

Application of Long Short Term Memory Networks for Long- and Short-term Bus Travel Time Prediction

Osama A. Osman, PhD

ORCID: <https://orcid.org/0000-0002-5157-2805>

Research Associate

Center for Sustainable Mobility

Virginia Tech Transportation Institute, Blacksburg, Virginia 24060

Email: osman@vtti.vt.edu

Hesham Rakha, PhD, P.Eng (Corresponding Author)

ORCID: <https://orcid.org/0000-0002-5845-2929>

Samuel Reynolds Pritchard Professor of Engineering and Director

Center for Sustainable Mobility

Virginia Tech Transportation Institute, Blacksburg, Virginia 24060

Email: hrakha@vtti.vt.edu

Archak Mittal, PhD

ORCID: <https://orcid.org/0000-0001-6186-4513>

Research Scientist

Ford Motor Company

20000 Rotunda Drive

Dearborn, Michigan 48124

Email: amittal9@ford.com

ABSTRACT

This study introduces a comparative analysis of two deep learning (multilayer perceptron neural networks (MLP-NN) and the long short term memory networks (LSTM)) models for transit travel time prediction. The two models were trained and tested using one-year worth of data for a bus route in Blacksburg, Virginia. In this study, the travel time was predicted between each two successive stations to all the model to be extended to include bus dwell times. Additionally, two additional models were developed for each category (MLP of LSTM): one for only segments including controlled intersections (controlled segments) and another for segments with no control devices along them (uncontrolled segments). The results show that the LSTM models outperform the MLP models with a RMSE of 17.69 sec compared to 18.81 sec. When splitting the data into controlled and uncontrolled segments, the RMSE values reduced to 17.33 sec for the controlled segments and 4.28 sec for the uncontrolled segments when applying the LSTM model. Whereas, the RMSE values were 19.39 sec for the controlled segments and 4.67 sec for the uncontrolled segments when applying the MLP model. These results demonstrate that the uncertainty in traffic

Osman, O. A.; H. Rakha, and A. Mittal

conditions introduced by traffic control devices has a significant impact on travel time predictions. Nonetheless, the results demonstrate that the LSTMN is a promising tool that can have the ability to account for the temporal correlation within the data. The developed models are also promising tools for reasonable travel time predictions in transit applications.

Keywords: Travel Time Prediction, Deep Learning, Long Short Term Memory Networks, transit, temporal correlation

INTRODUCTION

The advent of advanced vehicle and smartphone technologies has changed the face of transportation and will continue to do so. The availability of unprecedented amounts of real-time information enabled by these technologies has made it possible for travelers to proactively plan their trips, alter their choices and decisions before and during their trips, and act as assisting agents to other travelers in a smart and intelligent transportation system (ITS). Advanced traveler information systems (ATIS) is one application of ITS. In ATIS, provision of real-time information is supported by a hierarchical process that starts with the collection of sensor data which goes through comprehensive data processing, before the resulting synthesized information is disseminated in real-time to travelers through several relaying devices/services, such as dynamic message signs, 511 calling services, radio, and smart phones, among others. Every step in this hierarchical process is important and must be performed accurately and in a timely manner to result in reliable information that can help travelers make informed and timely decisions. Collected data must be obtained via reliable and calibrated sensors; data processing and information synthesis require well-structured and well thought out calculations and models; and information dissemination requires reliable communication technologies.

Transit agencies have been relying on ATIS for effective trip planning and improved service. In doing so, data collection devices have been installed on buses for collection of data that are later used to aid decision-making by both travelers and the transit agency. Travel time and bus arrival times are two pieces of information passengers/travelers need for departure time, route, and mode choice decisions. Using the collected data, these pieces of information can be estimated/predicted and travelers can effectively plan their trips. While there have been many attempts to predict bus travel and arrival times, this task is complicated and not easy to achieve because of the involvement of many stochastic variables in the process. Thus, it is very important to account for such stochasticity to realize reasonable predictions. This study attempts to develop a bus travel time prediction model that takes into consideration such stochasticity through accounting for the temporal correlation between the timeseries observations of the controlling variables.

LITERATURE REVIEW

Traffic forecasting is one of the important components for successful implementation of ATIS. Specifically, prediction of travel time - hence Estimated Arrival Time (ETA) - is a critical building block for determination of the shortest and/or most eco-friendly routes in navigation systems. Additionally, travel time is better perceived by travelers and more helpful in their decision-making process. Predicting travel time minimizes the effect of uncertainty in traffic conditions when optimizing routes. There is a plethora of literature pertaining to travel time predictions for ATIS and traffic management, which can be categorized as model-based or data driven.

Osman, O. A.; H. Rakha, and A. Mittal

Model Based Approaches

Model-based approaches are defined as any techniques that incorporate simulation as a core component for travel time prediction. This means that these approaches rely on traffic flow models to describe traffic propagation on roadways for prediction of travel time. Model-based approaches are flexible, as they can account for traffic dynamics as model parameters and can measure effects of future changes, such as expansions and developments [1]. These approaches are also known to be robust and efficient, especially for real-time online applications. Therefore, they require comprehensive data pre-processing and powerful computational methods to maintain reliable performance. Model-based approaches can be categorized according to their level of detail as macroscopic, mesoscopic, cellular automata (CA), and microscopic. Regardless of the category, model-based approaches can predict travel times over a network for a given horizon even with relatively small number of detectors [1].

Data Driven Approaches

With the introduction of many technological advancements in transportation and the vast availability of data, attention has shifted towards data-based (data-driven) approaches for ATIS applications. Data-driven approaches predict travel times by assuming that traffic patterns will remain similar to the historical data [2]. These approaches rely on historical data collected from different sources that require a significant pre-processing effort before future travel time predictions are obtained. Because of the huge amount of data involved, data-driven approaches have considerably high computational costs. Unlike model-based approaches, the real-time applicability of the data driven approaches is limited unless specific treatments are considered [3, 4]. Data driven approaches can be categorized as naïve, parametric, or non-parametric.

