the top of a 30-story building adjacent to the freeway recorded vehicles passing through the study area. A total of 45 minutes of data are available in the full dataset, segmented into three 15-minute periods: 4:00 p.m. to 4:15 p.m.; 5:00 p.m. to 5:15 p.m.; and 5:15 p.m. to 5:30 p.m. These segments represent the transition periods between uncongested and congested traffic conditions. We have used the total data (total 45mins) which describes the instantaneous movements of around 3000 vehicles.

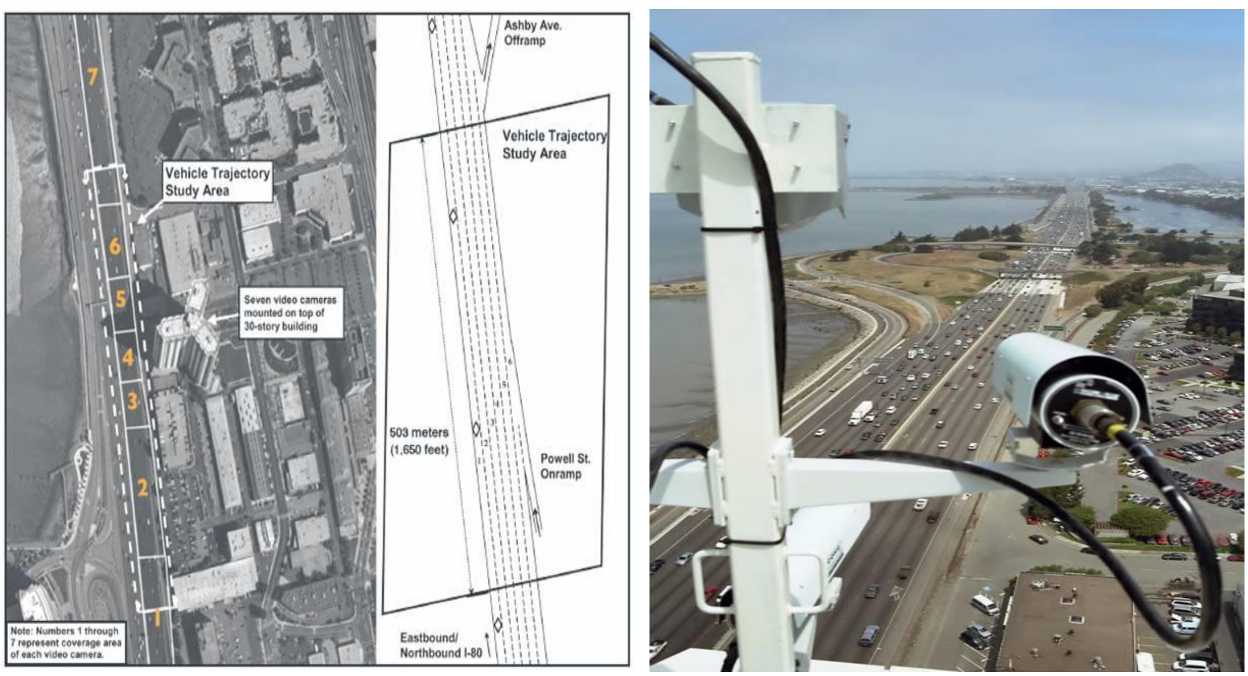


Figure 4. Aerial view of the subject area (left) and a digital video camera mounted on top of a building recording vehicle trajectory data of I-80 (right)

We selected data having time headway greater than 1 second. This is a safety constraint that was added so that the model can also be used in trajectory planning of vehicles such as connected and autonomous vehicles (CAVs).

4.2. Data preparation

Before feeding into the model data has to be preprocessed, making things easier for the machine to learn and analyze. As it is a panel data, some feature engineering techniques were also applied to extract spatial and temporal dependency. Traffic time headway, as a mesoscopic parameter of traffic engineering, can be described by both microscopic and macroscopic simulation.

4.2.1. Feature engineering

Traffic flow, defined as the number of vehicles passing a certain location on a highway per unit time, is one of the most important influencing macroscopic factors of time headway. Vehicle flow in every one-minute interval was calculated and used as a feature. Headway varies during different traffic flow and conditions. Flow rate also describes the building up congestions and how it can correlate with vehicle velocity and headway.

In addition, time headway of a subject vehicle heavily depends not only on its own velocity but also on the velocity of the preceding vehicle (The front vehicle). For example, if the velocity of the preceding vehicle is too high, it will certainly take more time than usual for a subject vehicle to pass that vehicle. So considering the velocity of the preceding vehicle as an important feature, it was calculated using the ID numbers of preceding vehicles in each 0.1s of the data. This microscopic feature also reflects the behavioral tendency of the driver whether the driver drives slow, fast, or overtakes regarding the velocity of the preceding vehicle at that time. This is how the macro and microscopic information with Spatio-temporal dependency were extracted from the raw data.

A dummy variable is a numerical variable taking only 0 or 1 that is used to represent subgroups or categories of the sample data. Two categorical variables in the dataset, representing the vehicle class (namely, cars, trucks, and motorcycles) and the lane number/position of the moving vehicle (seven lanes), have been transformed into a number of dummy variables. For example, three dummy variables are defined to represent the categorical variable associated with vehicle class. That is, one dummy variable is defined for each vehicle class (e.g., the dummy variable of cars takes 1 if the vehicle of interest is a car and takes 0 otherwise).

4.2.2. Preprocessing

Raw data has to be preprocessed into a useful and efficient format to promote the extraction of meaningful insights from the data before implementing the model. Data preprocessing has some necessary steps but it may vary depending on the dataset and the model to feed-forward into. As NGSIM data suffer from some noises it was properly cleaned first and then a lower bound constraint of 1s was set to time headway for safety. It means we used the data of vehicles having more than 1s time headway. Most of the machine learning and deep learning algorithms need categorical values to be encoded before feeding for training. Ordinal categorical values are ordered numeric values, for example, 0,1,2 numbers can be referred to good, better, and best respectively. As vehicle ID numbers are unique for each vehicle, it has to be treated as a nominal categorical column. For this reason, the column which represents vehicle ID number was encoded so that the machine can interpret these numbers as nominal categorical values rather than ordinal values.

Furthermore, to ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, raw data has to be properly scaled in a specific range before feeding it to the model. Data transformation is a crucial step in preprocessing which can be done through several techniques such as standardization, normalization, etc. This is also known as feature scaling. Normalization refers to scaling the data into a specific range (for example, [-1, 1] or, [0, 1], etc.). In this study, the data were normalized into zero to one value range ([0, 1]). This can be represented by the following equation.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{11}$$

Where X is the raw and X' is the scaled value of each feature.

4.2.3. Splitting and reframing

After finishing the preprocessing of the data, a total of 4,122,130 records of the whole dataset has been split into three groups for training, validation, and, testing purposes. 65% of the dataset has been used for training and 15% was used to validate the model. The rest of the data (20%) is used for prediction so that the testing data are unseen and new to the model. The summary of the dataset is represented in table 4.

Table 4 Data overview

Dataset	Number of samples	Mean (sec)	Standard Deviation	Minimum (sec)	Maximum (sec)	1st quartile (Q1) (sec)	2nd quartile (Q2) (sec)	3rd quartile (Q3) (sec)
Training	2,677,951	4.63	10.34	1.01	399.54	2.17	3.01	4.41
Validation	618,551	4.87	10.21	1.01	398.85	2.32	3.23	4.79
Testing	825,628	4.77	11.48	1.01	399.72	2.09	2.88	4.23

As it is a time series panel data, the data was sorted properly reshaping into a multi-index dataset. A multi-index dataset refers to data having more than one index. Here, we have chosen time and the vehicle ID columns as two indexes. By doing this, data having the same vehicle IDs were stacked together by the index ‘Vehicle_ID’ and after that, it was sorted by ‘Global_Time’, which is another index. Data of each vehicle was split into training, validation, and test sets sequentially by time. In this way, the model can learn easily and predict more accurately.

As it is a supervised problem, the dataset contains information about both the independent variables ($x_1, x_2, x_3, \ldots$) which are also called features and the output/dependent variable which is also known as labels *(y)*. As mentioned in chapter 2, we selected the contributing factors of time headway such as vehicle velocity, flow rate, type and shape of the vehicles, lane numbers, etc. as independent variables and time headway as the dependent variable. The goal of the model is to map between input features ($x_1, x_2, x_3, \ldots$) and the output (y) finding correlations and learning from the historical data to predict over unseen features.

Training the Long-Short-Term-Memory networks (LSTM NN) requires each data reshaped as three-dimensional tensors. To this end, we first converted the data frame into a supervised stationary structure by lagging the feature columns and shifting the target column

(Time Headway) l^{th} step into the future. The lagged structure of the whole data frame can be written as follows:

$$\begin{bmatrix} {x_1}^{t-l} & {x_2}^{t-l} & {x_3}^{t-l} & & {x_n}^{t-l} & {x_1}^{t-(l-1)} & x_2^{t-(l-1)} & & x_n^{t-(l-1)} & x_1^{t-(l-(l-1))} & x_2^{t-(l-(l-1))} & & x_n^{t-(l-(l-1))} & x_n^t \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ {x_1}^{T-l} & {x_2}^{T-l} & {x_3}^{T-l} & & {x_n}^{T-l} & {x_1}^{T-(l-1)} & x_2^{T-(l-1)} & & x_n^{T-(l-1)} & x_1^{T-(l-(l-1))} & x_2^{T-(l-(l-1))} & & x_n^{T-(l-(l-1))} & x_n^T \end{bmatrix} \tag{12}$$

To illustrate, let us assume that *x* represents all the variables present in the dataset *(x_1, x_2,…, x_n)*, 'n' being the number of variables, where we want to predict the variable x_n and the rest of the variables are independent time series features. Here, *l* is the lag length which is one of the hyperparameters that has to be trialed to get the optimum lag length. It varies within different types of model implementation and data structure. Lag length indicates the historical data that the model would look back to predict the future. Generally, complex relationships between the variables need greater lag length to be understood accurately by the machine. For this reason, some trials and errors were performed to see how many lag length is optimum for the model. We have concluded that the model works better with a single lag unit (*l=1*). While using different lag numbers (*l=3,5,10,60,80)*, it was found that accuracy did not get higher than using a single lag length, and also it was very time-consuming. T is the total time frame of the study which is total 45 minutes (4:00 p.m. to 4:15 p.m.; 5:00 p.m. to 5:15 p.m.; and 5:15 p.m. to 5:30 p.m.) having a minimum timestep of 0.1s. From this structure, the n^{th} variable, x_n is representing the dependent variable (time headway) which was used as the output *(y).* The rest of the variables with their lagged time frame including lagged time headways were considered as the input of the model.

$$X = \begin{bmatrix} {x_1}^{t-1} & {x_2}^{t-1} & {x_3}^{t-1} & & {x_n}^{t-1} & {x_1}^{t} & x_2^t & & x_{(n-1)}^t \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ {x_1}^{T-1} & {x_2}^{T-1} & {x_3}^{T-1} & & {x_n}^{T-1} & {x_1}^{T} & x_2^T & & x_{(n-1)}^T \end{bmatrix} \tag{13}$$

Here, X is the input matrix consisting of the historical and present values of the independent variables throughout time *T*, including the historical values of the dependent variable (time headway). For example, if *t* denotes as present time, ${x_1}^{t-1}$ defines as the value of the variable x_1 at the previous time step.

It means the model will be experiencing the different headways in the previous and the present states and predict accordingly to the next step. So, the output or label matrix can be shown as:

$$y = [x_n^t, x_n^{t+1} \ldots \; x_n^T] \tag{14}$$

Predicting farther steps can be transited from one to many. Which means more the one step ahead prediction to the future. Although the accuracy might get lower while predicting more than one step. In this study, a single-step prediction has been performed.

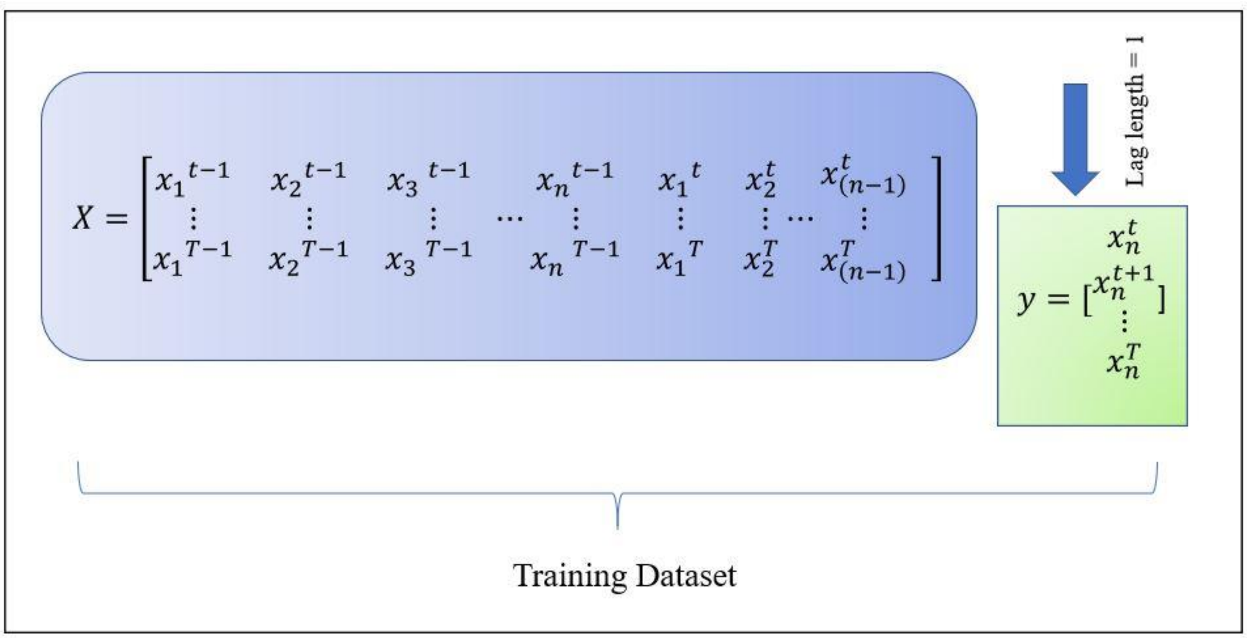


Figure 5. Spliting & reframing the dataset

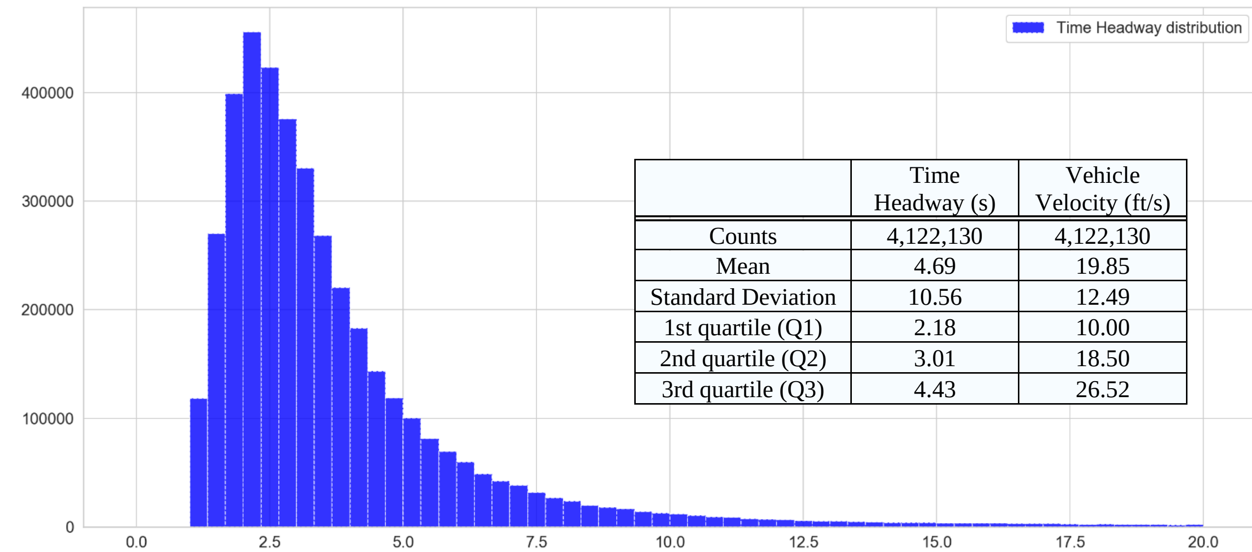
Figure 5 represents the summarized version of reshaping and splitting the training data into features (X) and target data (Y) which is time headway.

After reframing the data frame into the stationary structure, the values of the three split datasets (training, validation, and, testing) were reshaped into 3-D Tensors like (X, Y, Z) where X, Y, and Z represent the total length of the data, time steps (lag number), and, number of features, respectively, where the input of the model was shaped like (Y, Z). It is because the Long Short-Term Memory networks require the input shapes to be in 3-D tensors. For example, our training data consisted of 2,677,951 numbers of data having 31 features. These 31 features are cumulative of lagged and present data of each variable after the final reformation of data. As we considered lag length to be 1, the final 3-D shape of the input data was (2677950, 1, 31); where the shape (historical step, feature numbers) = (1, 31) was fed into the model as input for every 2677950 times. It means the model will experience the headways of lagged variables (31 features) in every step, and learn from each scenario of each vehicle ID to predict the next time step. This is where the LSTM network can be very powerful as it can memorize the present and previous states to predict the next state.

