



An Artificial Intelligence Approach to Estimate Travel Time along Public Transportation Bus Lines

Mohammad S. Ghanim

mohammad.ghanim@jacobs.com
CH2M HILL International B.V., Doha, Qatar

Khaled Shaaban

kshaaban@qu.edu.qa
Department of Civil and Architectural Engineering/Qatar Transportation and Traffic Safety Center, Qatar University, Doha, Qatar

Motsem Miqdad

mm080855@qu.edu.qa
Department of Civil and Architectural Engineering, Qatar University, Doha, Qatar

ABSTRACT

Public transportation sectors have played significant roles in accommodating passengers and commodities efficiently and effectively. The modes of public transportation often follow pre-defined operation schedules and routes. Therefore, planning these schedules and routes requires extensive efforts in analyzing the built environment and collecting demand data. Once a transit route is operational as an example, collecting and maintaining real-life information becomes an important task to evaluate service quality using different Key Performance Indicators (KPIs). One of these KPIs is transit travel time along the route. This paper aims to develop a transit travel time prediction model using an artificial intelligence approach. In this study, 12 public bus routes serving the Greater City of Doha were selected. While the ultimate goal is to predict transit travel time from the start to the end of the journeys collected over a period of one-year, route-specific inputs were used as inputs for this prediction. To develop a generalized model, the input variables for the transit route included the number and type of intersections, number of each type of turning movements and the built environment. An Artificial Neural Networks (ANN) model is used to process 78,004 valid datasets. The results indicate that the ANN model is capable of providing reliable and accurate transit travel time estimates, with a coefficient of determination (R^2) of 0.95. Transportation planners and public transportation operators can use the developed model as a tool to estimate the transit travel time.

Keywords: Public transportation; Artificial neural networks; Travel time prediction

1 INTRODUCTION

The population in the State of Qatar has increased dramatically during the last few decades and the lifestyle for the people living in the country became a car-dependent one (Shaaban & Khalil, 2012; Shaaban & Khalil, 2013). These conditions caused a substantial pressure on the transport system. Moreover, the current National Vision 2030 for Qatar calls for a reliable multimodal transport system to support the national economic and social development of the country. Furthermore, Qatar is currently planning to host different major events in the near future including the FIFA World Cup

2022 event. Therefore, it is necessary to have an efficient, accessible and sustainable transport system in the country (Shaaban & Hassan, 2014; Shaaban & Maher, 2019).

Previous studies indicate that efficient public transportation is one of the methods to improve traffic congestion (Beaudoin, Farzin, & Lawell, 2015, Dewan & Ahmad, 2007; Shaaban & Khalil, 2012). In Qatar, the public bus service provides an affordable and clean mode of transport to many people (Shaaban & Khalil, 2012; Shaaban & Khalil, 2013). Mowasalat, the only public transport bus system operator in the country, manages more than 45 bus routes around Qatar. Some of these routes connect the city of Doha with other cities.

Nevertheless, the reliability of the bus system, which is defined as the abidance of the bus arrival time to the announced scheduled timings, is considered an issue. The public bus users have reported the need to improve the reliability of the existing bus system (Shaaban & Khalil, 2012; Shaaban & Khalil, 2013). This problem could be tackled using different methods, including improving route planning and scheduling (Nam & Park, 2018), using more realistic and empirical models (Li et al., 2018), improving quality control (Friman & Felleson, 2009), using Automated Vehicle Location (AVL) systems (Hickman, 2004), and notifying the users of actual or better estimation of arrival time using Advanced Traveler Information Systems (ATIS) (Kumar et al, 2005).

The purpose of this study is to develop a prediction model that can be used by public agencies and transport operators to provide more accurate travel and arrival times for the bus system. The development of such a model using historical data for different routes in Qatar can be useful to improve the travel time reliability issue. The model can be utilized for planning and ATIS purposes. Such a model could resolve the current gap that public transport bus operators in Qatar have to deal with due to the lack of regionally developed models needed to estimate realistic travel time based on empirical efforts.

2 RESEARCH OBJECTIVE

Some of the key performance indicators that are often used in evaluating the quality of public transportation services are schedule adherence (Levine et al., 2000), transit travel time (Kumar at al, 2019), service frequency (Ingvardson et al., 2018) , and out-of-pocket cost (Koppelman, 1983). It can be noticed that accurately predicting transit travel time contributes towards improved service quality. While those travel time values can be estimated at the planning stage, comparing those values against the operational ones can lead to an improved planning process. The improvement process can be further appreciated if prediction models based on real-life observations are readily available. Furthermore, such models can be integrated within the Intelligent Transportation Systems (ITS) framework (Shaaban & Khalil, 2012; Shaaban & Khalil, 2013).