Naïve Approaches

Naïve approaches are defined as those that do not rely on a defined model structure or parameters to predict travel times [5]. These approaches are rather subjective as they rely on only the collected data and simple well-known physical relationships [6]. Because of their simplicity, naïve approaches have a low computational cost, and are therefore widely implemented in practice. However, their accuracy is not highly reliable [6]. Examples of naïve approaches include instantaneous and historical averages. Instantaneous travel time prediction approaches are based on the main assumption that prevailing traffic conditions will remain unchanged. Given this assumption, instantaneous approaches can provide reasonable travel time predictions only if traffic conditions are extremely stationary and homogeneous over time. Relying on historical averages is another naïve approach for obtaining travel time predictions. Similar to the instantaneous predictions, historical averages are usually used to compare the performance of other more superior and sophisticated methods.

Parametric Approaches

Parametric approaches predict travel times based on pre-structured relationships (models), while some parameters in the models may be estimated from the data [2, 5]. Examples of parametric approaches include time series approaches and the previously discussed model-based approaches. Time series approaches model travel time as a function of its past observations. In time series approaches, the travel time prediction problem should be treated categorically based on the season and other factors to guarantee accuracy [5]. Time series approaches have long been researched and implemented for on-line applications [7-10]. Examples of time series approaches include linear

Osman, O. A.; H. Rakha, and A. Mittal

regression, Kalman filtering, and Auto-Regressive Integrated Moving Average (ARIMA), among others.

Kalman filter performs travel time predictions based on a continuous update of the traffic state variables and by assuming that the future traffic state is a function of the current state and the estimated state in the previous time step [2, 5]. In a study by Chein et al. [11], a Kalman-filtering-based approach was developed to predict travel time for selected OD pairs in a network over different prediction horizons. In another study by Park and Rilett [12], automatic vehicle identification (AVI) data were used to develop a Kalman-filtering-based travel time prediction model.

Linear regression assumes that travel time is a linear function of covariates. Because of the simple linear form, linear regression prediction models are known for fast predictions with good prediction accuracy in most cases [2, 5]. As an example, Kwon et al. [13] developed a travel time prediction model based on linear regression with a step-wise variable selection method. The linear regression model was tested considering departure time and the day of the week as covariates and for travel time prediction over prediction horizons ranging from a few minutes to one hour. The results indicated that accurate predictions over periods of up to 20 minutes can be obtained using the current traffic states, while to maintain the same performance over longer prediction horizons, historical data should be used. In another study by Rice and Van Zwet [14], a linear-regression-based travel time prediction model with time-varying coefficients was developed. This study showed that using linear regression can achieve reasonably accurate travel time prediction over a one-hour prediction horizon with a root mean square value of 10 minutes or less for a trip of 48 miles. This finding is different from Kwon et al.'s [13] conclusions, which could be related to the fact that this study used time-varying coefficients in their model. These findings confirm that the performance of this type of prediction model is subject to many factors and requires careful treatment.

The ARIMA Model treats traffic data in travel time prediction problems as a series of sequential and noisy observations that require noise minimization. This perspective of analyzing data has enabled many researchers to use the ARIMA model as a way to account for the stochastic traffic nature in prediction problems [15-17]. Billing and Yang [18] applied the ARIMA model for short-term travel time prediction in a congested urban arterial using probe data. The analysis of prediction results showed that the models can provide more reasonable predictions for sections with higher speed limits and/or longer travel distances. These results indicate that it is harder to accurately predict travel time on roadways with lower speeds, as these roadways have higher crossing-traffic that may lead to more uncertainty that is hard to capture with the ARIMA model. Similarly, crossing-traffic would have a larger effect on shorter sections that could lead to the same uncertainty problem.

Non-Parametric Approaches

Unlike the parametric approaches where the model structure is predetermined, non-parametric travel time prediction approaches have flexibility in the number and nature of their parameters, or, in other words, the model structure and parameter values are all determined based on the traffic patterns in the data [5]. The non-parametric nature of these approaches require more data compared to other approaches. While these approaches have the advantage of being dynamic, they are not very efficient in accounting for unseen incidents (since the models are formed using the data) [5]. Among the non-parametric approaches is the artificial neural networks. Kisgyorgy and Rilett [19] developed a FF-MLP model for travel time prediction using loop detector and GPS data. The developed model was tested on a 26-mile freeway segment to predict travel times 25

Osman, O. A.; H. Rakha, and A. Mittal

minutes into the future. The results showed that the FF-MLP model achieved a robust performance at real-time travel time prediction with an error value of 7.6%. Similar performance was achieved when applying FF-MLP in another study by Mark et al. [20], where simulation data was used to predict travel time. The FF-MLP model in that study was trained using speed, traffic flow, and incident data as inputs collected from a simulation model of a freeway. The developed model in Mark et al.'s study was robust and achieved the best performance using the speed and incident data as inputs with a 4.2% prediction error when predicting travel times up to 20 minutes into the future.

More advanced NN models were also applied for transit travel time prediction. In [21], a Long Short Term Memory Networks (LSTMN) model was developed to predict arrival times for five bus lines with the objective of minimizing wait times. In another study by Pang et al. [22], a Recurrent Neural Networks (RNN) based model with LSTM block was developed to predict bus arrival time considering a large-scale bus trajectory data. In both studies, an average error on the order of 10% was achieved. There have been several attempts to improve the NN models' promising performance in making travel time predictions. For instance, Innamaa [23] applied the Fletcher-Reeves training procedure, which is based on the Conjugate Gradient Algorithm for weight adjustment instead of the widely used back propagation procedure, and the results were comparable to the back-propagation-based NN models. Evolutionary Learning is another training procedure that has been applied instead of back propagation [24]. Unlike the Fletcher-Reeves based NN, the Evolutionary NNs proved to be computationally faster while training and more accurate in their predictions. In another attempt to improve the performance of NNs, researchers applied methods to pre-process and prepare the data before feeding them into the models. The Spectral basis Neural Networks (SNN) [25], Fuzzy C-means Clustering Neural Networks (FCNN) [26], and Principal Component Analysis Neural Networks (PCANN) [27] models are other example improvements that rely on pre-transformation of the input data to the NN models.

The neural networks approaches have been widely applied to solve the traffic forecasting problem. Yet, uncertainty in traffic conditions is one issue that usually stands in the way of achieving a prediction performance that makes those models useful for a vast range of applications. Such uncertainty is usually resulting from multiple exogenous factors including traffic control, weather conditions, and many others. Recently, the introduction of several models (e.g. RNN and LSTM) that can account for the temporal and spatial correlation between different events in the data could help resolve that issue. This study will perform a comparative analysis between LSTMN and MLP-Neural Networks (NN) to explore effect of uncertainty-causing factors for bus travel time prediction.