4.3 Time Headway Analysis

The distribution of time headway has been studied over a long period introducing various mathematical models. Varieties of study regarding this distribution have been established using different traffic flow and scenarios. Most of the ancient studies indicated *negative exponential distribution* as a well fit probabilistic model for describing time headway distribution. It is because at that time traffic flow was low and beyond today’s high and mixed traffic scenarios. To establish a boundary, Al-Ghamdi stated that the *negative exponential distribution* describes the time headways well under low traffic conditions (<400 vehicles/hour) whereas *shifted* *exponential* and *gamma distributions* were found to be well fit in moderate traffic flow conditions (400~1200 vehicles/hour) [8]. Similar comments were proposed by Kumar and Rao where *negative exponential distribution* was found suitable for low and moderate traffic flow [72]. But headway distribution in a congested flow generally does not follow exponential distributions. They also formed a relationship between the vehicle platooning and mean time headway which stated that vehicles are considered to be in platoon if the time headway is less than 2 seconds. Platooning of the vehicles leads to a car-following situation. Mei and Bullen described this scenario very well and proposed *log-normal distribution* to be the best fit for congested traffic flow [8]. According to them, although individual time headways are always changing over time, as vehicles enter into a congestion state, drivers tend to maintain a specific range of headway. This condition arises the car-following situation where drivers are mostly biased on their front vehicle's movement.

The difference in time headways of the vehicle at congested states being very small leads to a *shifted/ skewed log-normal distribution*. From figure 6, it can be seen that the velocity is very low in the 3rd quartile (around 17 mph), which leads to a congested state. As congested situations are stochastic in mixed traffic flows, the *skewed log-normal distribution* was found while plotting the time headway distribution which aligns with the recent literature on headway distributions. Log-normal distributions are positively skewed with long right tails due to low mean values and high variances in the random variables which tells us that time headway is stochastic in nature. From Figure 6 it can be seen that percentiles (Q1, Q2, Q3) of time headway are increasing and which means congestion is being occurred gradually. As the standard deviation seems greater than the mean value it can be stated that the gathered data has a good amount of diversity which is good for models to learn.



	Time Headway (s)	Vehicle Velocity (ft/s)
Counts	4,122,130	4,122,130
Mean	4.69	19.85
Standard Deviation	10.56	12.49
1st quartile (Q1)	2.18	10.00
2nd quartile (Q2)	3.01	18.50
3rd quartile (Q3)	4.43	26.52

Figure 6. Statistics & distribution of Time Headway (x-axis: time headway, y-axis: counts of headway)

Features that were used as the input matrix are correlated with time headway. These relationships also satisfy the previous literature about traffic flow theory. We will discuss how these explanatory variables affect time headway both spatially and temporally.

4.3.1. Velocity and flow rate

Vehicle velocity, traffic flow, and time headway are highly correlated with each other maintaining a complex relationship. But this relation can vary across different traffic densities. Li et al. (2017) described this relationship vastly with the use of the probability density function (pdf) [71]. Traffic phenomena being stochastic as nature, these relationships can be stated differently for different traffic and vehicle condition. It can be explained in both microscopic and macroscopic ways.

By ‘velocity’ in general, we understand the average or mean velocity of the vehicles traveling on the same road. When this average velocity is used for modeling purposes, the model becomes a ‘macro’ simulated model. If we want to perform a microanalysis, the instantaneous velocity of the vehicles will be used. There can be seen differences describing the relationships in the micro and macro analyses. Generally, vehicles follow the relationship $s = hv$, where s is the travel distance (i.e., inter-vehicle spacing), h is the vehicle time headway (i.e., inter-vehicle time distance), and v denotes the vehicle speed. So, the reciprocal relationship between velocity and time headway can be stated which is shown by a graph in figure 7. There is a highly negative correlation between average velocity and time headway which follows the previous findings.

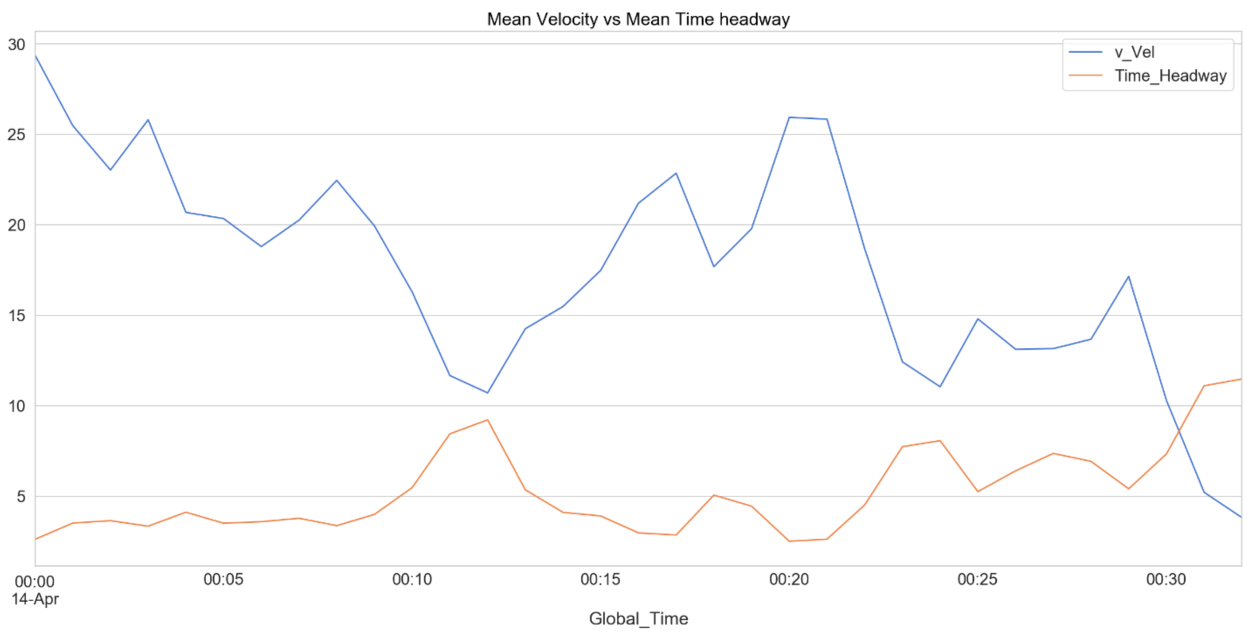


Figure 7: Velocity-headway relationship

Data also explained that not only the velocity of the subject vehicle but also the velocity of the preceding vehicle has an impact on the time headway of the subject vehicle. In mixed flow and congested conditions, drivers tend to be more alert and follow the preceding vehicle as it is the main object the driver can view. It can be further visualized through a microscopic analysis showed in figure 8. When the velocity of the vehicle in front of a subject vehicle (preceding vehicle) increases, the time headway of the subject vehicle decreases and vice versa. This explains that driver maintains a greater distance when the difference in velocity of the subject and the preceding vehicle gets smaller.

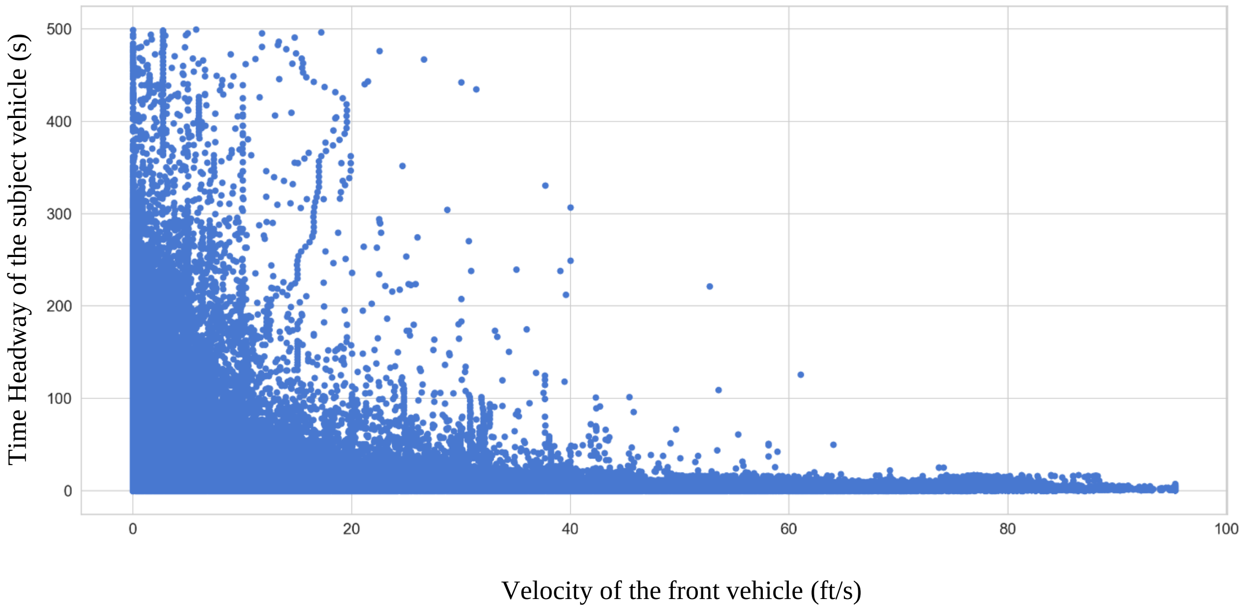


Figure 8 Velocity of preceding vehicle vs Time headway of subject vehicle

While analyzing at the macro-level, it can be seen that the flow rate has a stochastic impact on headway and vehicle velocity. This impact cannot be specifically stated as positive or negative on headway. Because traffic phenomena and driver behaviors are different in free flow and congested flow. This complex relationship can be viewed clearly by plotting the whole time-series data for better interpretation. We have shown the summed graph to zoom in for a detailed analysis. Velocity and the flow rate have been scaled into a similar range (o to 1) for showing correlation with the time headway in Figure 9. The graphs indicate that, when flowrate increases the time headway increases drastically after attenuating some durations. Whereas the vehicle velocity gradually decreases, while the time headway increases. This means vehicles were entering more in the road segment and as time goes by, congestion was about to occur. When the road segment was gradually being filled with more vehicles the time headway increased and velocity decreased, which indicates the congestion state of the road segment.

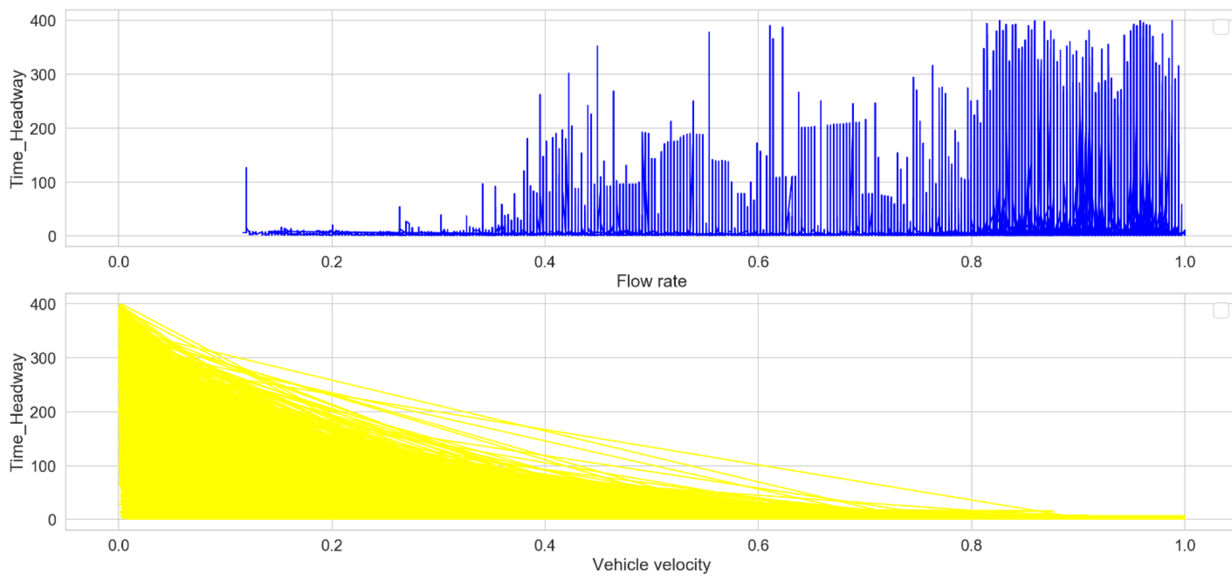


Figure 9. Correlation of time headway with flow rates and velocity

The reciprocal relationship between velocity and time headway can be stated as follows:

$$h \propto \frac{1}{v} \tag{15}$$

Here, h and v are represented as time head way and vehivle velocity respectfully. Time headway is inversely proportional to the instantaneous velocity of the vehicle. It means that driving at a higher speed leads the vehicle to pass a reference point more quickly and vice versa.

In the last few minutes, as there have been added more vehicles on the segment, the flow rate gradually becomes to decrease which made higher time headway and lower velocity (close to zero). At the end of the study time, it can be farther viewed in a zoomed figure (Figure 10) that the flow rate is very close to ‘0’ which is stated as the total congestion where time headway is the largest and average velocity is in a lower range (0~5 ft/s).

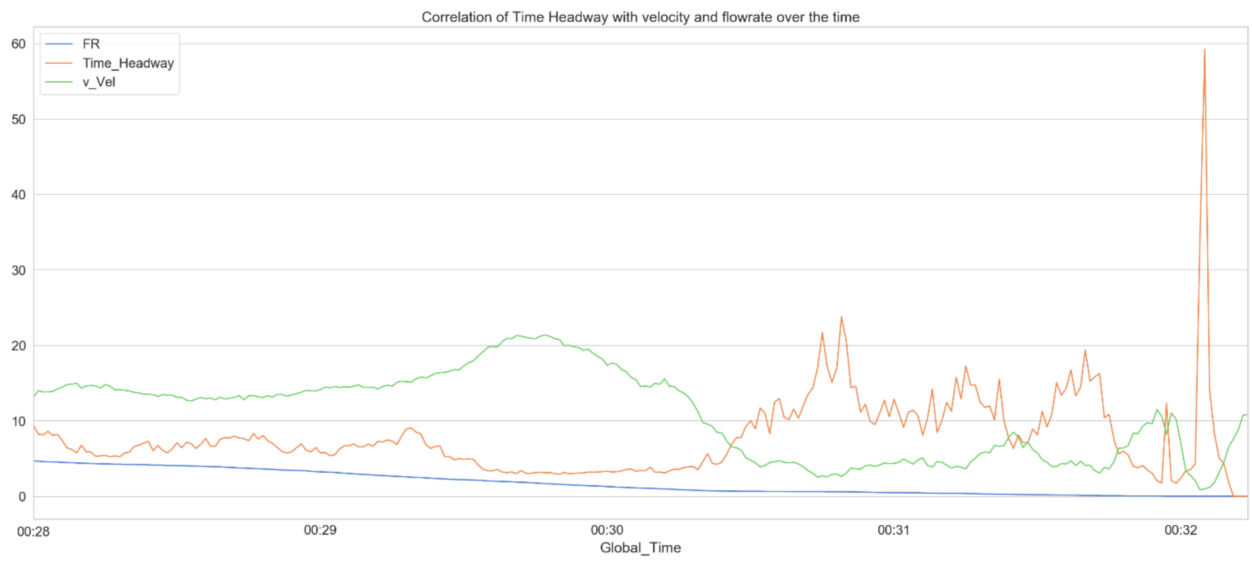


Figure 10. Correlation of headway with velocity and flowrate (last four minutes)

This flow-velocity-headway relationship can be also stated microscopically. For analyzing microscopically we selected a single vehicle within a specific study time and plotted the same graph for that vehicle. It can be seen in figure 11 that the relationship of flow rate with the vehicle velocity and time headway conflicts the macroscopic point of view. Because now, ‘velocity’ is considered as instantaneous velocity, not the average velocity.

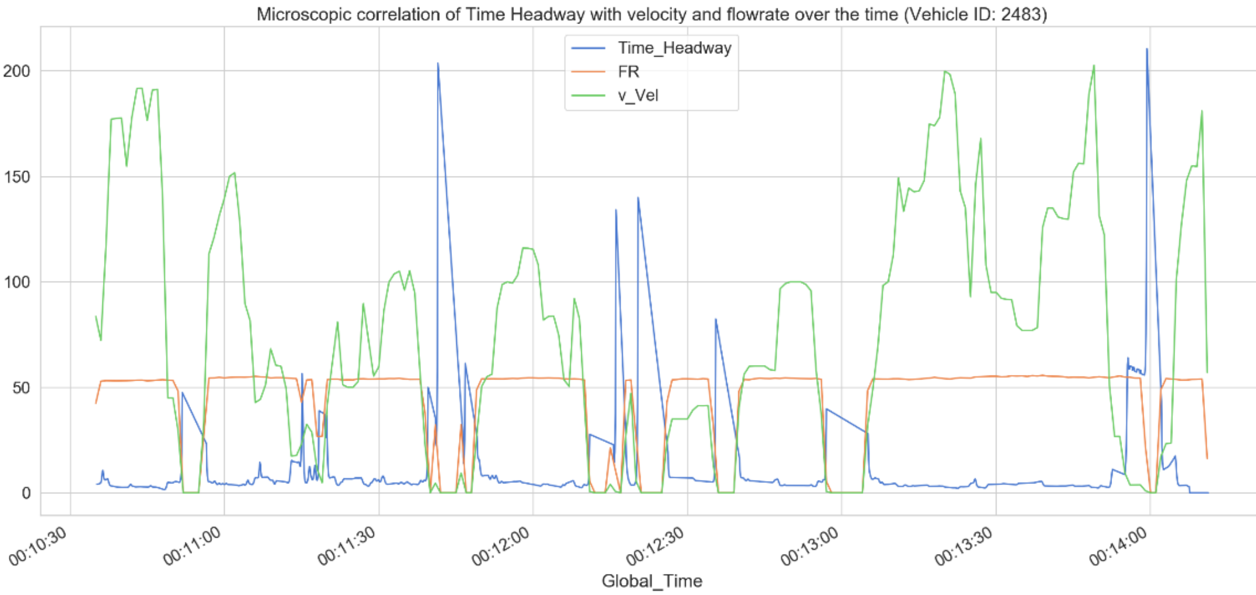


Figure 11. Microscopic analysis (for vehicle ID 2483)

)

This graph shows a positive correlation of flow rate with the instantaneous velocity. It means as the instantaneous velocity is increasing, there will be more vehicles entering the road segment which will begin to rise the flow rate. As there will be more vehicles on the road the traffic density will be gradually increasing.

$$\text{Traffic density} = \frac{\text{Number of vehicles on a specific road segment}}{\text{Length of that road segment}} \tag{16}$$

Eventually, the gaps between the vehicles will be getting shorter which will lead traffic time headway to get higher (at 10:38 to 10:39).