The research objectives can be addressed by implementing the following procedures:

- Collect, process, and analyze historical bus journey trips at selected sites, covering differently built environmental conditions.
- Identify the potential quantified input variables that can be used to estimate the transit travel time for different routes.
- Develop transit travel time prediction model using artificial intelligence techniques and methods based on the previously collected and processed input variables.

3 RESEARCH METHODOLOGY

Public transportation data were collected from two different sources. The first one was Mowasalat, which is the governmental entity in Qatar that manages and operates the public bus system in Qatar. The second source was collecting additional data manually as shown in Table 1. The collected data considered the characteristics of the bus routes in terms of both fixed features (such as route length, number of turning movements, number and type of intersections) and traffic-related features (such as the day of the week and peak periods). It should be noted that the research team could not get other useful data (such as number of and rates of passenger ridership and boarding and alighting passengers at bus stops) due to logistics and security reasons. Accordingly, this study utilized the available data only.

Table 1: Description of the independent variables

<i>Variable</i>	<i>Variable Type</i>	<i>Encoding Description</i>
<i>Input Variables</i>		
<i>Distance</i>	Numerical	Distance as recorded by AVL; rounded to nearest 1 km
<i>Month</i>	Binary	12 Columns (one for each month) were defined at which each column accepts either 0 or 1 as value
<i>Weekday</i>	Binary	7 Columns (one for each weekday) were defined at which each column accepts either 0 or 1 as value
<i>Weekend</i>	Binary	1 = Weekend (Friday or Saturday) 0 = All other days
<i>Peak</i>	Binary	4 Columns (AM, MD, PM, and Off-Peak) were defined at which each column accepts either 0 or 1 as value
<i>Starting Time</i>	Numerical	The timing bus started the trip, rounded to nearest 30 mins for ease of analysis and interpretation
<i>Through</i>	Numerical	Number of Movements per trip/route
<i>Left</i>	Numerical	Number of Movements per trip/route
<i>Right</i>	Numerical	Number of Movements per trip/route
<i>U-Turn</i>	Numerical	Number of Movements per trip/route
<i>CBD</i>	Binary	3 Columns (Inner-CBD, Outer-CBD, and Non-CBD) were defined at which each column accepts either 0 or 1 as value
<i>Road Class</i>	Binary	1= Major roads (arterials, expressways...etc.) 0= Local Roads
<i>Land use</i>	Binary	4 Columns (Residential, Commercial, Industrial, and Mixed-use) were defined at which each column accepts either 0 or 1 as value
<i>Stops</i>	Numerical	Number of Stops per Route/Trip
<i>Signals</i>	Numerical	Number of Traffic Signals per Route/Trip
<i>Roundabouts</i>	Numerical	Number of Roundabouts per Route/Trip
<i>Output Variable</i>		
<i>Travel Time</i>	<i>Numerical</i>	Trip duration in minutes rounded to 1 min

After grouping and filtering the collected data, the average transit travel times are used for matching records. For instance, if multiple records have identical input variables but their outputs are different, then the average values of these outputs are used after achieving a 95 percent confidence level that there are no statistical differences between these outputs. Then the outliers are excluded from the analysis.

After filtering the data records, it was found that 78,004 of valid data records were available for the analysis. Due to the nature of the independent variables (which has both numerical and categorical variables), an artificial intelligence approach, Artificial Neural

Networks (ANN), has been used to process and model valid records. This approach has been used due to its capability of analyzing big data and understanding the underlying relationship between the independent variables (Jun Lee & Siau, 2001; Ghanim & Abu-Lebdeh, 2018; Ghanim & Shaaban, 2018). The ANN approach performs well in predicting various combinations of input variables that were not previously observed in the analyzed data (Yahia & El-taher, 2010). ANN techniques have several applications in transportation-related studies and research, as can be found in (Ghanim, Abu-Lebdeh et al., 2013a; Ghanim, Abu-Lebdeh et al., 2013b; Ghanim, Dion et al., 2014; Ghanim & Abu-Lebdeh, 2015).

Inspired from the natural neuro-system that exists in humans, ANN processes the stimuli (i.e., inputs) and transfers them into reactions (i.e., outputs) (Yahia & El-taher, 2010). A typical ANN model consists of three different layers. The first layer is for the input variables, where the inputs are usually normalized via transfer functions to avoid any potential bias of the input vectors. The second layer is the hidden layer, which consists of several neurons that would receive the weighted signals from the input layers and transfer other sets of weighted signals to the third layer (i.e., the outputs layer). The third layer processes all the signals received from the neurons in the hidden layer and transfers those signals into an output vector.