METHODS

The literature is rich of studies focusing on predicting bus travel times. For instance, Ma et al. [28] developed a support vector machine (SVM) based model to predict travel time. In that study, achieving reasonable prediction required a multi-stage process that starts with identification of similar patterns and similar bus route segments before making any prediction. Similarly, Cristobal et al. [29] relied on travel time profile similarity to achieve reasonable short term bus travel time prediction. The study introduced a model that first performs clustering to identify similar patterns before reasonable short term travel time predictions can be made using neural networks and SVM. Kumar et al. [30] also needed the pre-prediction travel time pattern identification stage before performing short term bus travel time predictions using the K-Nearest Neighbors data mining technique. These studies, and the majority of the literature, rely on a pre-prediction stage that

Osman, O. A.; H. Rakha, and A. Mittal

requires identification of similar patterns and profiles to achieve, mostly short term, reasonable predictions. While these approaches could be successful for AITS applications, the multi-stage modeling is usually time consuming. In efforts to overcome that issue, Petersen et al. [31] developed an expert system that predicted bus travel time 0-1.5 hrs into the future using a convolutional LSTM model. The study predicted travel times with RMSE values as low as 4 seconds. The main advantage of the Conv-LSTM is that it is a one stage process, where identification of similar patterns is built in. In fact, LSTM and convolutional neural networks have a built-in capability to learn temporal and spatial patterns in the data, which makes them advantageous to many other techniques in the literature. Thus, this study aims to develop data-driven prediction models for inter-stations (between each two stations\stops) bus travel time based on the LSTMN technique. The study aims to achieve short and long term travel time predictions which can help travelers plan their trips not just hours in advance, but also days and weeks in advance.

Data Description and Preprocessing

To achieve the study objectives, one-year (from September 2017 through August 2018) transit data were acquired from Blacksburg Transit in Blacksburg, Virginia. The data contains several features including the bus arrival and departure times, and date and time, and were obtained for one route, Heathwood B, which is shown in Figure 1. This route has two modes of service: the full service in which a bus takes off from the main checkpoint/station every 10 minutes and passes through 22 other stations, and the intermediate service in which buses take off every 30 minutes and pass through only 15 stations. It is worth pointing out that bus drivers are allowed to decide not to stop at a station when no passengers are waiting.

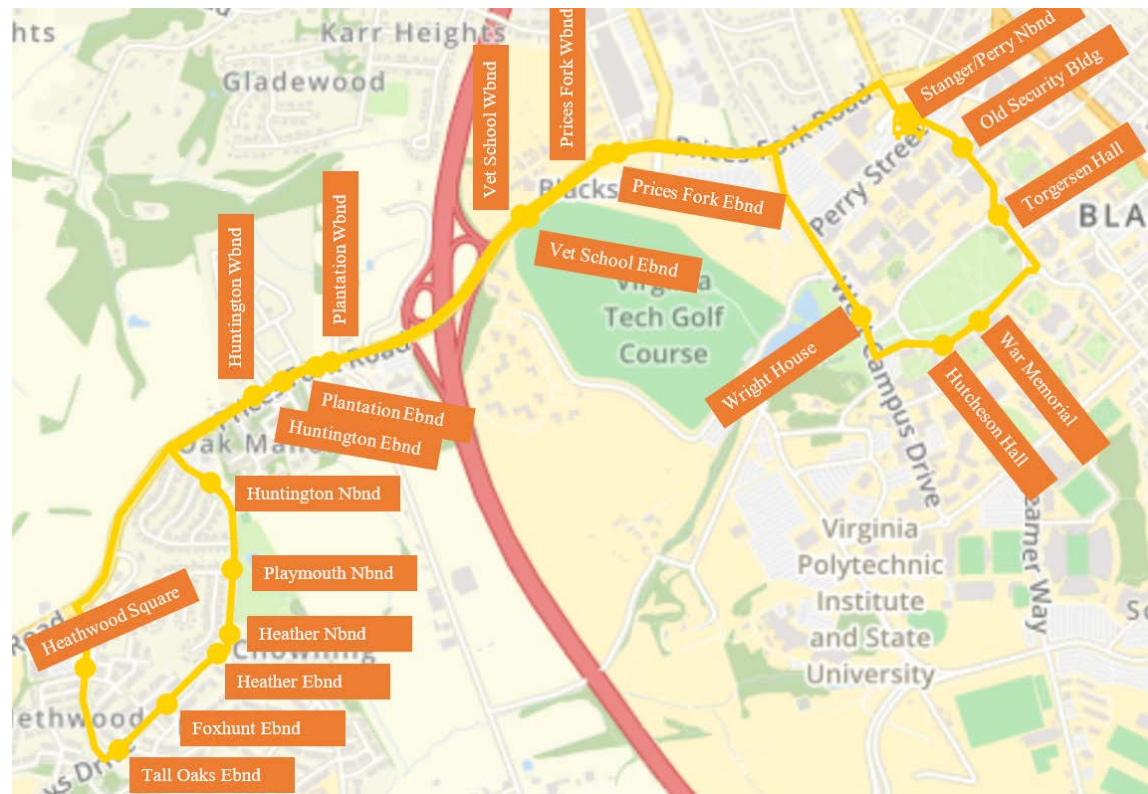


Figure 1 Heathwood B Bus Route

Osman, O. A.; H. Rakha, and A. Mittal

In addition to the bus route information, data about the weather conditions throughout the year, the numbers of controlled intersections (by a stop sign, traffic signal, or pedestrian crossings) between each two stations, speed limits, and the General Transit Feed Specification (gtfs) files were obtained. Using the available information in the data, the travel times between each two stations and the dwell times at each station were calculated. The data went through several preprocessing steps for cleaning from unrealistic arrival or departure times, repeated rows, and unrealistic travel and dwell times. Then, the sequences of stations along the route for each mode of service were identified, followed by assignment of distances between each two stations using the information from gtfs files. The data were then imputed to overcome the problem of missing arrival or departure time observations. The imputation was performed based on the commonly used historical averages method. The data were then sorted based on the time stamp, weekday, and time of day so that a time series is obtained for the inter-station distances. The data was then split into three datasets: one for the models training from September 2017 through April 2018, one for the models validation from May 2018 through July 2018, and the last set is for the month of August 2018 for the models testing.