4.3.2. Vehicle type-wise analysis

As the diversity of passenger types is incrementing day by day, the mode choices of passengers have been gradually changing which leads to present more than one type of vehicle on a road segment. Having different width and length of the vehicle certainly affects the total traffic phenomena. Especially in attendance of a mixed type of vehicles on a road segment describe the stochastic nature of traffic theory. In our dataset, there are three types of vehicles on the segment- Truck, Passenger car, and Bike. Expectedly the shape of time headway distribution over time of each type of vehicle has shown similarity (Log-normal distribution with different means and standard deviations).

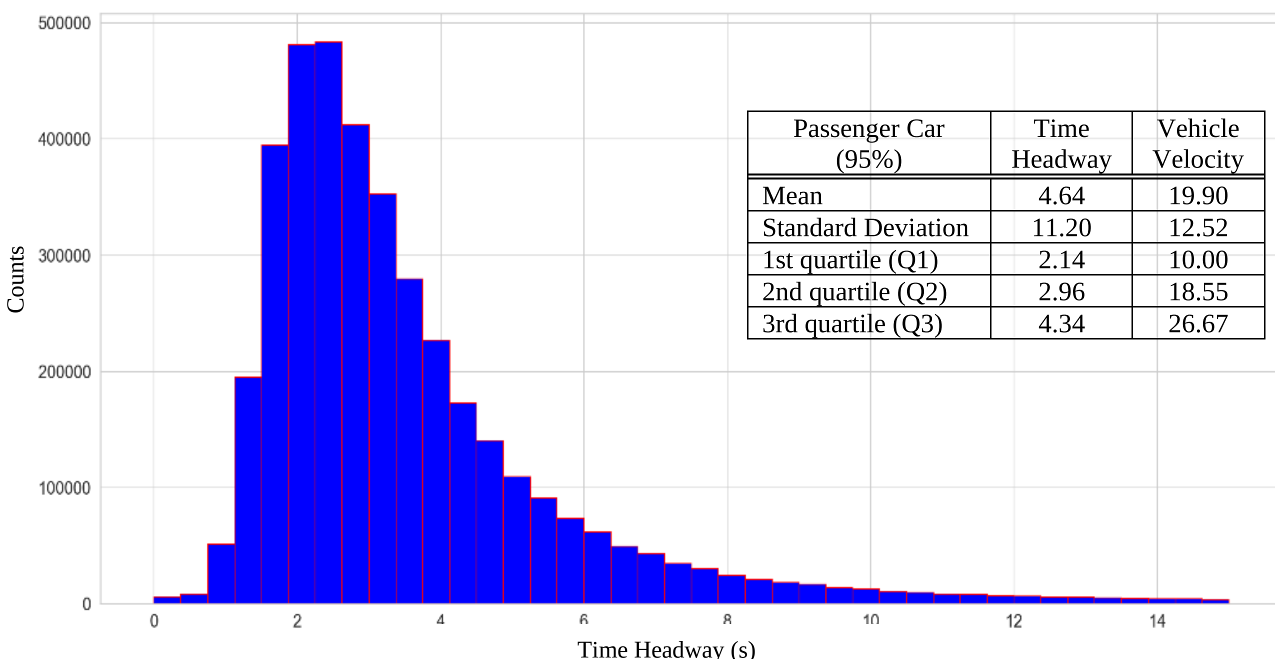
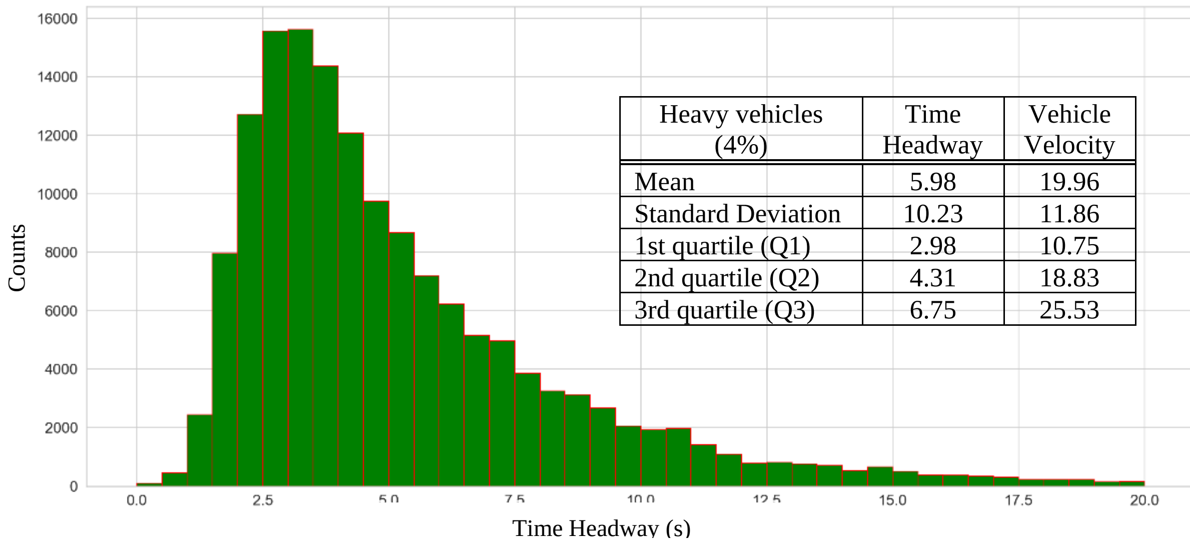


Figure 12 Time Headway distribution of Vehicle Class-2 (Passenger Car)

Figures 12 and 13 both show the same shape which are time headway distributions for passenger cars and trucks respectfully. But clearly, time headways (mean and percentiles) of trucks are higher than passenger cars. Also, they have different standard deviations. As heavy vehicles tend to maintain a similar range of velocity, the standard deviation is lower than the class-2 vehicles. It implies passenger cars have more stochastic time headway than trucks.



Heavy vehicles (4%)	Time Headway	Vehicle Velocity
Mean	5.98	19.96
Standard Deviation	10.23	11.86
1st quartile (Q1)	2.98	10.75
2nd quartile (Q2)	4.31	18.83
3rd quartile (Q3)	6.75	25.53

Figure 13. Time Headway distribution of Vehicle Class-3 (truck)

The length of the vehicle has a considerable effect on the headway of the surrounding vehicles. From theory, we know that generally larger vehicles tend to have longer time headway as they move at a constant speed most of the time and take a longer time to pass the next vehicle and vice-versa. Also, the type of preceding vehicles affects the trajectories of subject vehicles traveling in the same lane, which results in variations on the headway. It is due to the vehicle

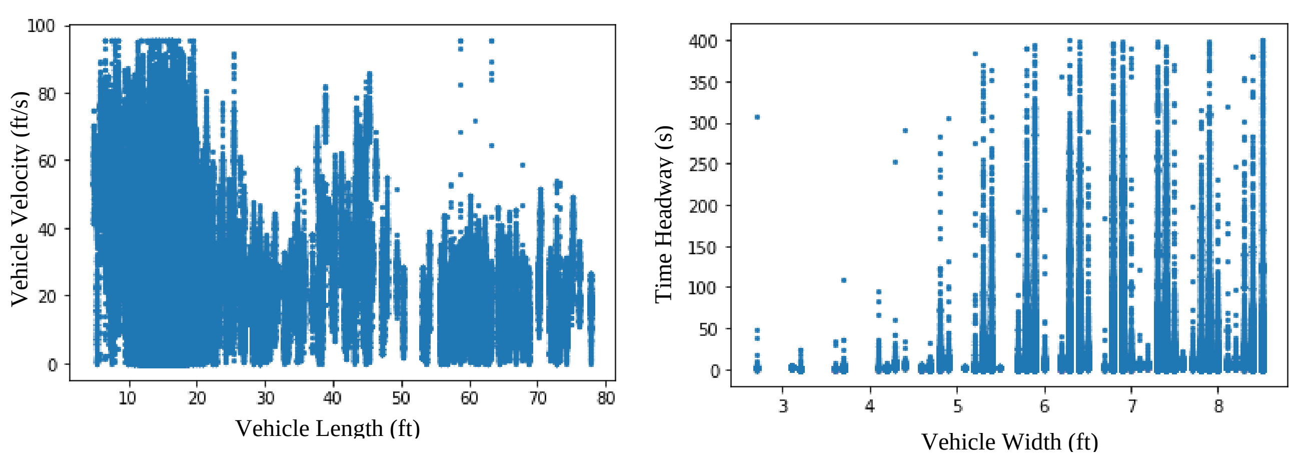
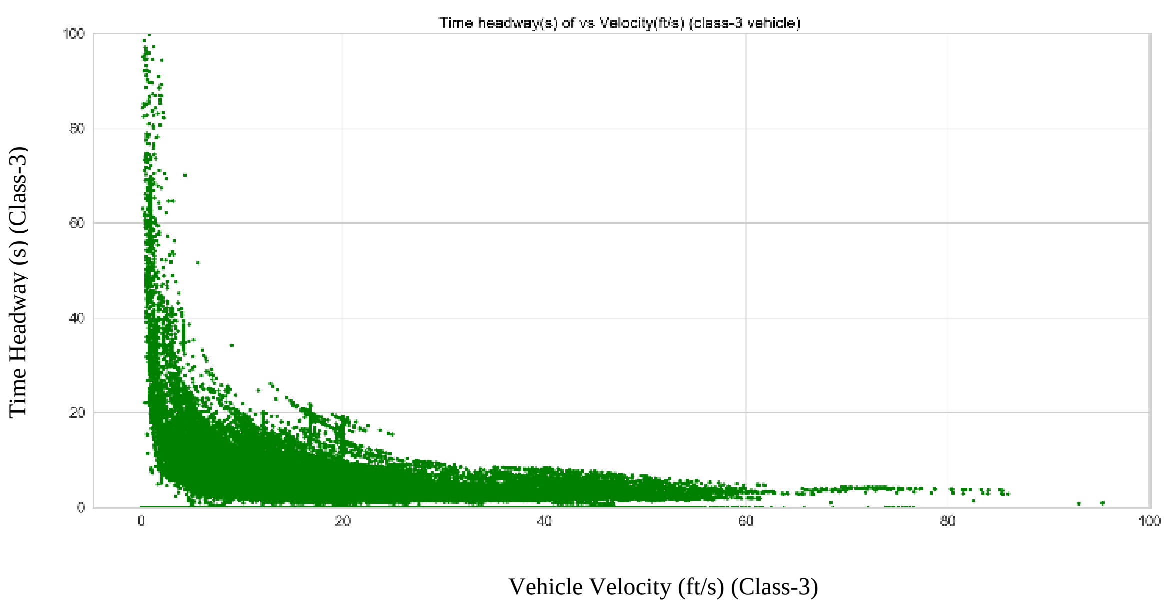


Figure 14. Relationship of headway and velocity with vehicle shape

length and shape.

Drivers tend to be more cautious and generally maintain a greater distance when a truck or a bigger vehicle is at the front. This phenomenon can be explained in Figures 14, where we show two scatter plots regarding changes in velocity and headway with vehicle width and length. Time headway is greater when the width of the vehicle increases and instantaneous velocity is lower when the length of the vehicle increases. The velocity of passenger cars has a longer range than heavy vehicles. It means class-2 vehicles provide more variation in time headway than the class-3 vehicles. It can be further visualized in figure 15, where the velocity of class-2 vehicles seems more scattered than class-3 vehicles which cause the diversity in their headways as well. Also, it can be seen that the velocity of class-3 vehicles has a velocity range of 10 ft/s to 60 ft/s whereas class-2 vehicles have a velocity range of 0 ft/s to 90 ft/s which also affects the heterogeneity of their time headways.



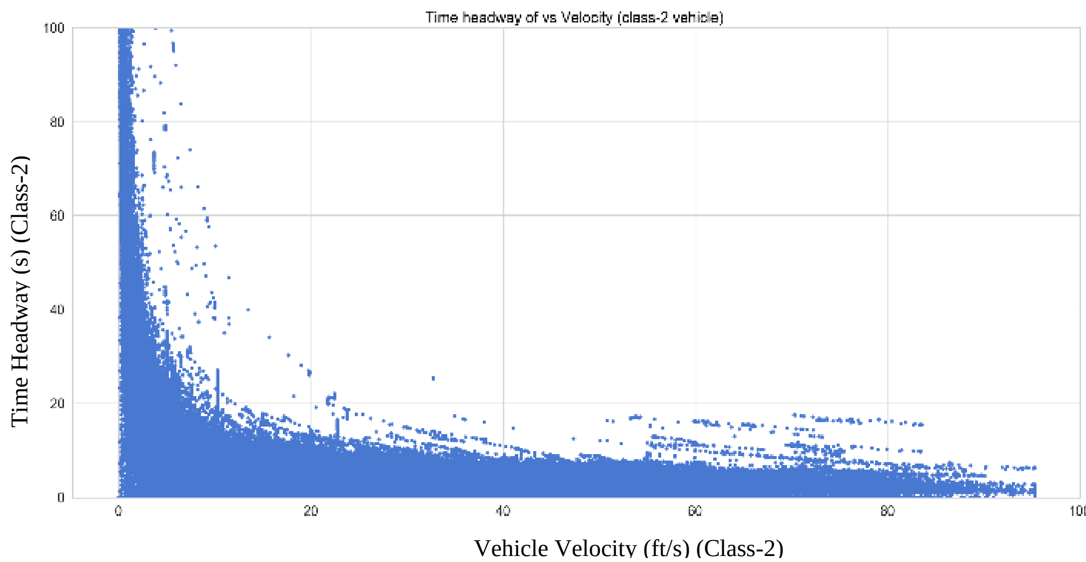


Figure 15. Headway diversity of different vehicle classes
(Class-3: upper figure, Class-2: lower figure)

This also shows heavy vehicles maintain lower velocity and greater time headway. When a vehicle of bigger length is moving next to a passenger car, the lower velocity of the larger vehicle will certainly affect the headway of the following passenger car. Also, drivers tend to keep a good distance from heavy vehicles and avoid being in the same lane. Passenger cars are more dependent on their preceding vehicles than heavy vehicles.

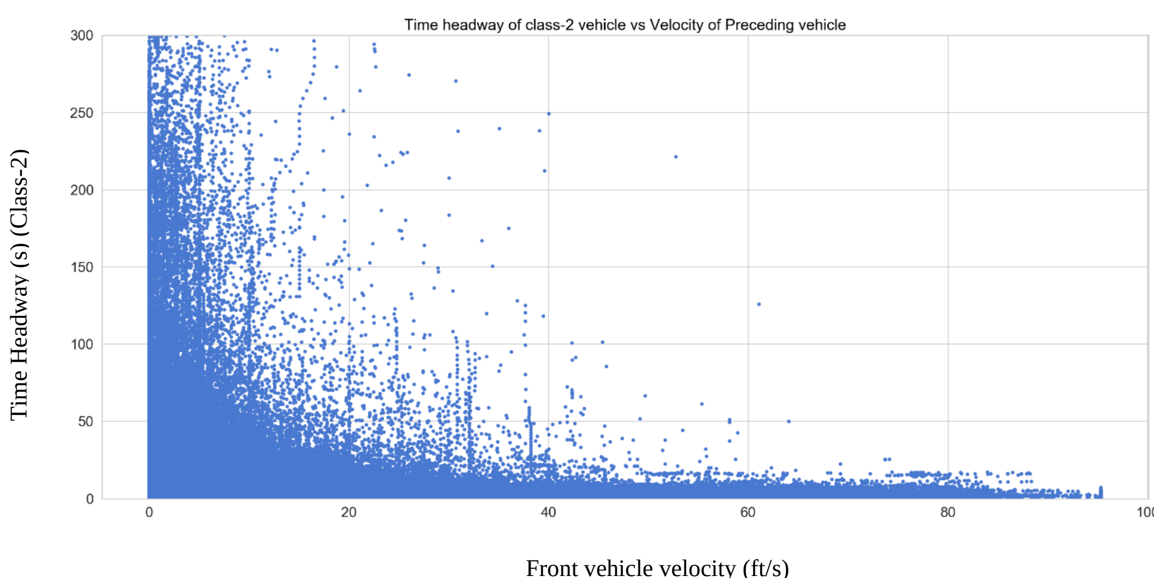


Figure 16. Effect of the velocity of preceding vehicle on Passenger Cars

Figures 16 and 17 explain that the velocity of a passenger car varies a lot concerning its preceding vehicle's velocity but as heavier vehicles try to maintain a steady speed range, the type or the velocity of its preceding vehicle do not affect that much on their headways.