The valid records were randomly divided into three different groups. The first one is the training group that is used to train and calibrate the weights between the three different layers through an iterative process. The second group is the validation group that is used to monitor the convergence process of the overall training process, and hence avoiding the model “overfitting” during the training process. The third group is the testing group, which is an independent set of data that were not previously experienced through the ANN model development. In other words, the testing group is used to evaluate the performance of the trained ANN model.

- **Training records:** 54,602 records, which are approximately 70% of all the valid records, and used to train the ANN.
- **Validation records:** 11,701 records, which are approximately 15% of all the valid records and used to validate and monitor the performance of the ANN training process.
- **Testing records:** 11,701 records, which are approximately 15% of all the valid records and used to evaluate the prediction performance of the trained ANN model.

Several training and learning algorithms were evaluated, and the selected one was Levenberg-Marquardt algorithm due to its fast convergence. The hidden layer that was used in the final ANN model consists of 10 neurons. Signals from the input layer are transferred to the output layer using the hyperbolic-tangent transfer function, while a linear transfer function is used to transfer the signal from the hidden layer to the output layer.

Figure 1 and Table 2 summarize the performance of the developed ANN model in predicting the route transit travel time for the different conditions.

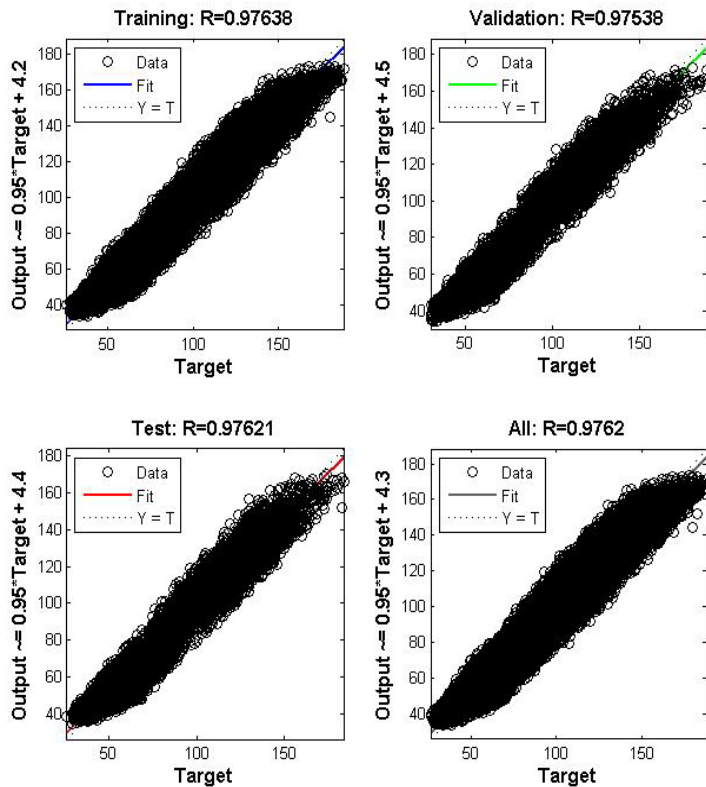


Figure 1: Performance of ANN for training, testing and validation datasets

Table 2: Performance measures for the developed ANN model training process

Number of Epochs	R	R ²	Mean-Squared Error (min ²)	Mean of Absolute Error (min)	Standard Error (min)
123	0.9762	0.9530	55.21	5.94	0.13

4 VALIDITY OF TRANSIT TRAVEL TIME DEVELOPED MODEL

Following the development of the ANN prediction model, its prediction accuracy has been evaluated by comparing the observed travel time to the predicted ones of the testing group (recall that the testing group has 11,701 records that are used only to assess the performance of the developed model, and they were not part of the learning process at any stage).

To do so, two histograms were created. The first one shows the number of records where the error in predicting the journey's transit travel time is within a given time range. The second histogram shows the number of records where the error is within a given percentage of the trip's observed travel time, as can be seen in Figure 2 and Figure 3.