In this study, the travel time is predicted between each two bus stations to minimize the level of uncertainty when multiple stops are made between two spaced apart stations. Considering two stations i and j (where the bus travels from i to j), the input features for the travel time prediction included the departure time from station i , the dwell time at station i , the travel distance between the two stations, the free flow travel time between the two stations, the day of the week, the hour of the day, the total number of controls (stop signs, traffic signals, and pedestrian crossings) between the two stations, and whether rain/snow took place. To develop the prediction models, the data was normalized such that all features have comparable ranges of values. Once the datasets and input features were ready, the two models were developed to predict travel time between each two stops. In this study, the travel time prediction problem is treated differently compared to the majority of the literature in that no clear prediction horizon is set. In other words, the main objective of the study is to be able to predict/estimate the bus travel time when certain conditions are present (departure time from the previous station, dwell time at the previous station, weather condition, number of controlled points along the distance between the two stations of interest, ... etc) given how this travel time was over the past.

Model Development

To develop the prediction models, it was important to understand the data in hand. Figure 2 presents sample travel time data over a few months of the year (Figure 2-a) as well as for one day (Figure 2-b). As the figures show, the data are noisy and non-stationary. The inter-station travel times vary significantly over the day as well as throughout the year. This is because of the uncertainty caused by the traffic conditions in the network, in addition to the control points (cross-traffic) along the bus route. Yet, the data show signs of both long term seasonality (daily) and short term seasonality (hourly). Such seasonality are signs of temporal correlation between the data points, hence accounting for it can help overcome the effect of the uncertainty factor.

Osman, O. A.; H. Rakha, and A. Mittal

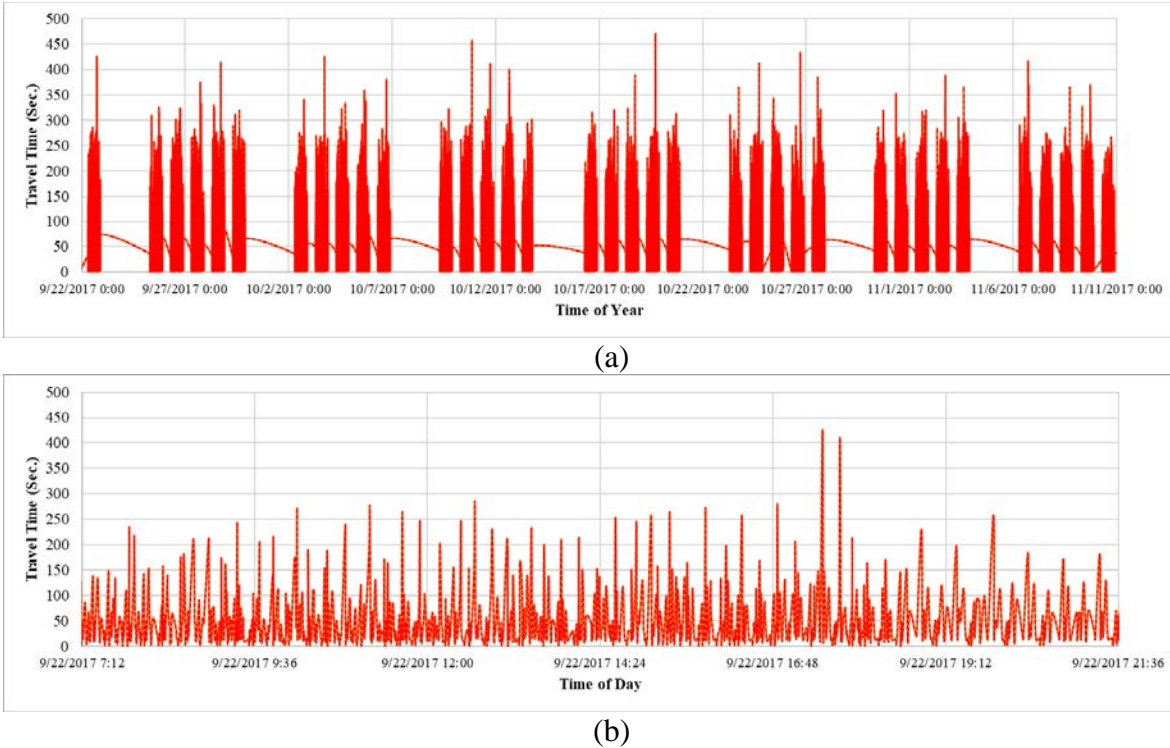


Figure 2 Travel Time Data; (a) over a few months of the year, (b) for one day of the year LSTMN Model

Two models were developed and compared. Both models fall under the non-parametric deep artificial neural networks category. Artificial neural network (ANN) is a modeling approach inspired by how the human brain works. It is an adaptive technique that has been used in several detection and pattern recognition studies (20, 21, 22, 23, 24, 25). ANN has been recognized for its ability to detect patterns in datasets and find the best non-linear function to fit these data. In the following subsections, each model is described in details.

Multi-Layer Perceptron (MLP) Model

The multi-layer perceptron (MLP) is the main and basic type of ANNs in which all nodes are fully connected. In this study a supervised feed-forward MLP network with backward propagation (FFBP-MLP) was applied as one solution to predict inter-station bus travel times. The number of input features (independent variables) was six since the rain/snow and free flow travel time variables did not affect the model performance, hence they were removed from the data. For the transfer function, several functions were tested including Tanh, Sigmoid, Softplus, softmax, SeLu, Elu, and ReLU. The best performance (in terms of prediction accuracy) was achieved with the Tanh transfer function for all layers, except for the output layer where the sigmoid function was used. The model training was performed using the stochastic gradient descent (SGD) optimization algorithm, which enables a faster and more efficient process to locate the global optima of the weights in each layer.

To speed up the learning process, several callbacks were applied: (a) the learning rate was reduced by 10% when the validation loss (loss was calculated as the mean squared error since the problem of interest is a regression problem) [32] did not decrease for more than 50 epochs (an epoch is a training iteration) [33]; and (b) training was stopped when no further reduction in the validation loss was achieved for 350 epochs.