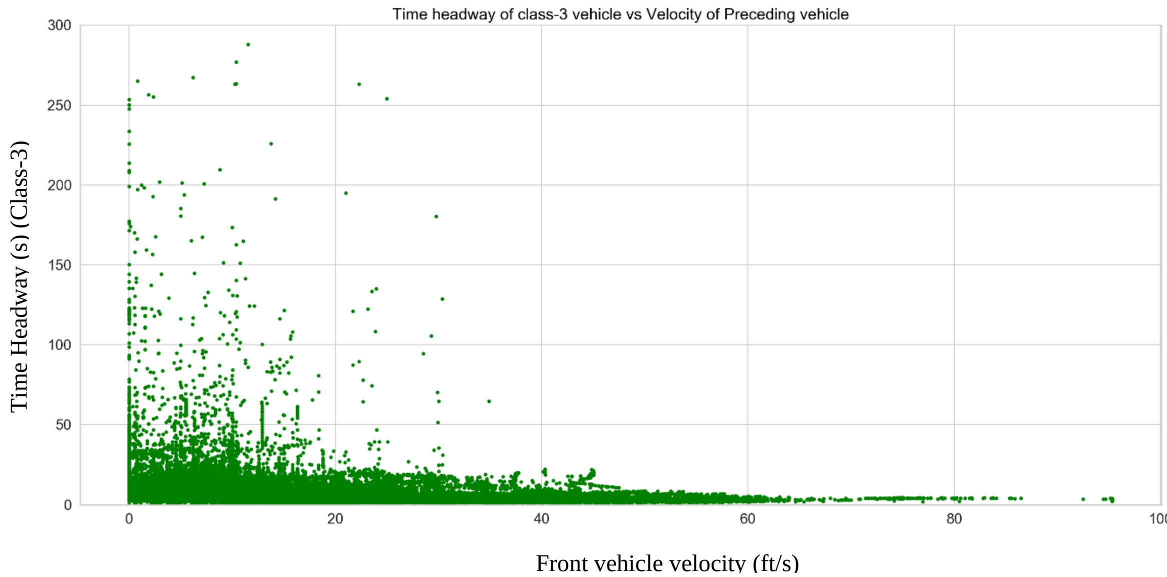


Figure 17. Effect of the velocity of the preceding vehicle on Heavy vehicles

4.3.3. Lane-wise Analysis

Different classes of vehicles have been seen to use some specific range of lanes in recent years. There are even rules for some roads also for different classes of vehicles to travel on specified lanes. It is because of the variant speeds and characteristics of the vehicles. As discussed earlier, preceding vehicles have a significant effect on the headways of the subject vehicle, thus it is useful to travel similar types of vehicles on the same lanes for traffic safety and maintaining regular headways. Irregularity in headways can lead to traffic jams and affect driver behaviors also. For example, if a heavy vehicle is in front of a bike or passenger car, the driver of the subject vehicle may get frustrated as the heavier vehicle in front of him is moving slowly. So, the driver may end up overtaking the front vehicle. As the heavy vehicles are generally bigger, it is always risky to overtake these types of vehicles which may arise crashes.

In our dataset, it can be visualized that not all the lanes were used equally, and each class of vehicles tried to maintain specific lanes to travel. From figure 19, lane 4 was mostly used by all of the vehicles together where lane 7 was least used. It is expected because lane 7 is the on-ramp lane, not one of the main lanes. On-ramp is a lane that connects to a highway merging into the auxiliary and main lanes. As there was less density on lane 7, the characteristics of this lane are different from the main highway lanes. It is used to get prepared for a highway. While entering the highway, drivers need to gradually speed up to adjust with the auxiliary and highway lanes. As lane 1 is the farthest left lane, most of the class 1 vehicles (from figure 18, 80% of total class 1 vehicles) traveled within this lane. The density of class 1 vehicles being high in lane 1 also affected the headway of other vehicle types traveling on the same lane. From figure 20, it can be visualized that the average time headways of

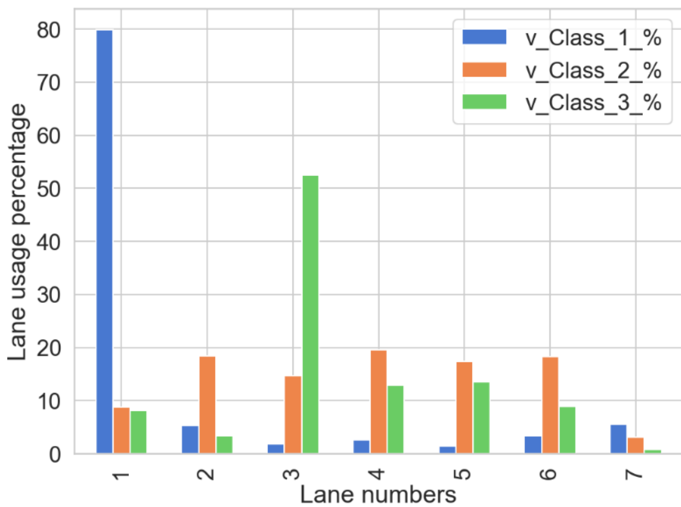


Figure 18. Lane usage percentage of each vehicle class

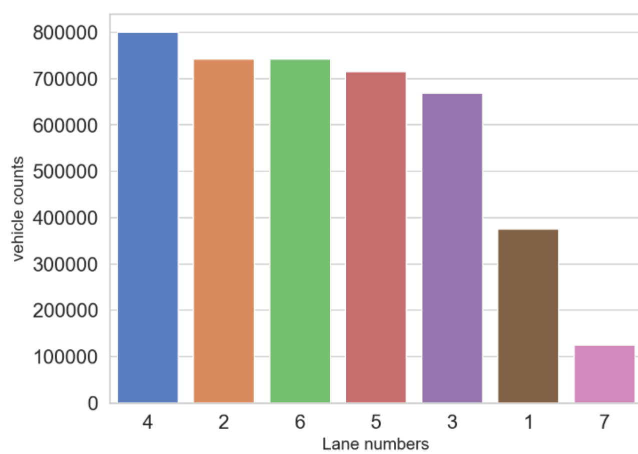


Figure 19. Lane usage

vehicle classes 2 and 3 are the lowest in lane 1. This indicates the effect of getting space in this lane smaller vehicles on time headway of other vehicles. Heavy vehicles avoided the 2nd lane because the lane was covered the most with passenger cars (class 2). As there were fewer heavy vehicles in the 2nd lane, the difference between the average headways of class 2 and class 3 vehicles was the least (figure 20). It is because being the lower percentage in numbers, heavier vehicles had to maintain a speed range with passenger cars to avoid congestion and confliction.

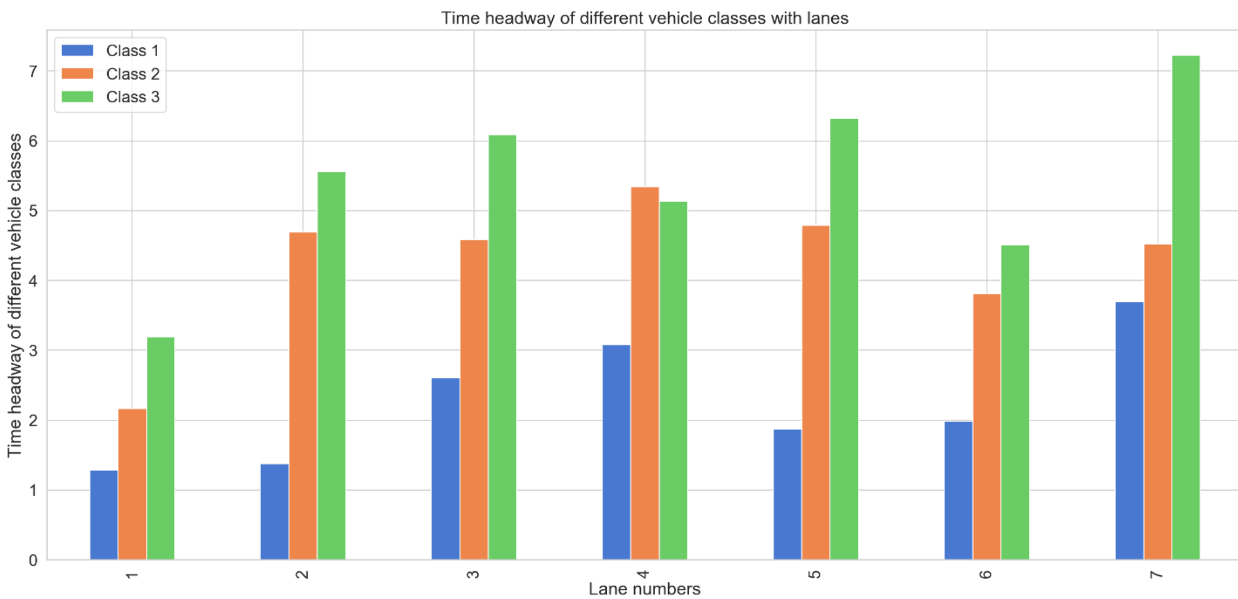


Figure 20. Average Time Headways in each lane

Around 55% of the class 3 vehicles used lane 3 (from figure 18) whereas class-2 vehicles were lower in percentage rather than in any other lanes. It describes the driver's behavior of avoiding lanes which are occupied by heavy vehicles. Traveling most of the heavier vehicles on lane 3 also increased their headways from lane 2 to 3. As lane 3 was less occupied by passenger cars because of the high percentage of heavy vehicles, they shifted to lane 4. It can be seen in figure 19 that lane-4 has the most number of vehicles moving by. Most of them are passenger cars. The highest percentage of passenger cars (20%) were traveling on this lane (from figure 18). The 4th lane had enough heavier vehicles which also had an impact on their (passenger cars) headways. We can see from figure 20 that, only in lane 4 the passenger cars had higher average time headway than any other lanes. It indicates lane-4 as a bit congested lane than other lanes. In lane-5, the percentage of heavy vehicles was a bit higher and the percentage of passenger cars was a bit lower than lane-4 which again made lane-5 consisting of a bit higher proportion of heavy vehicles than other lanes (except lane-3). For this reason, the average time headway of class-3 vehicles was increased and passenger cars were traveling by slightly higher velocity than lane 4. It is also because lane 5 is the lane next to the auxiliary lane.

The shape of the headway distribution of each lane is similar and the same shape of the vehicle type-wise headway distribution. These distributions also provide a log-normal distribution. But as the usage of lanes for various vehicles was not similar, the mean and variance of the lane-wise distributions vary lane to lane. From figure 21, it can be explained that the on-

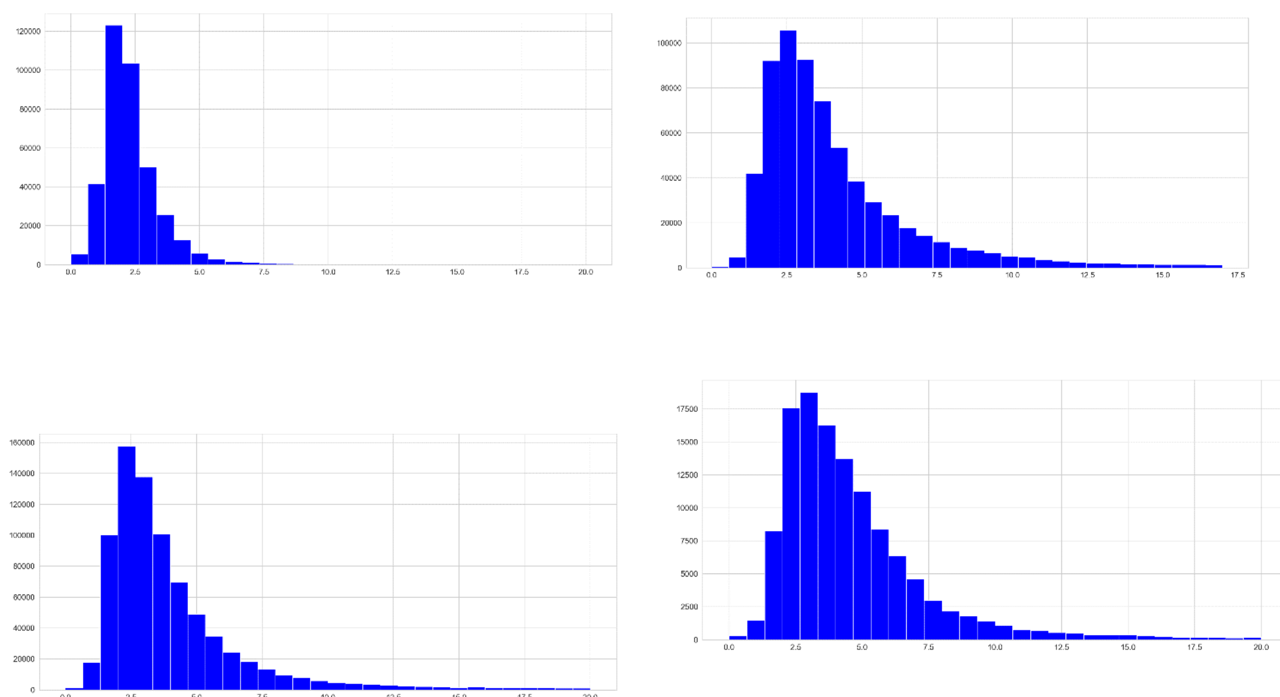


Figure 21. Lane-wise Time Headway distributions (upper left= lane 7, upper right= lane 1, lower left= lane 4, lower right= lane 3)

ramp lane (lane 7) had fewer vehicles indicating different flow rates with higher average time headway. In the on-ramp lane, drivers tend to transit the velocity to enter into highways which resulted in higher variance in headway distribution (figure 15, upper left). While entering into the auxiliary lane drivers increased the velocity to cope up with the highway lanes. But because of

Table 5. Lane wise statistics

	Mean		Standard Deviation		1st Quartile (Q1)		2nd Quartile (Q2)		3rd Quartile (Q3)	
Lane No.	Time Headway (s)	Velocity (ft/s)	Time Headway (s)	Velocity (ft/s)	Time Headway (s)	Velocity (ft/s)	Time Headway (s)	Velocity (ft/s)	Time Headway (s)	Velocity (ft/s)
1	2.33	44.78	1.56	11.84	1.62	36.42	2.08	43.83	2.75	51.75
2	4.88	18.67	13.14	9.67	2.13	11.04	2.87	19.50	4.09	25.63
3	4.99	17.22	9.94	8.71	2.42	10.00	3.37	16.72	5.06	23.85
4	5.54	15.93	14.30	9.29	2.34	8.63	3.25	15.00	4.85	22.33
5	5.04	17.80	11.80	10.03	2.31	10.00	3.18	16.95	4.62	24.89
6	3.97	18.93	7.71	10.31	2.02	10.02	2.81	18.47	4.17	25.75
7	5.51	12.42	9.85	9.69	2.78	5.83	3.97	9.68	5.73	15.05

the congestion in lane 4, vehicles were moving slower in lane 5 than the auxiliary lanes. It can be seen from Table 5 that, time headway was increased drastically in lane 5 from the auxiliary lane. Lane-1 did not face too much congestion because vehicles were moving fast and freely. This lowered the mean headway and variance. Investigating the quartiles (Q1, Q2, Q3) of headways, it can be seen from Table 5 that, lane 1 and lane 2 did not have many differences in their quartiles but in lane 3, there is a bigger increment in time headways. This means transiting into a more congested situation. As there was the highest number of vehicles on lane 4, vehicle density was higher which led to mean time headway to get higher (especially for class 2 and 3 vehicles). Also, as the number of heavy vehicles increased in lane 3, the mean time headway was a bit higher than lane-2 (Table 5). Because of the increased traffic density, the standard deviation of the time headway was higher too in lane 4, which means variations in speed and headway. In fact, the average time headway was the highest in lane 4 than in any other lanes.

From the above analysis, it can be justified that, variables (features) selected for training have significant influences on predicting traffic time headway.

CHAPTER V

MODEL IMPLEMENTATION & VALIDATION

5.1. Model Set-up and Configuration

Deep learning models perform well in solving complex relationships between features if all the parameters are tuned properly. Recently these models are being used in different sectors of ITS and prediction problems. The *Long Short-Term Memory networks (LSTM)* has been implemented in travel time prediction (Duan et al., 2016, Rosberg et al., 2018), car-following models (Huang et al., 2018, Zhang et al., 2019), speed predictions (Gu et al., 2019, Wang et al., 2019) and some other demand modeling and forecasting problems. In each problem, the model configurations were different as model setup completely depends on the shape and complexity of the data that is being fed into it.

5.1.1. Hyperparameters

LSTM networks have several hyperparameters to tune for better prediction performance. These hyperparameters are- historical time step, batch size, number of epochs, number of neurons and layers, learning rate, etc. As mentioned earlier, the historical time step refers to the number of previous steps to look back while training. Our model has provided higher accuracy prediction using a single historical time step which took a shorter time to train rather than using more than a single time step. Batch size defines how much data is being trained at a time.

For example, if *n* numbers of the sample are being trained with the batch size of *b*, the model will train *n/b* times in every epoch. It means it will train every *b* number of samples and update the weights until it reaches to n^{th} sample in every epoch. Epoch refers to how many times the model will be trained where the updated weights will be transferred to the next epoch. That's how the training and validation accuracy keeps decreasing which indicates that the model is learning well in every epoch and weights are being updated in every epoch. But overtraining the same sample may leed to an overfitting problem. To eradicate this situation we used a callback function named 'Early stopping' which stops the iteration when validation accuracy does not improve. It has been described in detail later. The learning rate is one of the most important hyperparameters of optimization problems in machine learning. It indicates the time steps to be taken by an optimizer while moving towards a minimum of the loss function. Initial arbitrary weights are generated by the model at the beginning of the training. As it is being trained, the weights will be changed throughout time. This change of weights is the result of the learning rate, which indicates how many time steps to learn for reaching into the minimum loss.