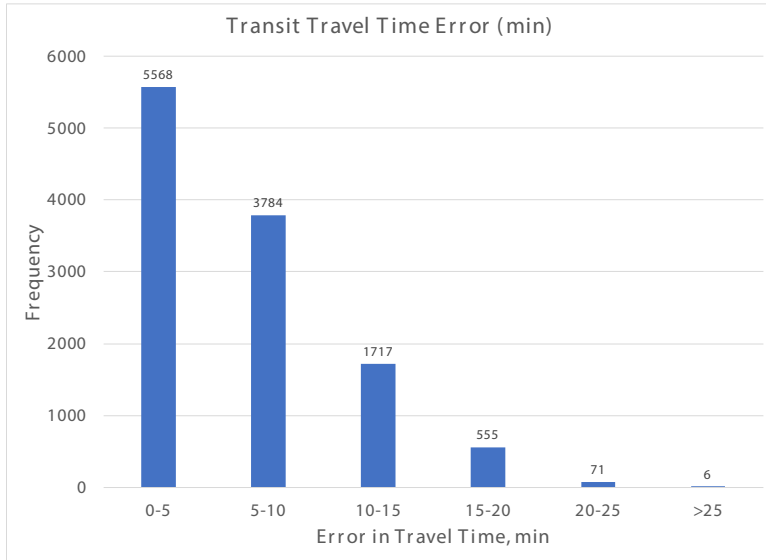


Figure 2: Performance of ANN for training, testing and validation datasets

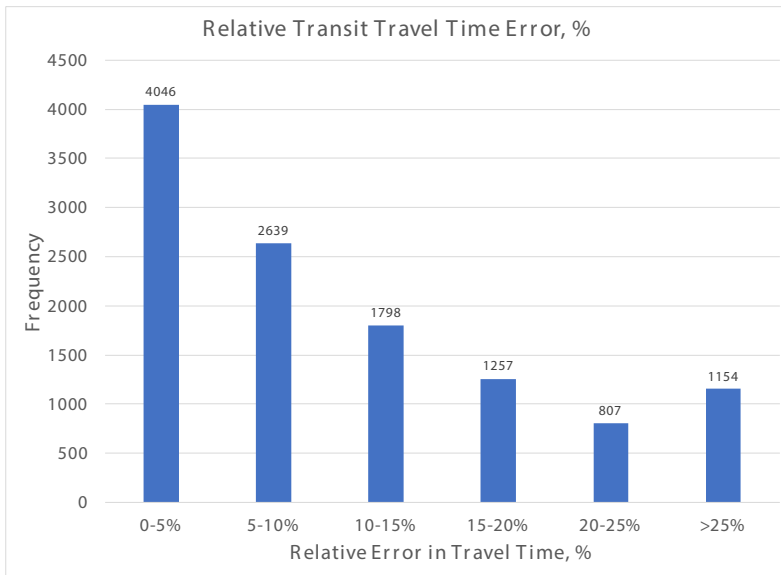


Figure 3: Performance of ANN for training, testing and validation datasets

Figure 2 and Figure 3 clearly reveal that the developed model is performing well in providing a reliable estimate of travel time. For instance, for approximately half of the transit trips, the prediction error is less than five minutes, and around 95% of the trips have an error of 15 minutes or less.

In terms of the percentage error of travel time prediction, more than two-third of the records have an error of 15% or less, with only 10% of all the trips having experienced an error that is more than 25% of its travel time. This observation is an indication that another input variable describing the trip length (i.e., short or long trip for instance) can

be introduced to improve the prediction capabilities of the model.

5 CONCLUSION AND RECOMMENDATIONS

The objective of this research is to evaluate the potential implementation of ANN in predicting model the travel time of different bus routes using a set of variables describing the route characteristics. The model can be later used as a tool by public transportation operators and planners. To do so, a one-year worth of data from 12 different routes were collected, processed and modeled.

Approximately 150,000 records were analyzed. The sum of 78,004 records were found valid and used in the analysis. An ANN model is developed to predict the travel time along the transit route. The results show that the ANN was able to provide a reliable estimation of travel times.

One of the key limitations of this study is the absence of traffic volumes or other traffic congestion indications along the transit travel routes due to the unavailability of such data. Nonetheless, the developed model furnishes the overall framework that can be enriched when traffic information or congestion level indicators are nested in the future.

While the study shows promising results regarding the use of the developed model by stakeholders to plan different public transit routes, there is a significant potential of expanding the scope of this study, where more information about ridership can be integrated into the model. It can also be expanded to be integrated within a real-time advanced traveler information system or an advanced public transportation system.

ACKNOWLEDGMENT

The authors acknowledge the valuable cooperation of Mowasalat in making their data available. All views and results presented in the paper are solely the authors'.