Osman, O. A.; H. Rakha, and A. Mittal

Long Short Term Memory Model

The long short term memory networks (LSTM) is a branch of the recurrent neural networks (RNN) which are designed to learn patterns in sequences of data as they add the temporal dimension to the model architecture. LSTM are more advantageous to RNN as they have feedback connections (forget and memory gates) that enable identification of long sequences of data points. This in turn helps LSTM capture the temporal correlation between the data points, hence makes it well-suited to solve pattern recognition problems based on time-series data.

To determine the optimal values of the hyperparameter of the LSTM model, several trial and error attempts were made. For the temporal dimension, several trials were made to capture the aforementioned long-term and short-term seasonality in the data by changing the look-back time window. Similar to the MLP model, the rain/snow and free flow travel time variables were removed from the data as it did not make any impact in the model performance. Additionally, the Tanh transfer function for all LSTM layers and the sigmoid function for the dense layer achieved the best performance. Finally, the model training was performed using the stochastic gradient descent (SGD) optimization algorithm, and the same callbacks applied in the MLP model were used.

RESULTS

As pointed out earlier, both deep learning models achieved a better performance when the weather condition variable was removed. The reason for this is the limited number of adverse weather observations throughout the year, making its effect, if any, hard to capture by the models. The other variables, on the other hand, especially the total control added significantly to the models performance. This can be explained by the fact that the selected bus route travels through several controlled intersections that affect the bus travel time. Additionally, since the bus route travels through the Virginia Tech campus, the pedestrian crossings are usually of high demand which also adds to the uncertainty in the bus travel time. Hence, accounting for such a variable can help overcome that uncertainty, and improve the travel time prediction accuracy.

To evaluate the performance of the models, two measures were used: the root mean square error (RMSE), the mean absolute error (MAE), and the percentage MAE (PMAE). The three measures are calculated as in equations 1-3, where \hat{y}_i is the predicted travel time at time step i , y_i is the ground truth travel time, and n is the sample size.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (2)$$

$$PMAE(\%) = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (3)$$

In the process of developing the two models, it was noted that achieving low prediction errors was not an easy task. Since the total control was the variable that improved the accuracy significantly, the data was split into two sets: one for the controlled segments, and another for the uncontrolled segments. Accordingly, two additional models were developed for each category (LSTM and MLP): one for only the controlled segment and another for the uncontrolled segments. After several trials to achieve the best prediction performance, the models' hyperparameters were determined as depicted in Table 1. The LSTM models had 2 to 4 hidden layers with one dense output layer for the predicted travel time. The input layer size was 5x6 since the look-back time window was 4 and the number of input features was 6 (representing the departure time from station

Osman, O. A.; H. Rakha, and A. Mittal

i , the dwell time at station i , the travel distance from station i to j , the day of the week, the hour of day, and the total number of controlled points along the segment). The look-back time window of size 4 indicates that the short-term seasonality has a higher impact on future travel time values. For the MLP models, the number of hidden layers and the sizes of each layer increased to achieve a comparable performance to the LSTM models. One possible explanation is that the MLP models are not well suited to capture temporal correlations as in the LSTM models, hence larger size models may be required to be able to capture the high non-linearity in the data and achieve a reasonable performance.

Table 1 Developed Models' Hyperparameters

Model	Parameter	Model for All Segments	Model for Controlled Segments	Model for Uncontrolled Segments
LSTM	Input Shape	5×6	5×6	5×6
	Hidden Layers	3	4	2
	Hidden Layers Size	8×4×2	16×8×8×4	8×4
	OutPut Layer	1 Dense	1 Dense	1 Dense
MLP	Input Shape	6	6	6
	Hidden Layers	4	4	3
	Hidden Layers Size	20×30×10×5	20×30×10×5	20×15×10
	OutPut Layer	1	1	1

Given these models architectures, the performances of the various models are presented in Table 2. As shown in the table, the all-segments LSTM model outperforms the MLP model with 6.3% reduction in the RMSE, 9.7% reduction in the MAE, and 11.4% reduction in the PMAE. These results were expected because of the LSTM models' ability to capture temporal correlations in the data. When splitting the data into controlled and uncontrolled, further improvements were achieved in the prediction performance. Compared to the models for all segments, slight improvements in the prediction performance were achieved for the controlled segments. For instance, the MAE for the LSTM model went from 11.55 sec in Model 1 (all segments) to 10.79 sec in Model 2 (controlled segments), and from 12.67 sec in MLP-Model 1 to 11.67 sec in MLP-Model 2. For the uncontrolled segments, on the other hand, more significant improvements were achieved as the MAE went down to 4.28 sec when applying the LSTM neural networks, and to 4.76 sec when applying the MLP neural networks.

The level of improvement achieved in each type of segments can be explained by the level of uncertainty in the segments. In other words, when the segments are controlled by either stop signs, traffic signals, or pedestrian crossing, the buses may experience highly stochastic levels of delays that would increase the level of uncertainty in those segments. Such uncertainty may not be easy to capture, with different degrees, by the developed models. This uncertainty is significantly minimized when the segments are uncontrolled, hence the developed models were able to accurately predict travel times compared to the controlled segments. In such a case, a complex model such as the LSTM or the MLP may not be even required. In this case, a model that relies on the physical relationships (distance/speed limit) could provide a better performance, which is clear in Table 2. As shown in the table, the physical relationships were able to predict the travel times on the uncontrolled segments with a considerably high accuracy (RMSE = 3.95 sec, MAE = 1.97 sec, and PMAE = 12.56%).

Osman, O. A.; H. Rakha, and A. Mittal

Overall, the various performance measures show the LSTM models superior to the MLP models. The LSTM neural networks models' ability to capture the temporal correlations in the data enabled them to minimize the effect of uncertainty in the data, thus achieve better prediction performance compared to the MLP models. This explanation is supported by the prediction performance graphs in Figure 3. The figure shows that the MLP models are not able to capture the peaks in the travel time curves. All the MLP models seem to have set upper and lower boundaries for the predicted travel times. However, the LSTM models do not have the problem of the upper and lower boundaries. Additionally, when splitting the data, the LSTM models are more able to capture the high and low peaks in the travel time values compared to the MLP models.