5.1.2. Neurons and Hidden Layers

Selecting the appropriate number of neurons and hidden layers in a deep learning network is a challenging first task for training. Though there are several rules of thumb for selecting the optimal number of neurons and layers, in reality, it is all about performing several trial and errors. After completing several trials, we can point out that increasing the number of layers does not necessarily improve the training accuracy. Increasing the number of layers from 1 to 2 enables the model to learn some complex correlations. But the performance of the model deteriorates when the number of layers is increased to more than two. Thus, two hidden layers and one dense layer (output layer) have been chosen. We obtained the optimal model performance while setting up the number of neurons to 512 and 256 in 1^{st} and 2^{nd} layers respectively.

5.1.3. Activation function

The neurons of our human body transmit information from different parts of the body to the brain which eventually leads to an output. Like the human body, artificial neural networks map each node's inputs to its corresponding output within a specific range (for example, 0 to 1, -1 to +1, etc,). This is referred to as an activation function. There are several individual activity functions for different purposes. For example, *ReLU* (Rectified Linear Unit) is usually used for regression problems and *Softmax* for classification problems. Generally, for regression problems, there are two kinds of activation functions- linear and non-linear.

In this study, the activation function that was used in hidden layers is ReLU (Rectified Linear Unit). It is a nonlinear function which is as follows:

$$ReLU(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{17}$$

The basic use of ReLU is to transfer negative values to 0, making it efficient and easy for computation. One of the most important reasons to use ReLU is that it resolves the vanishing gradient problem, which other functions like *sigmoid* and *hyperbolic tangent* fail to do so.

5.1.4. Loss function

The objective function refers to the absolute difference between the predicted and the true value. When this function is being minimized in a model, its called *loss function* or *cost function*. There are different loss functions for regression and classification problems. This function has to be estimated repeatedly in every iteration so that the weights used in the function can be updated to minimize the error in the following iteration.

For regression, Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPL), and Mean Squared Logarithmic Error Loss (MSLE) have been used as loss function in previous studies. Among them, MSE and MAE perform well for LSTM and other deep neural networks as DL models generally are trained using the stochastic gradient descent optimization algorithm. Some regression problems have outliers, e.g. large or small values far from the mean value. For example, our target variable (Time Headway) has lots of outliers as it contains data from mixed vehicle types and diverse traffic conditions. The Mean Absolute Error (MAE) is an appropriate loss function in this case as it is more robust to outliers. It is calculated as the average of the absolute difference between the actual and predicted values. So, in this study Mean Absolute Error (MAE) has been used as a loss function.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y} - y| \tag{18}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y} - y) \tag{19}$$

Here, N denotes the number of training samples, $\hat{y}$ and y refer to the predicted and true values respectively. By plotting the graph of loss function over the iterations the fluctuation and shape of the error can be visualized. To get more information on the gradual accuracy of training we considered Mean Squared Error (MSE) as to accuracy metrics. We used this metric to observe how the model learns after each epoch in the different loss functions.

5.1.5. Optimizer

The basic idea of implementing neural network models is to learn and teach gradients of the original network which is called meta-learning. Optimization algorithms enable the model to exploit the structure automatically by adjusting network weights. Different optimizers have been used in different studies. Among them, Adam, RMSprop, and AdaGrad optimizer are being vastly used as these optimizers can detect nonlinearity. As headway modeling is a stochastic gradient-based optimization problem, the Adaptive Moment Estimation (Adam) optimizer (reference: https://keras.io/api/optimizers/adam/) has been used in this study. Having impressive computational efficiency, low memory requirement, and low loss make adam optimizer one of the brightest choices for predicting through LSTM NN. The update rule for Adam optimizer is stated in the following equation.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}}\hat{m}_t \tag{20}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{21}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{22}$$

Here, m_t and v_t are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients respectively. β_1, β_1, and ϵ are hyperparameters of Adam which are set to be 0.9, 0.999, and $10^{\wedge}-8$ respectively. The learning rate (lr) is set to be $1e^{-2}$. But as the 'Early Stopping' callback function was included in the model, the learning rate would change according to the improvement or deterioration of the loss functions. The summary of the selected model parameters is shown in the following table.

Table 6. Model summary

Model parameters	Descriptions	Value/ Name
Input features	Number of variables (reshaped data frame)	31
Output dimension	Dimension of Output	1
Training size	Number of training samples	2,677,950
Hidden layers	Number of Hidden Layers	2
Neurons	1st Hidden layer	512
	2nd Hidden layer	256
Loss function	Loss calculator of training and validation	MAE
Activation function	Obtaining outputs from nodes in hidden layers	ReLU
Optimizer	Adjust weights to minimize loss	Adam
Initial earning rate	The initial rate for the machine to learn	$1e^{-2}$
Batch size	Number of samples trained in each iteration	192
Epochs	Number of training update	35

5.2. Training the model

The model is a sequential model where two LSTM hidden layers have been stacked together and connected with a Dense layer which enables the output of one value. After compiling the loss function and the optimizer, the model structure has been finalized. But before training, we need to set up some callback functions. The callback is a set of functions that enables automating some tasks after every training epoch that helps us to have control over the training process. We have used two callback functions. One of them is called "Early Stopping" which allows us to specify an arbitrarily large number of training epochs and stop training once the model performance stops improving on a hold-out validation dataset. It also protects the model from overfitting restoring the best weights after each epoch. As mentioned earlier, we have set the initial learning rate *(lr)* to be $1e^{-2}$. But training on a dataset having stochastic nature, this rate should not be set permanently throughout the training process. To overcome this issue we used another callback function named 'Reduce Learning-rate), which monitors the validation loss and adjusts the learning rate if the validation loss deteriorates. Callbacks provide a way to execute code interacting with the training model process automatically.

Batch size and number of epochs are other hyperparameters. This pair of values were obtained by performing several trials. We selected a batch size of 192 and trained the model for 35 epochs which gave us the optimum result. As deep learning and machine learning algorithms are not deterministic the output will vary every time training on the same model structure. This is due to the random initialization of the weights. So, the model was repeated 10 times and selected the one that returned the lowest error (both training and validation losses).

5.3. Validation

65% of the total data was trained and validated with 15% data. The rest 20% of the data was applied to test the model. While training, the validation loss was monitored closely over time. The model stopped after 14 epochs restoring the best weights. From Figure 22, we can see the declination of both training and testing loss. Validation loss was a bit higher than training loss on initial epochs which is expected. But it gradually decreased over epochs and maintained a close gap with training loss. This means the model was capable to learn the data well. Both losses declined sharply within two epochs and after that declination rate became slow. Eventually, validation loss slightly decreased than training loss at the last epoch after preserving minimal distance with the training loss.

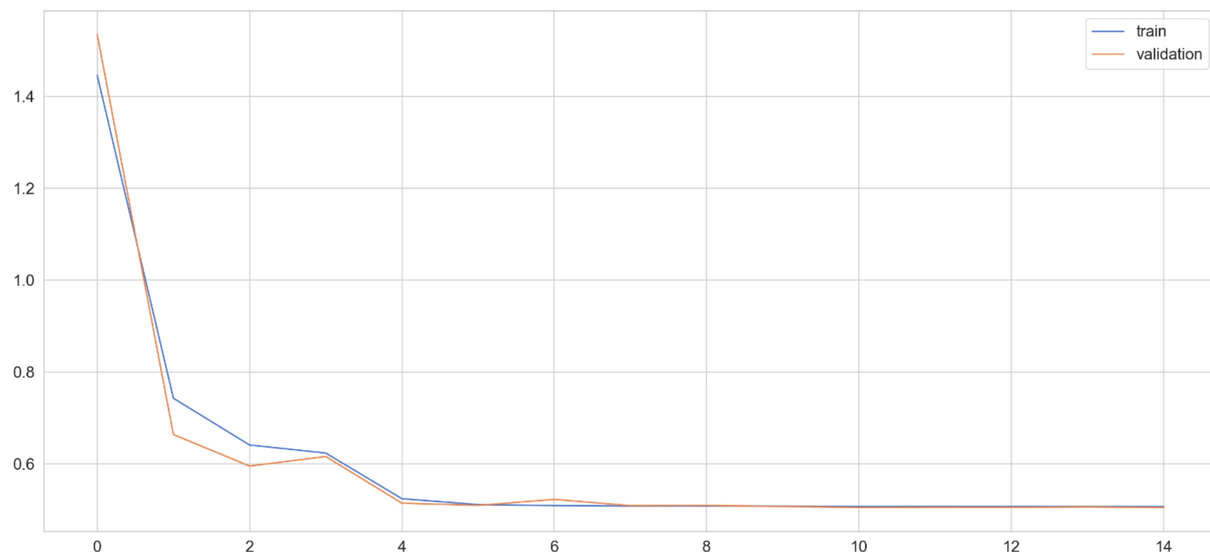


Figure 22. Loss functions over epochs

After validating the model with the validation data our model is ready to predict over unknown data. It is always recommended to use different validation and test data so that the model can be evaluated without any bias. To get an overall accuracy on predicting over unseen data (test set), *Root Mean Squared Error (RMSE)* was calculated between the actual and predicted value. We chose to calculate RMSE value as the model is sequential regression in type. To be more precise, it was one of the important parameters to look at after each training and prediction to see if or which parameter(s) needed to be tuned farther. *Root Meaned Squared Error (RMSE)* indicates the standard deviation of the residuals (predicting errors) which can be also stated as the difference between the actual and predicted values. It is calculated by taking the square root of the average squared residuals. The equation is as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y} - y)\,2} \tag{23}$$

Here, $\hat{y}$ is predicted and y is the actual value (test values). *N* denotes the number of predicted values which is the length of the test set. Our model gives the RMSE value of 0.088, which is an acceptable value for regression and time series problems.

A model can be treated as a general model if it has higher predictability on different sets of data. To be more precise and accurate, we tested the model with a completely different type of roadway data. Data from the US-101 road segment was used for further validation. It's consisted of 43,98,832 rows, each representing unique information of around 3500 vehicles. The RMSE value was close to 0.07 which is also in an acceptable range. Despite having different flow rates and other trajectory parameters our model was able to map the relationship between the variables and predicted time headway quite accurately. This generalized our model, having good prediction capabilities on unseen data. In both cases, the time headway distribution of predicted and actual values was almost overlapped completely. The distribution of predicted headway also gave us the log-normal distribution curve with almost similar to the mean and standard deviation of the actual time headway distribution.

CHAPTER VI

RESULTS AND ANALYSIS

The Long Short-Term Memory (LSTM) networks have been producing a very high accuracy in predicting time series data in various fields. The implemented model of this study shows the highest accuracy in analyzing and predicting time headway until now. In this chapter, the results of our model will be visualized and analyzed. The correlations of the variables with the predicted time headway will also be visualized for both I-80 and US-101 datasets including the comparison of the headway distribution of actual and predicted values.

6.1. Prediction

The predicted time headway of highway I-80 is shown in figure 23, where green points indicate historical data by which the model was trained. The blue and red lines indicate the actual and predicted values. The void space on the graph represents the range of validation tests (which

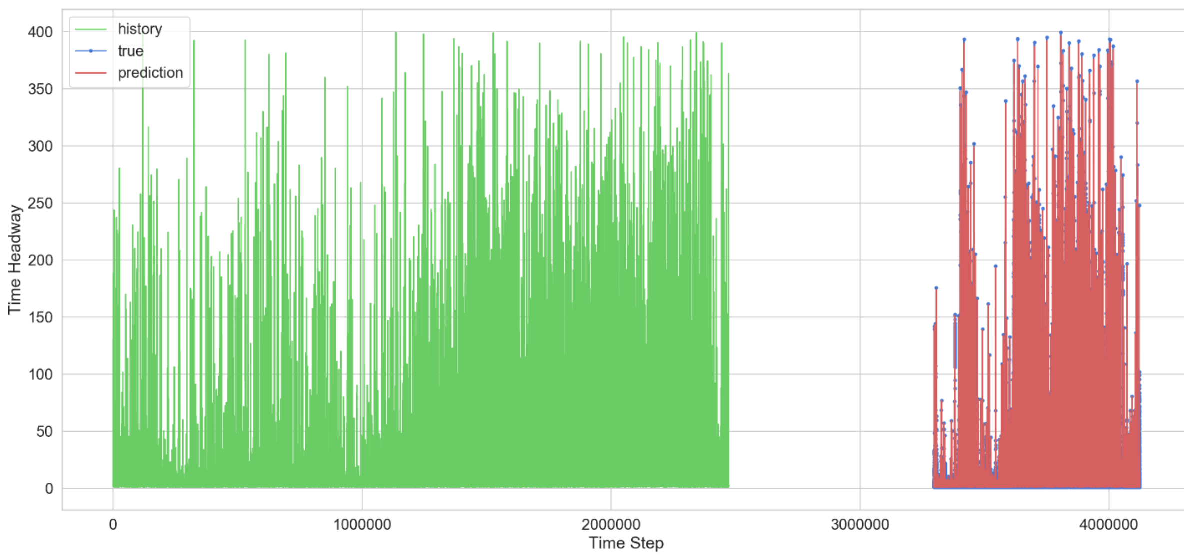


Figure 23. Predicted Time Headway on test set

was 15% of the total data). This is a scatter plot of each vehicle's time headway wherein each time steps the time headways of multivariate vehicles have been presented. It can be seen that red lines almost fully covered and overlapped the blue lines. It indicates the high prediction accuracy on the test set. Different sets of datasets have been used for validation and prediction to avoid overfitting. Overfitting is suspected when the training accuracy is high but the validation and prediction accuracy drops significantly. It implies the model knows the training data well and training loss is very small but it fails to generalize and predicts poorly on unseen data. To be more precise, we validated the model with one set of data and tested it with a different set of data. High accuracy on testing data refers to the ability of the model to generalize the overall scenarios for predicting time headway.

The predicted average and summed time headways of all the vehicles throughout a specific time range (last 30 minutes data) are also visualized in figures 24 & 5. As our data consisted of two separate time-series data (one for 15 minutes and another series is for 30 minutes), in figure 24 the 30 minutes time-series data with predictions has been shown.

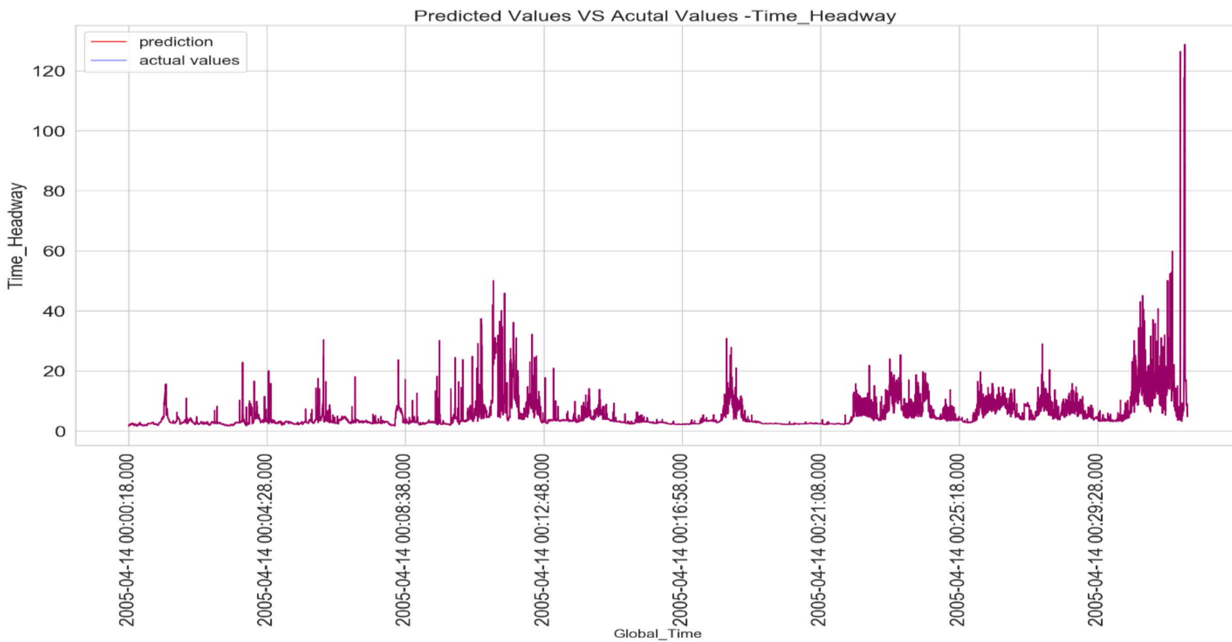


Figure 24. Actual Time Headway vs Predicted Time Headway (Mean value)

Figure 24 presents the average time headway prediction. Here the same scenario can be observed where almost all of the predicted values (red lines) have converged with the actual values (blue lines) of time headway. To be more precise, the summation of the predicted time headways in each timestep have been calculated and is shown in figure 25, where the legend of the graph is similar to the previous graphs.