REFERENCES

- Beaudoin, J., Farzin, Y., & Lawell, C. (2015). Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality. *Research in Transportation Economics*, 52, 15-22.
- Dewan, K. & Ahmad, I. (2007). Carpooling: a step to reduce congestion. *Engineering Letters* 14(1): 61-66.
- Friman, M., & Fellesson, M. (2009). Service supply and customer satisfaction in public transportation: The quality paradox. *Journal of Public transportation*, 12(4), 4.
- Hickman, M. (2004). Bus automatic vehicle location (AVL) systems. In *Assessing the Benefits and Costs of ITS* (pp. 59-88). Springer, Boston, MA.
- Ghanim, M., Abu-Lebdeh, G., and Ahmed, K. (2013a). Microscopic Simulation Study of Transit Signal Priority Implementation Along an Arterial Corridor, *5th International Conference on Modeling, Simulation, and Applied Optimization*. Hammamet, Tunisia.
- Ghanim, M., Abu-Lebdeh, G. & Ahmed, K. (2013b). Modeling Historical Traffic Data using Artificial Neural Networks, *5th International Conference on Modeling, Simulation, and Applied Optimization*. Hammamet, Tunisia.
- Ghanim, M., & Abu-Lebdeh, G. (2018). Projected state-wide traffic forecast parameters using artificial neural networks. *IET Intelligent Transport Systems*, 13(4), 661-669.

- Ghanim, M. S. & Abu-Lebdeh, G. (2015). Real-time dynamic transit signal priority optimization for coordinated traffic networks using genetic algorithms and artificial neural networks. *Journal of Intelligent Transportation Systems*, 19(5): 327-338.
- Ghanim, M. S., Dion, F. & Abu-Lebdeh, G. (2014). The impact of dwell time variability on transit signal priority Performance. *Canadian Journal of Civil Engineering*, 41(2): 154-163.
- Ghanim, M. S. & Shaaban, K. (2018). Estimating turning movements at signalized intersections using artificial neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 20(5): 1828-1836.
- Ingvardson, J. B., Nielsen, O. A., Raveau, S., & Nielsen, B. F. (2018). Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis. *Transportation Research Part C: Emerging Technologies*, 90, 292-306.
- Jun Lee, S. & Siau, K. (2001). A review of data mining techniques. *Industrial Management & Data Systems*, 101(1): 41-46.
- Koppelman, F. S. (1983). Predicting transit ridership in response to transit service changes. *Journal of Transportation Engineering*, 109(4), 548-564.
- Kumar, P., Singh, V., & Reddy, D. (2005). Advanced traveler information system for Hyderabad City. *IEEE Transactions on Intelligent Transportation Systems*, 6(1), 26-37.
- Kumar, B. A., Jairam, R., Arkatkar, S. S., & Vanajakshi, L. (2019). Real time bus travel time prediction using k-NN classifier. *Transportation Letters: The International Journal of Transportation Research*, 11(7).
- Levine, J., Hong, Q., Edward Hug Jr, G., & Rodriguez, D. (2000). Impacts of an advanced public transportation system demonstration project. *Transportation research record*, 1735(1), 169-177.
- Li, Z., Hensher, D. A., & Ho, C. (2018). An empirical investigation of values of travel time savings from stated preference data and revealed preference data. *Transportation Letters*, 1-6.
- Nam, K., & Park, M. (2018). Improvement of an optimal bus scheduling model based on transit smart card data in Seoul. *Transport*, 33(4), 981-992.
- Shaaban, K. & Hassan, H. M. (2014). Modeling significant factors affecting commuters' perspectives and propensity to use the new proposed metro service in Doha. *Canadian Journal of Civil Engineering*, 41(12): 1054-1064.
- Shaaban, K. & Khalil, R. F. (2012). Proposed policies to support the new metro system in Qatar. *Procedia-Social and Behavioral Sciences*, 48: 2315-2324.
- Shaaban, K. & Khalil, R. F. (2013). Investigating the customer satisfaction of the bus service in Qatar. *Procedia-Social and Behavioral Sciences*, 104: 865-874.
- Shaaban, K. & Maher, A. (2019). *Using the theory of planned behavior to predict the use of an upcoming public transportation service in Qatar*. *Case Studies on Transport Policy*, <https://doi.org/10.1016/j.cstp.2019.11.001>.
- Yahia, M. & El-taher, M. (2010). A new approach for evaluation of data mining techniques. *International Journal of Computer Sciences*, 7(5): 181-186.

Cite this article as: Ghanim M. S., Shaaban K., Miqdad M., "An Artificial Intelligence Approach to Estimate Travel Time along Public Transportation Bus Lines", *International Conference on Civil Infrastructure and Construction (CIC 2020)*, Doha, Qatar, 2-5 February 2020, DOI: <https://doi.org/10.29117/cic.2020.0074>