1 **Table 2 Models' Performance Measures**

Model		LSTM			MLP			% Improvement			
Parameter		RMSE (sec)	MAE (sec)	MAPE (%)	RMSE (sec)	MAE (sec)	MAPE (%)	RMSE (sec)	MAE (sec)	MAPE (%)	
Model for All Segments (Model 1)		17.69	11.55	31.13	18.81	12.67	34.67	6.33	9.7	11.37	
Model for Controlled Segments (Model 2)	MOP	17.33	10.79	19.85	19.39	11.67	20.28	11.89	8.16	2.17	
	% Improvement	2.04	6.58	36.24	3.08	7.89	41.51	--	--	--	
Model for Uncontrolled Segments (Model 3)	DL	MOP	4.28	3.09	24.07	4.76	3.44	41.6	11.21	11.33	72.83
		% Improvement	75.81	73.25	22.68	74.69	72.85	19.99	--	--	--
	Physical Relationship	MOP	3.95	1.97	12.56	--	--	--	--	--	--
		% Improvement	77.67	82.94	59.65	--	--	--	--	--	--

2 ****DL: Deep Learning (LSTM or MLP)**3 **MOP: Measures of Performance**4 **% Improvements in the rows are calculated relative the All Segments Models**5 **% Improvements in the columns are calculated relative to the LSTM Models**

6

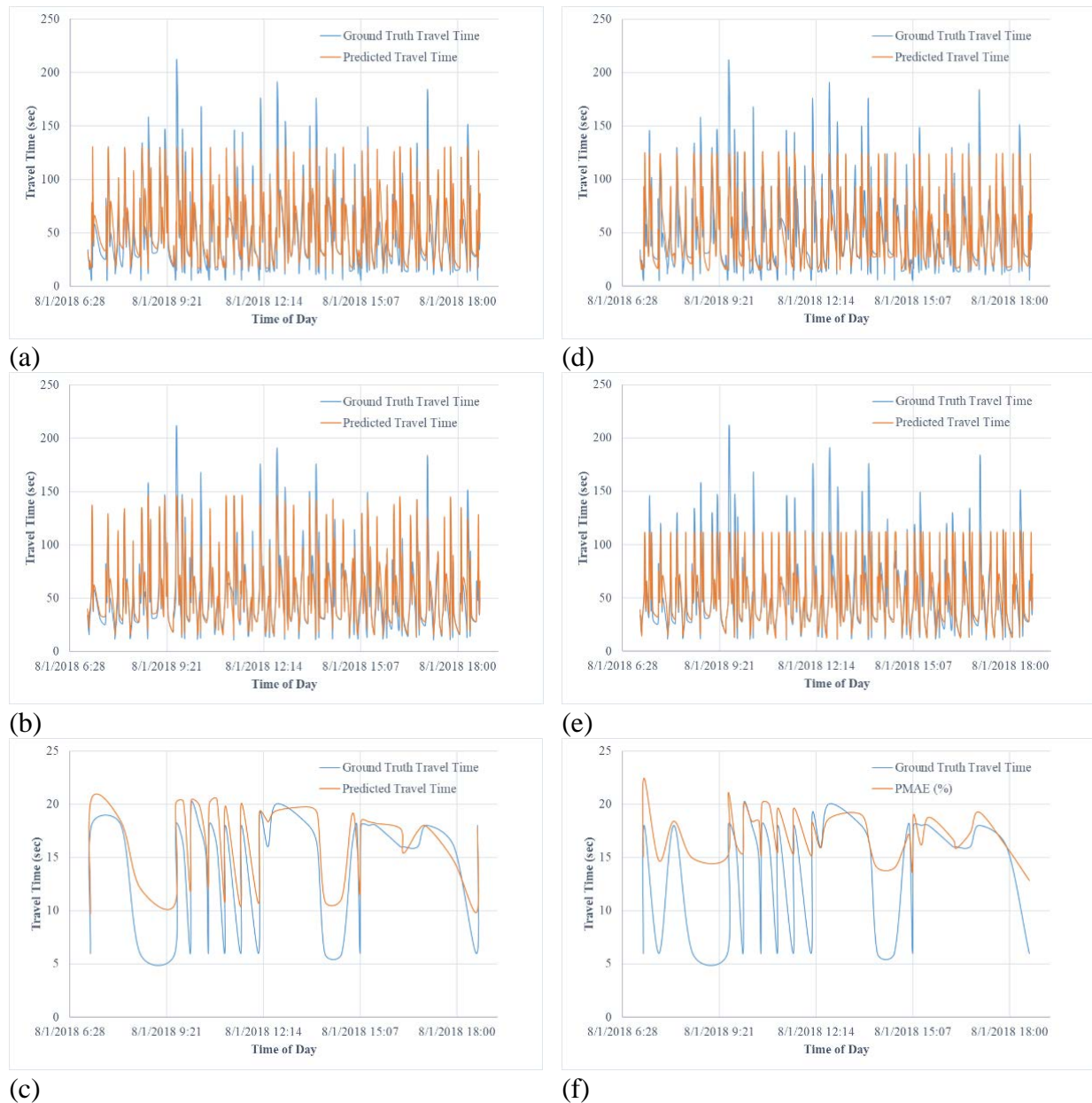


Figure 3 Prediction Performance by Time of Day for a Sample Day; (a) LSTM-Model 1, (b) LSTM-Model 2, (c) LSTM-Model 3, (d) MLP-Model 1, (e) MLP-Model 2, (f) MLP-Model 3

Although the prediction performance for the LSTM models is superior to the MLP models, there is still room for improvement, especially that the RMSE values for Models 1 and 2 are a little not comparable to that of Model 3. The reason for this performance can be explained by looking at Figure 4 which shows the models overestimating travel times at lower travel time values with PMAE values that are sometimes higher than 100%. When the travel times become higher, the models tend to give underestimated predictions with PMAE values that are mostly less than 50%. The overall outcome of the considerably overestimated predictions and the underestimations is the performance measures' values discussed earlier. This figure also shows two important observations: (a) the problem of overestimation and underestimation improved when the data was

Osman, O. A.; H. Rakha, and A. Mittal

split, especially for the uncontrolled segments, which supports the thought that the uncertainty in traffic conditions (to which the control adds) is among the main reason for that under- and overestimation problem; (b) the problem of under- and overestimations is less for the LSTM models compared to the MLP models, which supports the previously pointed out conclusion that accounting for the temporal correlation between the travel times can help minimize the problem of uncertainty in traffic conditions.