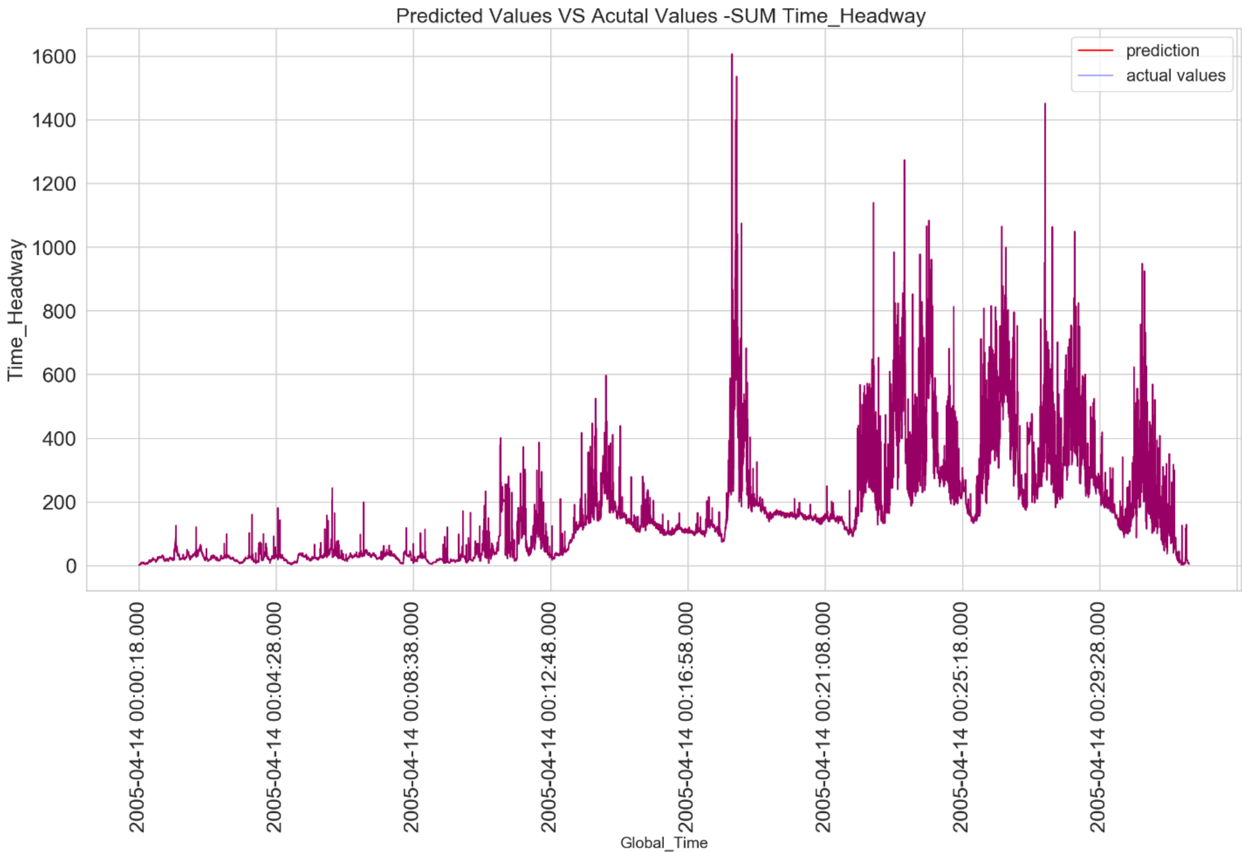


Figure 25.Actual Time Headway vs Predicted Time Headway (Summed value)

Though the graphs indicate overall precise time headway predictions, the goodness and overall error of the model have been also calculated to validate the anticipated time headway.

The goodness of the predicted values have been determined through Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). MAPE is a widely used statistical measurement for forecasting accuracy. It is calculated by taking the percentage of the average absolute error of the total dataset. It can be expressed in the form of the following equation.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y-\hat{y}|}{y} \times 100 \tag{24}$$

Here, y is the actual value, $\hat{y}$ is the predicted value, and N represents the total number of the predicted samples. The overall optimum RMSE and MAPE values of the model were found to be 0.088 and 2.36% respectively. These values give us satisfactory comments on time headway predictions.

But to be more consistent with accurate prediction, we further validated our model on another high way dataset (US-101) consisting of 4,398,832 numbers of data. Data was divided into train, test, and validation sets according to the previous ratio used in the I-80 dataset. The

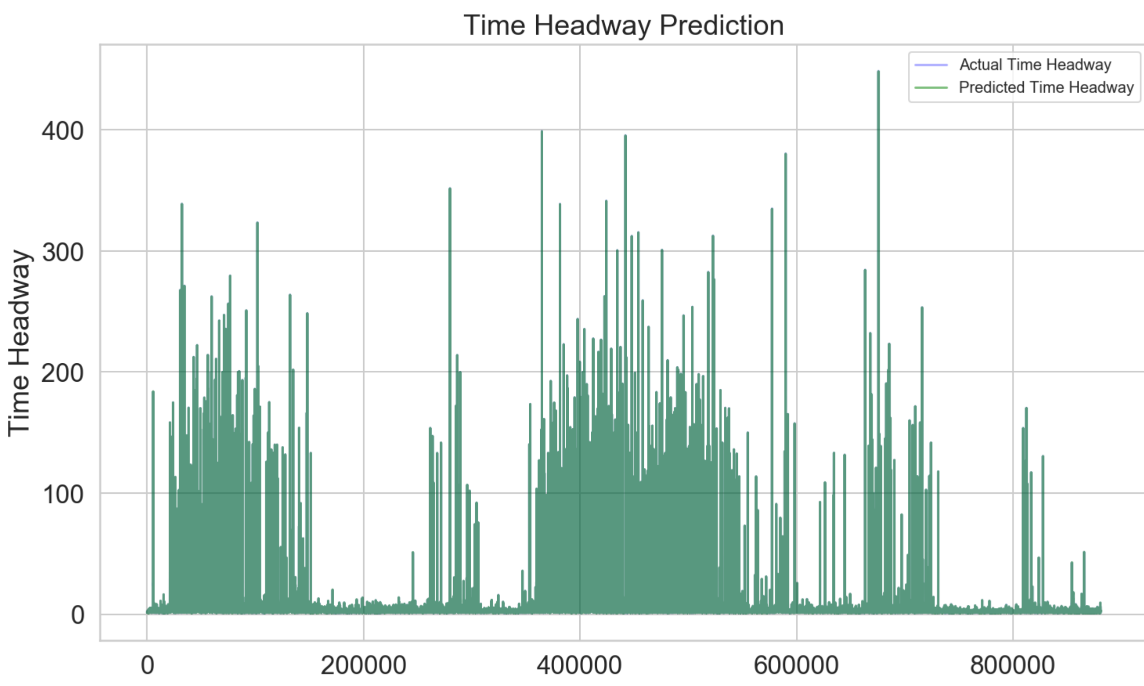


Figure 26. Time headway prediction of US-101 highway

model performs well on prediction having RMSE and MAPE value of 0.07 and 1.89% respectively. These results outperform the previous headway prediction models.

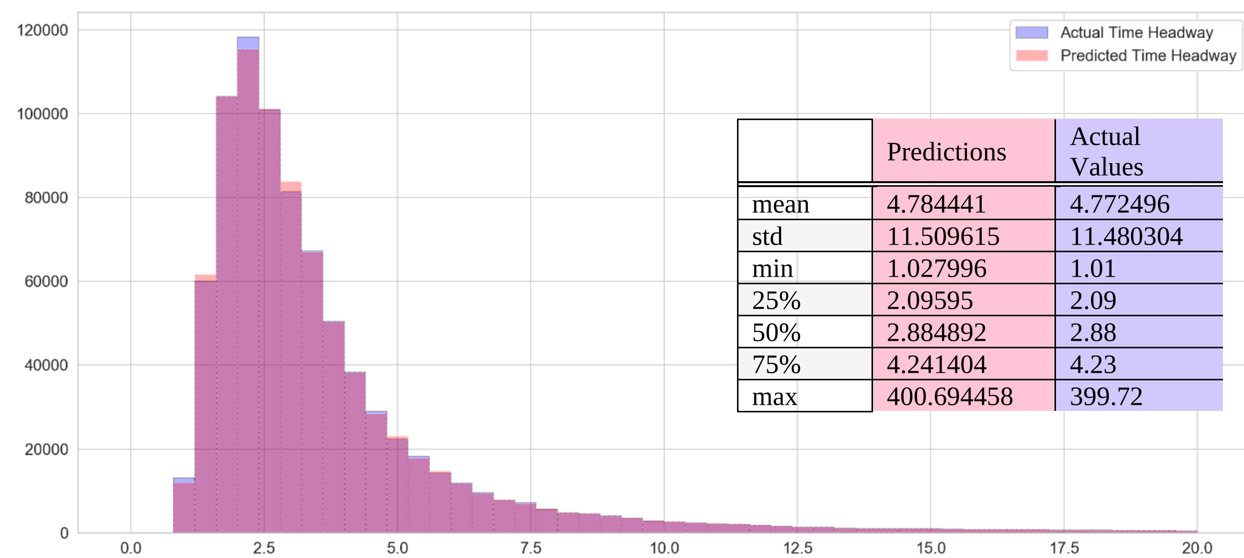
Figure 26 presents the predicted time headway of the US-101 highway over the actual time headway. Predicted headways again followed the actual values and overlapped almost all of them. It can be farther investigated that, the model can predict time headways nearly 100%

accurately which are below 100 seconds. Despite having the robustness and outliers, our model was able to detect the pattern and predict accordingly.

6.2. Prediction Analysis

6.2.1. Headway distribution

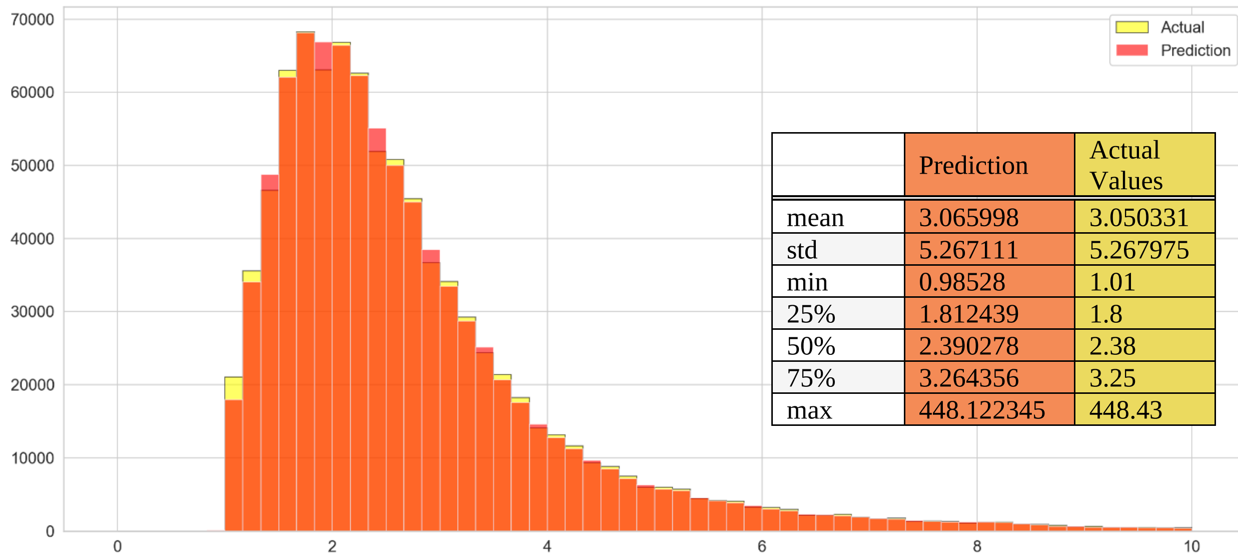
The predicted time headways were further analyzed visualizing the distribution and the correlations with the variables considered in the model. The I-80 and US-101 highways are different in their natures. Though the traffic density and flow rate were different, the distribution of the headway is similar in shape.



	Predictions	Actual Values
mean	4.784441	4.772496
std	11.509615	11.480304
min	1.027996	1.01
25%	2.09595	2.09
50%	2.884892	2.88
75%	4.241404	4.23
max	400.694458	399.72

Figure 27. Time headway distribution of I-80

The distribution of the predicted time headway follows the log-normal shape which matches the actual distribution. Figure 27 shows the actual and predicted time headway distribution of I-80. The graph shows that, both actual and predicted headways follow the shifted log-normal distribution. A similar trend can be found on the US-101 dataset. Figure 28 shows the actual and predicted time headway distribution of US-101. Both actual and predicted histograms also returned the log-normal shape. Comparing figure 27 and figure 28 we can comment that,



	Prediction	Actual Values
mean	3.065998	3.050331
std	5.267111	5.267975
min	0.98528	1.01
25%	1.812439	1.8
50%	2.390278	2.38
75%	3.264356	3.25
max	448.122345	448.43

Figure 28. Time headway distribution of US-101

though the mean and the variance of the distributions are different, they both followed a similar right-skewed distribution which is also known as log-normal distribution. This shape implies the basic nature of time headway on each roadway, lanes, and for each type of vehicle.

This distribution can also be visualized as a probability density function (pdf). Here we used kernel density estimation (KDE) to estimate the probability density function of time headway. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample.

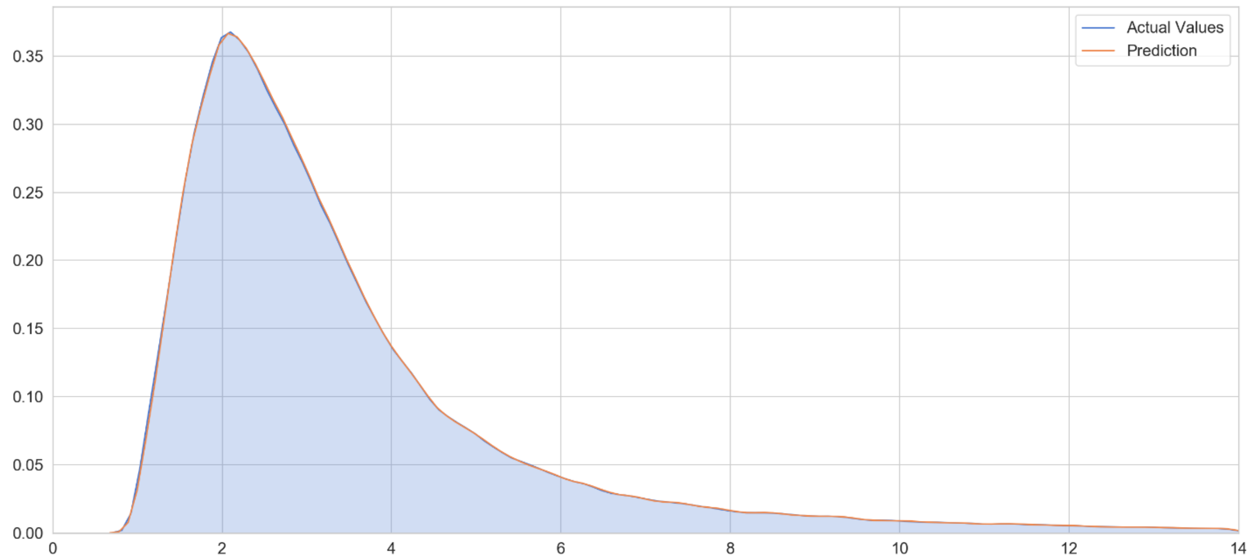


Figure 29. PDF of predicted and actual Time Headway (I-80)

From figure 29, it can be seen that the predicted density function overlaps the actual probability density function which indicates the accurate prediction of the time headway.

6.2.2. Correlations

To deeply investigate the predicted time headways, we visualized the correlations of the predicted time headway with the features that were used as inputs of the model. We wanted to explore how much our predicted values can be correlated with the explanatory variables and also the deviation in correlations while using the actual values. This will also give us a closer outlook about the important features of time headway too.

We would also like to show the Spearman's correlation coefficients of the important features of the headway with both predicted and actual headway. The Spearman rank-order correlation is a statistical procedure that is designed to measure the relationship between two variables on an ordinal scale of measurement. Spearman's correlation quantifies the degree to which ranked variables are associated by a monotonic function, meaning an increasing or decreasing relationship.

$$\rho_{rgX,rgY} = \frac{cov\ (rg_X, rg_Y)}{\sigma_{rg_X}, \sigma_{rg_Y}} \quad (25)$$

Here,

$\rho_{rgX,rgY}$ = The usual Spearman correlation coefficient

$cov\ (rg_X, rg_Y)$ = The covariance of the rank variables

$\sigma_{rg_X}, \sigma_{rg_Y}$ = The standard deviations of the rank variables.

Figure 30 shows the correlation of vehicle velocity with the actual time headway and predicted time headway in I-80 dataset. It can be seen that relationship between the time headway and velocity is similar for both actual and predicted values.

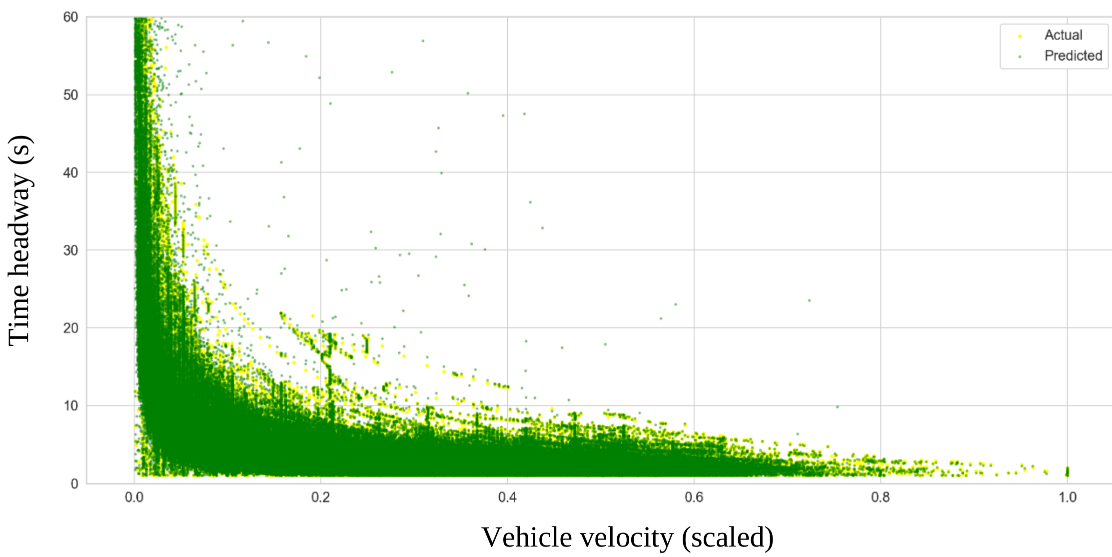


Figure 30. Correlation of vehicle velocity with time headway in I-80 (actual vs predicted)

Spearman's correlation coefficient of velocity with the predicted and actual time headway are -0.682 and -0.680 respectfully. Negative value implies the inverse relationship between time heaway and vehicle velocity. We can see the correlation coefficients for the actual and predicted values are closer to each other.