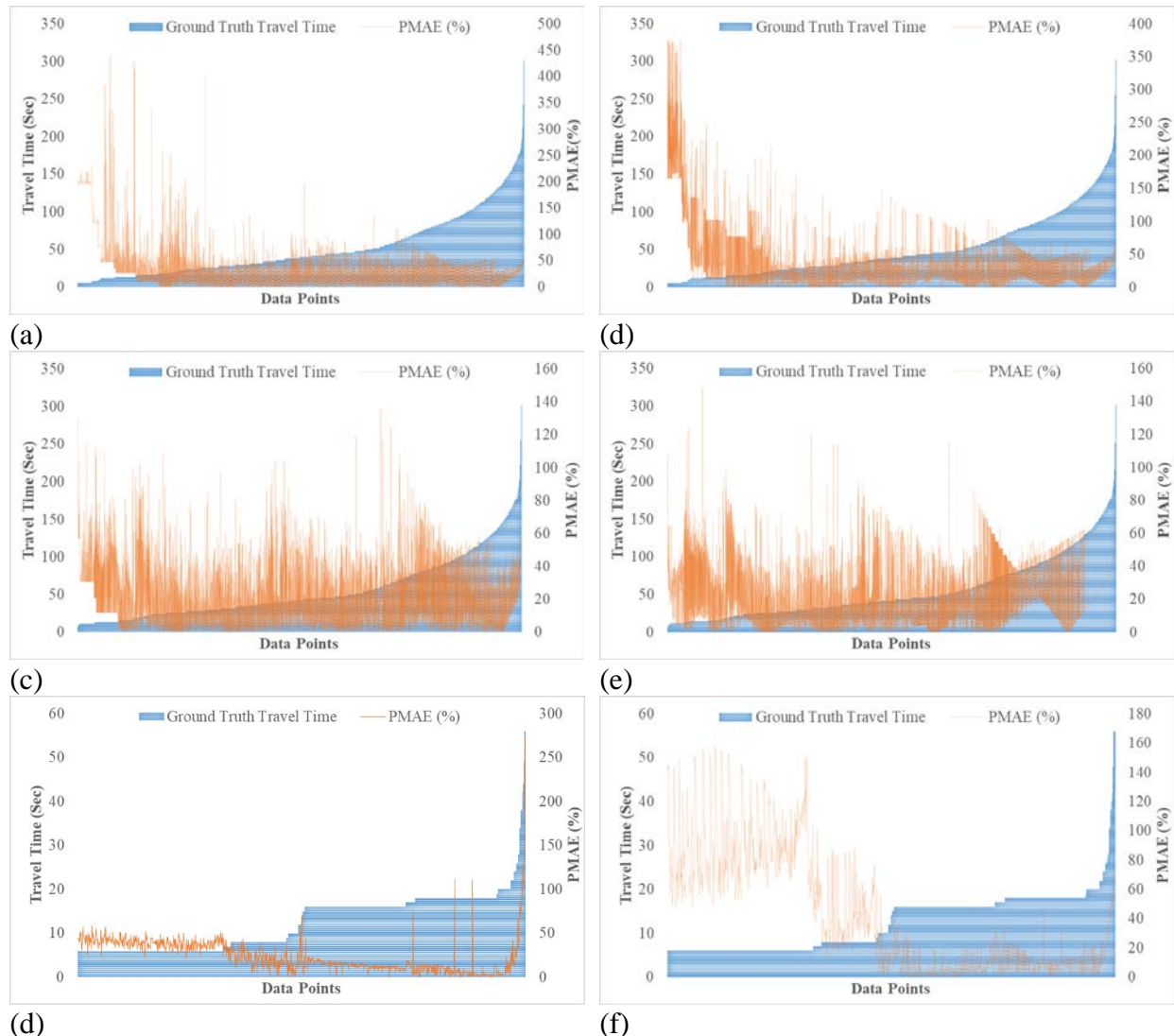


Figure 4 Prediction Performance by Travel Time Values; (a) LSTM-Model 1, (b) LSTM-Model 2, (c) LSTM-Model 3, (d) MLP-Model 1, (e) MLP-Model 2, (f) MLP-Model 3

CONCLUSIONS

Travel time prediction is a crucial task for advanced traveler information systems (ATIS). Transit agencies have been relying on ATIS to help riders properly plan their trips through reduction of wait times at bus stations and improve levels of service of the bus routes. Past and current research has made effort to improve the quality of bus travel time predictions. Yet, many

Osman, O. A.; H. Rakha, and A. Mittal

studies either focus on short term predictions, and/or multi-stage modeling that requires an additional step for identification of similar travel time patterns. Therefore, this study tries to overcome those limitations by developing a Long Short Term Memory Networks (LSTMN) model for short and long term bus travel time prediction. The study benefits from the built-in capability of LSTMN to identify similar patterns in travel time data. The study investigates the value of that capability (accounting for the temporal correlation between the travel time observations) in terms of overcoming uncertainty in the traffic conditions, hence improving the prediction performance. A comparative analysis is performed between the LSTMN model and the traditional multilayer perceptron (MLP) networks. The study treats the prediction problem as a prediction-horizon independent prediction to enable short and long term predictions. In other words, the travel time between bus stations i and j is predicted as a value of several factors (including, the departure time from station i , the dwell time at station i , the travel distance between the two stations, the day of week, the time of day, and total number of controls along the segment) without reliance on a specific prediction horizon.

The preliminary analysis of the two models results indicated that although the LSTM model outperforms the MLP model, the uncertainty factor still plays a role in holding the performance back. To investigate the possibility of achieving further improvement, the data was split into segments with control and segments without control. The models developed for the splitted data showed considerable improvement in the prediction performance, especially for the uncontrolled segments indicating that traffic signals, stop signs, and high-demand pedestrian crossings add to the uncertainty in bus travel times, hence significantly affect the prediction performance. In such a case, a deep learning model may not even be required and reliance on physical relationships can be more than enough. Nonetheless, the temporal component in the LSTM neural networks has the ability to suppress the effect of that uncertainty when it exists and achieve a reasonable prediction performance.