This implies velocity has a great impact on predicting time headway and thus it's considered one of the most important affecting features for analyzing headways. Figure 8 represents the same graph for US-101 dataset. By looking closely into figure 30 and figure 31 it can be interpreted that, the correlation shapes are not completely similar for two different highways.

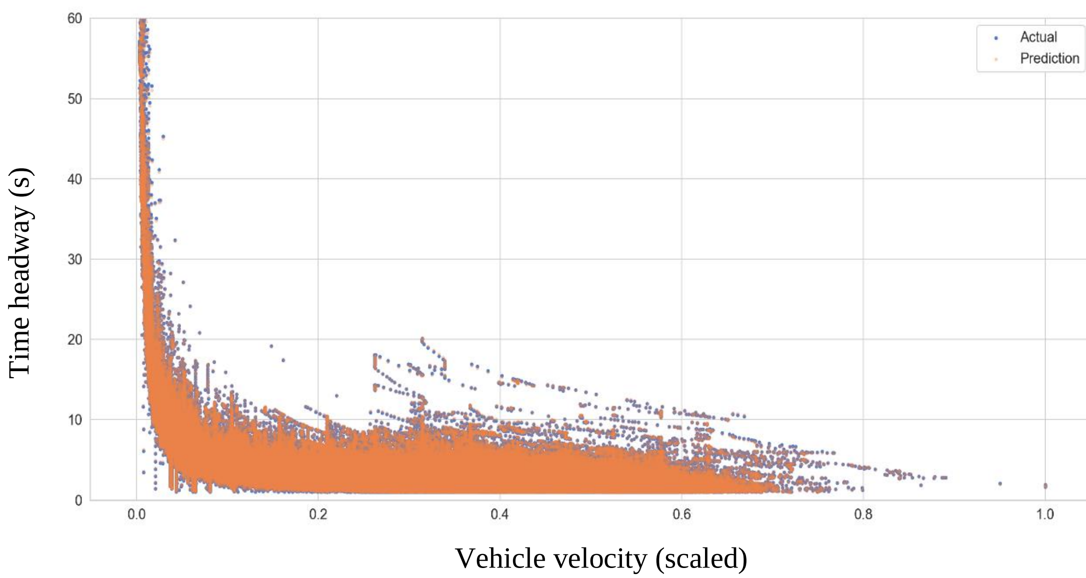


Figure 31. Correlation of vehicle velocity with time headway in US-101 (actual vs predicted)

It is because the average velocity and vehicle density were different on the two highways. But the correlation between actual and predicted time headway is completely similar in shape for an independent highway which is inversely proportional to the vehicle velocity. As discussed earlier, the velocity of the preceding vehicle (front vehicle) has been considered as another important feature for headway prediction,

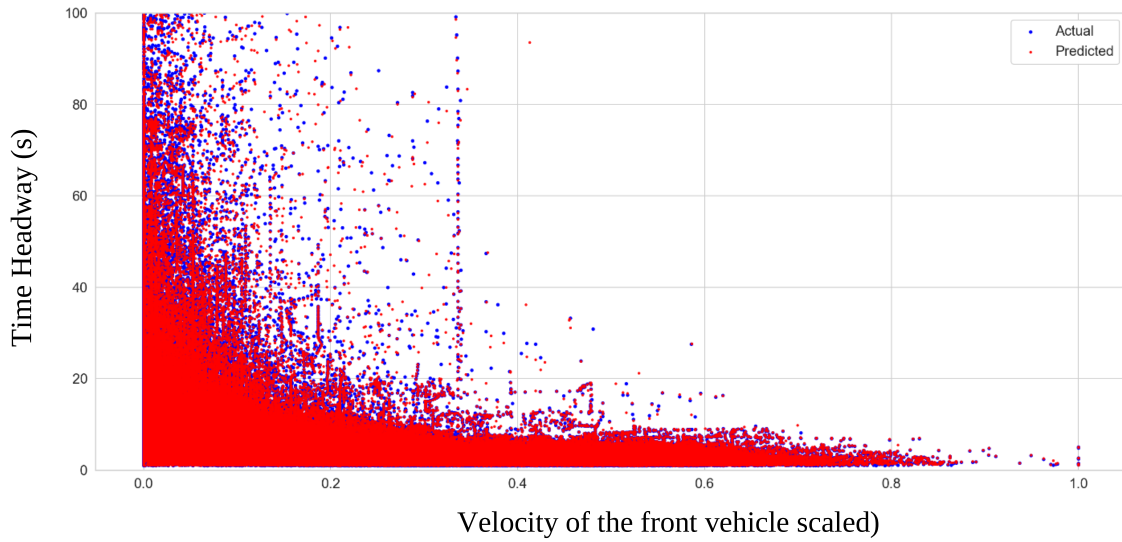


Figure 32. Correlation of Preceding (front vehicle) velocity with time headway in I-80 (actual vs predicted)

Figures 32 and 33 explain the correlation of the velocity of the preceding vehicle with time headway for different road networks. It can be visualized that the correlation of preceding velocity with predicted time headway matches with the actual time headway. The Spearman's correlation coefficients of the velocity of front vehicle with the predicted and actual time headways are similar which is -0.588. As the coefficients are similar for the predicted and actual time headway it means our model did well to extract this relationship between the headway and the velocity of the front vehicle.

In figure 33, the same relationship has been shown for US-101. As US-101 did not have too many varieties in velocity and headway like I-80, the shape is different. But the shapes of the correlation of actual and prediction values with the preceding velocity are very similar

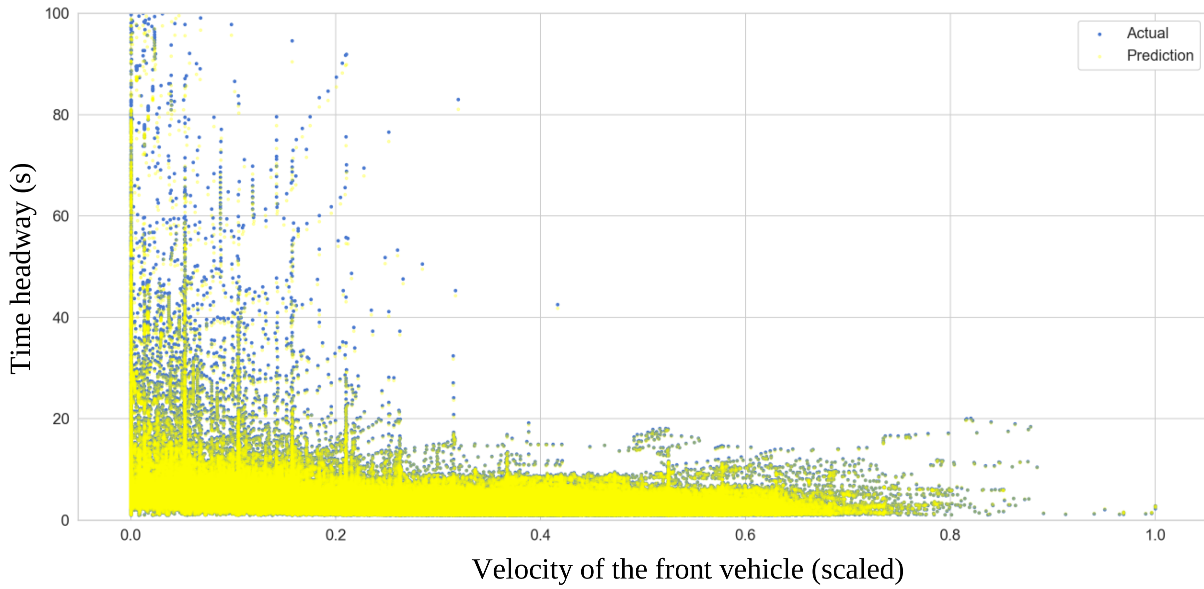


Figure 33. Correlation of Preceding (front vehicle) velocity with time headway in US-101 (actual vs predicted)

The flow rate of a road segment has been considered as an important feature for headway prediction in this study. Its importance has been validated by visualizing the correlation graphs for both highways. This is a non-linear complex relationship that was captured smoothly by our model.

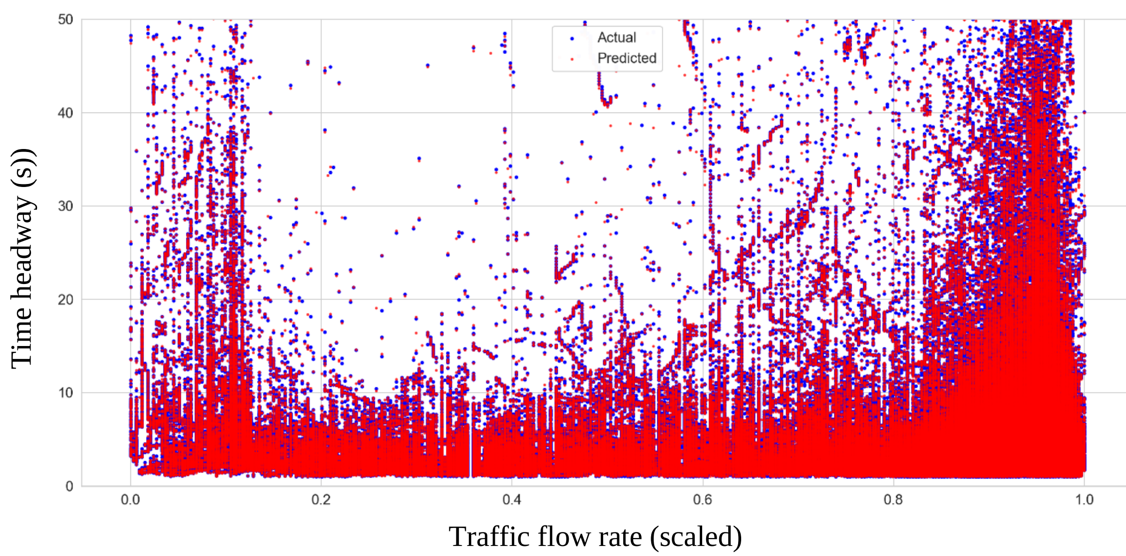


Figure 34. Correlation of time headway (actual vs predicted) with flow rate in highway I-80

The correlation for highway I-80 is represented by a scattered plot in figure 34. The Spearman's correlation coefficiants for the flow rate with the actual and predicted headways are 0.09 and 0.085. We can see that our model was able to find the relationship pattern very well though highway I-80 had a lot of variations in data. Figure 35 represents the correlation of time headway with different flow rates in the US-101 highway. As US-101 data provided less variation than highway I-80, the model could extract the complex relationship more accurately.

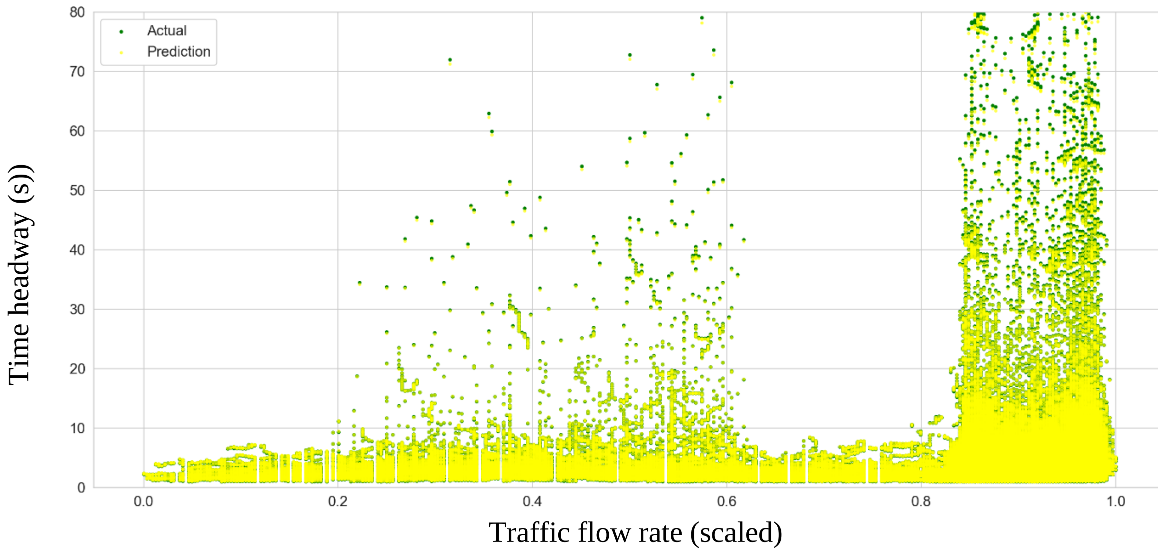


Figure 35. Correlation of time headway (actual and predicted) with flow rate in US-101

6.3. Model Comparison

There has been a lot of analysis in headway distribution through probabilistic methods. Recently some new and improved methods have been implemented to predict time headway which is described in chapter 2 (Literature Review). Among them, RVM, SVM, ANN, and some hybrid models have been mostly used. In those prediction models, there were some shortcomings which are as follows:

1. Almost all of the models lacked in high ranged headway prediction. It means not predicting greater time headways (>100 seconds). Also, the robustness of the headway was neglected.
2. The variance of the headway was not adequate to generalize the model. As most of the models were probabilistic models, high variance data was difficult to adapt and predict. By introducing novel deep learning techniques we have overcome this issue as deep learning and machine learning algorithms need more variations in data so that the machine can learn from different scenarios and adopt the stochastic robustness nature.
3. Not having enough data size. One of the most important features of artificial intelligence is that it can learn from a huge dataset and it is often recommended to use a bigger dataset for training.
4. Not validating on different road segments. Training and validating different sets of datasets provide the picture of the overall accuracy and learning ability of the model.

In this study previously faced problems and the above-listed lackings were overcome. The maximum predicted time headway of our model is close to 400 seconds (over 5 minutes). A deep learning approach has been made for the first time to predict and analyze traffic time headway. Deep learning models can capture complex relationships from big data which makes it easier to use numerous stochastic data. Moreover, LSTM networks are specialized in time series analysis and predictions. It has proved to be one of the finest predicting algorithms in recent years. Though it has already been introduced to predict traffic velocity, time, etc., still there are some sectors in traffic engineering where these kinds of deep learning algorithms haven’t been implemented yet. Time headway is one of those important parameters of traffic engineering which did not have any reliable predicting model with the use of new emerging technologies.

A comparison has been made between the result of this study and some recent headway predictions in Table 7. From the table, it can be seen that new techniques have been used recently which lowered the RMSE and MAPE values gradually over time.

Table 7. One step prediction comparison of time headway

Author & year	**Data usage**	**Algorithm(s)/ Model(s)**	**RMSE**	**MAPE (%)**	**Comment(s)**
Moridpour et al. (2014)	Highway I-80 (similar to this study)	Probabilistic method (Chi-Square tests)	N/A	N/A	Only distribution; no prediction.
Tong et al. (2000)	16,976 samples	ANN	N/A	6.8	Discharge headway estimation.
		Regression	N/A	10.4	
Wu et al. (2016)	Based on only smart card data	RVM	1.494	15.39	Bus arrival time estimation using headway prediction without considering other traffic parameters.
		SVM	1.5598	16.29	
		GA-SVM	4.0833	31.77	
		KF	9.7916	74.57	
		KNN	3.111	15.92	
		SNN	2.852	24.8	
Yu et al. (2016)	Transit smart card data	LS-SVM	1.3439	5.42	Fails to predict the extreme high bus headways; Bus arrival time prediction.
		KNN	2.0933	6.71	
		ANN	7.2992	9.99	
		RF	2.7804	8.7	
		GPR	4.1122	10.24	
Guo et al. (2016)	Four segments of Shanghai-Nanjing freeway (6,202 samples)	The maximum-likelihood estimation	N/A	N/A	Headway distribution in mixed traffic scenarios but no prediction.
Roy et al. (2018)	8,000 samples	Probabilistic methods	N/A	N/A	Modeling headway distribution on two-lane rural highways with mixed traffic.; no prediction.
Current study (2020)	4.1 million samples	LSTM	**0.088 (I-80)**	**2.36 (I-80)**	Modeling time headway distribution with one step prediction in mixed traffic scenarios.
			0.07 (US-101)	**1.89 (US-101)**	

The RMSE and MAPE values of the current study are the lowest so far in predicting time headway. This accuracy outperforms the previous models which show another dominant display of deep neural networks in the field of ITS.

CHAPTER VII

CONCLUSION & FUTURE RECOMMENDATIONS

Traffic congestion has become one of the serious problems in most of the countries worldwide. Increased waiting time for traffic leads to an overall loss in productivity and economy. For example, Americans lost an average of 97 hours a year due to congestion, costing them nearly $87 billion in 2018, an average of $1,348 per driver [1]. At the global level, Moscow topped the list of the world's most gridlocked cities (210 hours lost due to congestion) when weighting for population, followed by Istanbul, Bogota, Mexico City, and São Paulo. So, to eradicate this issue proper planning and modeling have become mandatory tasks nowadays. Traffic jam occurs due to the heterogeneity of the time headways which is the various arrival time of the vehicles. Thus scientists have been working on headway modeling for a long time. Appropriate headway models can give us insights about traffic congestion, pointing out the cause and solutions to overcome this problem. In congested flow, drivers maintain shorter headways (larger time headways). This condition leads to skewness in headway distribution. The log-normal distribution was the best fit for the time headway distribution.

There have been introduced a lot of probabilistic and mathematical models to analyze and predict time headway. Those models were fine but could not predict accurately for various traffic scenarios. It is because time headway is stochastic and it does not only depend on traffic conditions, it heavily depends on the driver's behavior during various traffic scenarios.

Mathematical models can not accurately describe human behavior. Recently, the invention and provision of artificial intelligence have enabled us to visualize and learn more about human nature. These new techniques are replacing the ancient mathematical and probabilistic models in various other disciplines as well. Deep learning algorithms, especially *Long Short-Term Memory* networks have been providing high accuracy prediction on time series analysis in various fields. This network has been also used in predicting other parameters of transportation such as traffic flow, vehicle velocity, etc. This technique has been introduced for the first time to analyze and predict time headway in this study. The prediction error has been decreased to a great extent using this deep learning method.

After analyzing, several facts were revealed regarding time headway. Vehicles having higher length and width are considered as heavy vehicles. Heavy vehicles (Class-3 vehicles in this study) have greater time headway than passenger cars. This is due to operational limitations (acceleration, deceleration, etc) of heavy vehicles. The effect of the positions of heavy vehicles on time headway has been explored and visualized. Not only do the drivers of heavy vehicles maintain higher time headway, but also the vehicles around a heavy vehicle tend to maintain greater time headway. This behavior explains the safety concerns of the drivers. Thus velocity and the velocity of the preceding vehicle (front vehicle) have been considered as the two most important features for time headway. The behavior of the drivers also changes with the different traffic flow. For example, in a freeway, drivers do not follow their front vehicles that much whereas, during a congested flow, the driver follows the front vehicle to a great extent. This situation is explained by the car-following behaviors of the drivers. Time headway differences in each lane were also visualized through separated lane-wise and vehicle-type wise analysis. Though there are differences in headway values in each lane, and for each vehicle type, the distribution of the time headway was similar in each case (shifted log-normal distribution). This describes the similar nature of time headway.

Though our model provided high accuracy predictions, there were some flaws that can be easily modified and updated in the future. Future recommendations can be pointed out as follows:

1. The data we used in this study was not up to date (gathered from 2005). As traffic conditions and human behaviors are constantly changing, it is best to build a model using the most recent instantaneous data.
2. If the instantaneous data can be gathered while traveling, the instantaneous predictions can be also made through our model scripts which will enable drivers to visualize the overall traffic conditions and vehicle movements instantly. This visualization can prevent traffic accidents and also provide the accurate arrival time of the vehicles instantly.
3. The scripts can also be used in modeling trajectories of Connected Autonomous Vehicles (CAV) and for more advanced CAV modeling such as how CAVs form platoons, interact across lanes in freeway merge areas. This opens a new door to explore research on CAVs using time headway prediction.
4. Weather conditions of the road segment were not considered in this study. It can be further added to the data to get more precise results and human behaviors.
5. The model can also be used in describing lane-changing behaviors.

Time headway analysis can provide both macroscopic and microscopic analysis which is very useful for interpreting different parameters of traffic engineering.

REFERENCES

1. Adams, W.F., 1936. Road traffic is considered as a random series. J. Inst. Civil Engin. 4 (1), 121–130.
2. Greenberg, I., 1966. The log-normal distribution of headways. Austral. Road Res. 2 (7), 14–18.
3. Tolle, J.E., 1976. Vehicular headway distributions: testing and results. Transp. Res. Rec. 567, 56–64.
4. Luttinen RT. Properties of Cowan's M3 headway distribution. Transp Res Record 1999; 1678: 189–196.
5. Hagring, O., 1996. The use of the Cowan M3 distribution for modeling roundabout flow. Traffic Eng. Control 37 (5), 328–332.
6. Roess et al. (2004). Traffic Engineering. Prentice Hall Polytechnic Series in Traffic Engineering, United State.
7. Mei M and Bullen AGR. Lognormal distribution for high traffic flows. Transp Res Record 1993; 1398: 125–128.
8. Al-Ghamdi AS. Analysis of time headways on urban roads: case study from Riyadh. J Transp Eng: ASCE 2001; 127: 289–294.
9. Faderico Pascucci (2012-2013). Mixed Time Headway Distributions: Endogenous Effects in Motorways and Rural Two-lane Two-way Roads. Ph.D. Thesis, Department of Structural and Transportation Engineering, University of Padua

10. T. V. Arasan and R. Z. Koshy, “Headway distribution of heterogeneous traffic on urban arterials,” Journal of the Institution of Engineers, vol. 84, pp. 210-215, 2003.

11. Krbálek, M., Šeba, P., Wagner, P., 2001. Headways in traffic flow: remarks from a physical perspective. Phys. Rev. E 64 (6), 066119.

12. Nishinari, K., Treiber, M., Helbing, D., 2003. Interpreting the wide scattering of synchronized traffic data by time gap statistics. Phys. Rev. E 68 (6), 067101.

13. Bham, G.H., Ancha, S.R.R. (2006). Statistical models for preferred time headway and time headway of drivers in steady state car-following. Applications of Advanced Technology in Transportation: Proceedings of the Ninth International Conference, Chicago, Illinois, USA, 13-16 August 2006, pp. 344-349.

14. Zwahlen, H.T., Oner, E., Suravaram, K.R. (2007). Approximated headway distributions of free-flowing traffic on Ohio freeways for work zone traffic simulations. Transportation Research Record 1999, pp. 131-140.

15. Ye F and Zhang Y. Vehicle type–specific headway analysis using freeway traffic data. *Transp Res Record* 2009; 2124: 222–230.

16. D. Kong and X. Guo, “Analysis of vehicle headway distribu-tion on multi-lane freeway considering car-truck interaction,” *Advances in Mechanical Engineering*, vol.8,no.4,pp.1–12,2016.

17. Penmetsa P, Ghosh I, Chandra S (2015) Evaluation of performance measures for two-lane intercity highways under mixed traffic condition. J Transp Eng-ASCE 141(10):04015021–04015028. https:// doi.org/10.1061/(ASCE)TE.1943-5436.0000787.

18. Rupali, R.; Saha, P. Headway distribution models of two-lane roads under mixed traffic conditions: A case study from India. *Eur. Transp. Res. Rev.* 2018,10, 1–12.

19. Chen, X., Li, L., Zhang, Y., 2010a. A Markov model for headway/spacing distribution of road traffic. IEEE Trans. Intell. Transp. Syst. 11 (4), 773–785.

20. Moridpour, S. 2014. Evaluating the time headway distributions in congested highways. Journal of Traffic and Logistics Engineering, 2(3): 224-229. doi: 10.12720/jtle.2.3.224-229.

21. F. Ho and P. Ioannou, “Traffic flow modeling and control using artificial neural networks,” IEEE Control Syst. Mag., vol. 16, pp. 16–27, Oct. 1996.

22. Tong, H.Y., Hung, W.T., 2002. Neural network modeling of vehicle discharge headway at signalized intersection: model descriptions and the results. Transportation Research Part A 36 (1), 17–40.

23. Y. Bin, Y. Zhongzhen, and Y. Baozhen, “Bus arrival time prediction using support vector machines,” J. Intell. Transp. Syst., vol. 10, no. 4, pp. 151–158, 2006.

24. B. Yu, W. H. K. Lam, and M. L. Tam, “Bus arrival time prediction at bus stop with multiple routes,” Transp. Res. C, Emerg. Technol., vol. 19, no. 6, pp. 1157–1170, Dec. 2011.

25. S. I.-J. Chien, Y. Ding, and C. Wei, “Dynamic bus arrival time prediction with artificial neural networks,” J. Transp. Eng., vol. 128, no. 5, pp. 429–438, 2002

26. H. Yu, D. Chen, Z. Wu, X. Ma, and Y. Wang, “Headway-based bus bunching prediction using transit smart card data,” Transp. Res. Part C Emerg. Technol., vol. 72, pp. 45–59, 2016.

27. Artificial Intelligence, Machine Learning and Deep Learning Pariwat Ongsulee Department of Computer Science Faculty of Science, Siam University Bangkok, Thailand.

28. S. Islam, “Estimation of annual average daily traffic (AADT) and miss-ing hourly volume using artificial intelligence,” M.S. thesis, ClemsonUniv., Clemson, SC, USA, 2016.

29. Laha, A.K., Putatunda, S., 2018. Real time location prediction with taxi-GPS data streams. Transp. Res. Part C: Emerging Technol. 92, 298–322.

30. Ba Y, Zhang W, Wang Q, Zhou R, Ren C. Crash prediction with behavioral and physiological features for advanced vehicle collision avoidance system. Transp Res Part C Emerg Technol. 2017;74:22–33. doi: 10.1016/j.trc.2016.11.009. [CrossRef] [Google Scholar].

31. Hasnat, M.M.; Hasan, S. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. Transp. Res. Part C Emerg. Technol. 2018, 96, 38–54. [CrossRef]

32. A. Cheng, X. Jiang, Y. Li, C. Zhang, H. Zhu, Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method, Physica A 466 (2017) 422–434.

33. F. Guo , J.W. Polak , R. Krishnan , Predictor fusion for short-term traffic forecast- ing, Transp. Res. Part C Emerg. Technol. 92 (2018) 90–100 .

34. Z. Liu, Y. Liu, Q. Meng, and Q. Cheng, “A tailored machine learning approach for urban transport network flow estimation,”Transp. Res. C, Emerg. Technol., vol. 108, pp. 130–150, Nov. 2019.doi: 10.1016/j.trc.2019.09.006.

35. L. Li, X. Qu, J. Zhang, Y. Wang, and B. Ran, ‘‘Traffic speed prediction for intelligent transportation system based on a deep feature fusion model,’’ J. Intell. Transp. Syst., vol. 23, no. 6, pp. 605–616, Nov. 2019.

36. Gao, X. and G. M. Lee (2019). Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning. Computers & Industrial Engineering 128, 60–69.

37. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F.Y. Traffic Flow Prediction With Big Data: A Deep Learning Approach. IEEE Trans. Intell. Transp. Syst. 2014, 16, 1–9. [CrossRef]

38. Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. Transportation Research Part C: Emerging Technologies 79, 1–17.

39. Wu, Y., Tan, H., Qin, L., Ran, B., Jiang, Z., 2018. A hybrid deep learning based traffic flow prediction method and its understanding. Transport. Res. Part C Emerg. Technol. 90, 166–180.

40. Dai, X., Fu, R., Zhao, E., Zhang, Z., Lin, Y., Wang, F.Y., Li, L., 2019. DeepTrend 2.0: a light-weighted multi-scale traffic prediction model using detrending. Transp. Res. Part C: Emerg. Technol. 103, 142–157.

41. Liu, L., Chen, R.C., 2017. A novel passenger flow prediction model using deep learning methods. Transport. Res. Part C: Emerg. Technol. 84, 74–91.

42. Zhang, Z., Li, M., Lin, X., Wang, Y., He, F. 2019, Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. Transport. Res. Part C: Emerg. Technol. 105, 297-322.

43. Yu, R., Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, Deep Learning: A Generic Approach for Extreme Condition Traffic Forecasting, 2017, pp. 777–785.

44. Ma, X., Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang. Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. Sensors, Vol. 17, No. 4, 2017.

45. Wang, J., Q. Gu, J. Wu, G. Liu, and Z. Xiong, Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method. 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, pp. 499–508.

46. Dabiri, S. & Heaslip, K. 2018, "Inferring transportation modes from GPS trajectories using a convolutional neural network", Transportation Research Part C: Emerging Technologies, vol. 86, pp. 360-371.

47. C. Mao and Z. Shen. "A reinforcement learning framework for the adaptive routing problem in stochastic time-dependent network". Transportation Research Part C: Emerging Technologies, 93, 179-197, 2018.

48. Jiang Z, Fan W, Liu W, Zhu B, Gu J (2018) Reinforcement learning approach for coordinated passenger inflow control ofurban rail transit in peak hours. Transp Res Part C EmergTechnol 88:1–16

49. Ma, X., H. Yu, Y. Wang, and Y. Wang. Large-Scale Transportation Network Congestion Evolution Prediction using Deep Learning Theory. PLoS One, Vol. 10, No. 3, 2015, pp. 1–17.

50. Wang, J., Q. Gu, J. Wu, G. Liu, and Z. Xiong, Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method. 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, pp. 499–508.

51. Yu, R., Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, Deep Learning: A Generic Approach for Extreme Condition Traffic Forecasting, 2017, pp. 777–785.

52. Zhao, Z., W. Chen, X. Wu, P. C. Y. Chen, and J. Liu. LSTM Network: A Deep Learning Approach for Short term Traffic Forecast. IET Intelligent Transport Systems, 2017.
53. Wu, Y., H. Tan, L. Qin, B. Ran, and Z. Jiang. A Hybrid Deep Learning Based Traffic flow Prediction Method and its Understanding. Transportation Research Part C: Emerging Technologies, Vol. 90, 2018, pp. 166–180.
54. Kuang, H., Z. P. Xu, X. L. Li, and S. M. Lo. 2017. "An Extended Car-Following Model Accounting for the Average Headway Effect in Intelligent Transportation System." Physica A: Statistical Mechanics and its Applications 471: 778–787.
55. Zhang, X.; Sun, J.; Qi, X.; Sun, J. Simultaneous modeling of car-following and lane-changing behaviors usingcdeep learning. Transp. Res. C Emerg. Technol. 2019, 104, 287–304. [CrossRef]
56. Wang, X., R., Jiang, L., Li, Y., Lin, X., Zheng, and F. Y. Wang. Capturing Car-Following Behaviors by Deep Learning. IEEE Transactions on Intelligent Transportation Systems, Vol. 99, 2017, pp. 1–11.
57. A. Khodayari, A. Ghaffari, R. Kazemi, and R. Braunstingl, "A modified car-following model based on a neural network model of the human driver effects", IEEE Transactions on Systems, Man, and Cybernetics, vol. 42, no. 6, pp. 1440–1449, 2012.
58. Lee, S., Ngoduy, D., Keyvan-Ekbatani, M., 2019. Integrated deep learning and stochastic car-following model for traffic dynamics on multi-lane freeways. Transportation Research Part C submitted.
59. Yanjie Duan, Yisheng Lv, and Fei-Yue Wang, "Travel time prediction with LSTM neural network," in 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, nov 2016, pp. 1053–1058.

60. Nguyen, H., C. Bentley, L. M. Kieu, Y. Fu, and C. Cai. Deep Learning System for Travel Speed Predictions on Multiple Arterial Road Segments. Transportation Research Record, Vol 2673, Issue 4, 2019, p. 03-6.
61. Wang, J., Chen, R., He, Z.: ‘Traffic speed prediction for urban transportation network: A path based deep learning approach’Transp. Res. Part C Emerg. Technol., 2019, 100, pp. 372–385.
62. Y. Gu, W. Lu, L. Qin, M. Li, and Z. Shao, ‘‘Short-term prediction of lane level traffic speeds: A fusion deep learning model,’’ Transp. Res. C, Emerg. Technol., vol. 106, pp. 1–16, Sep. 2019.
63. Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, “Lstm-based traffic flow prediction with missing data,” Neurocomputing, vol. 318, pp. 297–305, 2018.
64. Xu, C., Ji, J., Liu, P., 2018. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. Transp. Res. Part C Emerg. Technol. 95, 47–60. https://doi.org/10.1016/j.trc.2018.07.013
65. Schultz, M. and S. Reitmann. 2018. “Machine learning approach to predict aircraft boarding”. J. of Transp. Res. Part C: Emerg. Techn. (review)
66. Hochreiter, Sepp; Schmidhuber, Jürgen (1997-11-01). "Long Short-Term Memory". Neural Computation. 9 (8): 1735–1780.
67. Yang, Shuguan, et al. "A deep learning approach to real-time parking occupancy prediction in transportation networks incorporating multiple spatio-temporal data sources." Transportation Research Part C: Emerging Technologies 107 (2019): 248-265.
68. Stathakis, D. "How many hidden layers and nodes?." International Journal of Remote Sensing 30.8 (2009): 2133-2147.

69. Ma, Xiaolei, et al. "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data." Transportation Research Part C: Emerging Technologies 54 (2015): 187-197.

70. Huang, Xiuling, Jie Sun, and Jian Sun. "A car-following model considering asymmetric driving behavior based on long short-term memory neural networks." Transportation research part C: emerging technologies 95 (2018): 346-362.

71. Li, Li, and Xiqun Michael Chen. "Vehicle headway modeling and its inferences in macroscopic/microscopic traffic flow theory: A survey." *Transportation Research Part C: Emerging Technologies* 76 (2017): 170-188.

72. Kumar, V. M., and S. K. Rao. "HEADWAY AND SPEED STUDIES ON TWO-LANE HIGHWAYS." *Indian highways* (1998).

73. Highway capacity manual 2000. Washington, DC: Transportation Research Board, National Research Council, 1134 pp.

74. Bian, Y., Zheng, Y., Li, S.E., Wang, Z., Xu, Q., Wang, J. and Li, K., 2018, November. Reducing time headway for platoons of connected vehicles via multiple-predecessor following. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1240-1245). IEEE.

BIOGRAPHICAL SKETCH

Saumik Sakib Bin Masud received his B.Sc. degree in Civil Engineering from Bangladesh University of Engineering and Technology, Bangladesh in 2017 and joined in Transportation Laboratory (Dr. Noruzoliaee's research group) at the University of Texas Rio Grande Valley in 2019 as a graduate research assistant, pursuing his masters in Civil Engineering side by side. Previously he worked for construction industry for 2 years. His research interests are in Intelligent Transportation System (ITS), traffic forecasting, demand modeling, Connected Autonomous Vehicles (CAVs), etc. Saumik Sakib B. Masud is recently awarded: Outstanding Graduate Student of the year 2020. In 2019 he got the "Good Grade" scholarship and issued his Engineer In Training (E.I.T) certification from the Texas Board of Professional Engineers (TBPE). He received his Masters of Science degree in Civil Engineering in December of 2020 and joining in the University of Kansas as a Ph.D. student in the Department of Civil Engineering. He can be reached at saumiksakib@gmail.com.