In summary, traffic prediction problem has long been under research considering several approaches and all ended up with the same conclusion that such a problem should be treated carefully because of the stochastic nature of traffic conditions. Fortunately, accounting for correlations between the traffic conditions seems to help in that favor as concluded in this study. Yet, there is still room for improvement. For instance, how travelers respond to the predictions and how travelers' response affects the predicted traffic conditions are two questions their answers can help improve the prediction performance; this will be addressed in a future research by the authors.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: O. A. Osman and H. Rakha; data collection: O. A. Osman and H. Rakha; analysis and interpretation of results: O. A. Osman; draft manuscript preparation: O. A. Osman, H. Rakha, and A. Mittal. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Oh, S., et al., *Short-term travel-time prediction on highway: A review on model-based approach*. KSCE Journal of Civil Engineering, 2018. **22**(1): p. 298-310.
2. Oh, S., et al., *Short-term travel-time prediction on highway: a review of the data-driven approach*. Transport Reviews, 2015. **35**(1): p. 4-32.

Osman, O. A.; H. Rakha, and A. Mittal

3. Chen, H. and S. Grant-Muller, *Use of sequential learning for short-term traffic flow forecasting*. Transportation Research Part C: Emerging Technologies, 2001. **9**(5): p. 319-336.
4. Van Lint, J., *Online learning solutions for freeway travel time prediction*. IEEE Transactions on Intelligent Transportation Systems, 2008. **9**(1): p. 38-47.
5. van Hinsbergen, J. and F. Sanders, *Short Term Traffic Prediction Models*. 2007.
6. Van Lint, J. and C. Van Hinsbergen, *Short-term traffic and travel time prediction models*. Artificial Intelligence Applications to Critical Transportation Issues, 2012. **22**(1): p. 22-41.
7. Ishak, S. and H. Al-Deek, *Performance evaluation of short-term time-series traffic prediction model*. Journal of Transportation Engineering, 2002. **128**(6): p. 490-498.
8. Iwasaki, M. and K. Shirao. *A short term prediction of traffic fluctuations using pseudo-traffic patterns*. in *Intelligent Transportation: Realizing the Future. Abstracts of the Third World Congress on Intelligent Transport SystemsITS America*. 1996.
9. Jiang, Y., *Dynamic prediction of traffic flow and congestion at freeway construction zones*. Journal of Construction Engineering, 2002. **7**(1): p. 45-57.
10. Vojak, R., J.L. Vehel, and M. Danech-Pajouh. *A first step towards road traffic short-term prediction using multifractal tools*. in *DRIVE-II Workshop on Short Term Traffic Forecasting (2nd: 1994: Delft, Netherlands). Proceedings of the second DRIVE-II Workshop on Short Term Traffic Forecasting*. 1994.
11. Chien, S., X. Liu, and K. Ozbay, *Predicting travel times for the South Jersey real-time motorist information system*. Transportation Research Record: Journal of the Transportation Research Board, 2003(1855): p. 32-40.
12. Park, D. and L.R. Rilett, *Forecasting freeway link travel times with a multilayer feedforward neural network*. Computer-Aided Civil and Infrastructure Engineering, 1999. **14**(5): p. 357-367.
13. Kwon, J., B. Coifman, and P. Bickel, *Day-to-day travel-time trends and travel-time prediction from loop-detector data*. Transportation Research Record: Journal of the Transportation Research Board, 2000(1717): p. 120-129.
14. Rice, J. and E. Van Zwet. *A simple and effective method for predicting travel times on freeways*. in *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. 2001. IEEE.
15. Ahmed, M.S. and A.R. Cook, *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*. 1979.
16. Ahmed, S.A., *Stochastic processes in freeway traffic Part I. Robust prediction models*. Traffic Engineering & Control, 1983. **24**(HS-035 775).
17. Levin, M. and Y.-D. Tsao, *On forecasting freeway occupancies and volumes (abridgment)*. Transportation Research Record, 1980(773).
18. Billings, D. and J.-S. Yang. *Application of the ARIMA models to urban roadway travel time prediction-a case study*. in *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*. 2006. IEEE.
19. Kisgyörgy, L. and L.R. Rilett, *Travel time prediction by advanced neural network*. Periodica Polytechnica Civil Engineering, 2002. **46**(1): p. 15-32.
20. Mark, C.D., A.W. Sadek, and D. Rizzo. *Predicting experienced travel time with neural networks: a PARAMICS simulation study*. in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*. 2004. IEEE.

Osman, O. A.; H. Rakha, and A. Mittal

21. Huang, Z., et al., *A novel bus-dispatching model based on passenger flow and arrival time prediction*. IEEE Access, 2019. **7**: p. 106453-106465.
22. Pang, J., et al., *Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network*. IEEE Transactions on Intelligent Transportation Systems, 2018. **20**(9): p. 3283-3293.
23. Innamaa, S., *Short-term prediction of travel time using neural networks on an interurban highway*. Transportation, 2005. **32**(6): p. 649-669.
24. Annunziato, M. and S. Pizzuti. *A Smart-Adaptive-System based on Evolutionary Computation and Neural Networks for the on-line short-term urban traffic prediction*. in *EUNITE*. 2004.
25. Park, D., L.R. Rilett, and G. Han, *Spectral basis neural networks for real-time travel time forecasting*. Journal of Transportation Engineering, 1999. **125**(6): p. 515-523.
26. Park, D. and L. Rilett, *Forecasting multiple-period freeway link travel times using modular neural networks*. Transportation Research Record: Journal of the Transportation Research Board, 1998(1617): p. 163-170.
27. Ishak, S. and C. Alecsandru, *Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings*. Journal of Transportation Engineering, 2004. **130**(4): p. 452-465.
28. Ma, J., et al., *Bus travel time prediction with real-time traffic information*. Transportation Research Part C: Emerging Technologies, 2019. **105**: p. 536-549.
29. Cristóbal, T., et al., *Bus Travel Time Prediction Model Based on Profile Similarity*. Sensors, 2019. **19**(13): p. 2869.
30. Kumar, B.A., et al., *Real time bus travel time prediction using k-NN classifier*. Transportation Letters: The International Journal of Transportation Research, 2019. **11**(7).
31. Petersen, N.C., F. Rodrigues, and F.C. Pereira, *Multi-output bus travel time prediction with convolutional LSTM neural network*. Expert Systems with Applications, 2019. **120**: p. 426-435.
32. Goodfellow, I., et al., *Deep learning*. Vol. 1. 2016: MIT press Cambridge.
33. Jacobs, R.A., *Increased rates of convergence through learning rate adaptation*. Neural networks, 1988. **1**(4): p. 295-307.