QATAR UNIVERSITY

COLLEGE OF ENGINEERING

TRAVEL TIME PREDICTION MODEL FOR PUBLIC TRANSPORT BUSES IN QATAR USING ARTIFICIAL NEURAL NETWORKS

By

MOTASEM K. MEQDAD

A Thesis Submitted to

the Faculty of the College of

Engineering

in Partial Fulfillment

of the Requirements

for the Degree of

Masters of Science in Civil Engineering

January 2018

COMMITTEE PAGE

The members of the Committee approve the Thesis of Motasem K. Meqdad defended on 13/12/2017.

Dr. Mohammad Shareef Mohammad Ghanim
Thesis/Dissertation Supervisor

Dr. Khaled Salah Shaaban
Committee Member

Dr. Mehrnaz Ghamami
Committee Member

Prof. Murat Gunduz
Committee Member

Approved:

Khalifa Al-Khalifa, Dean, College of Engineering

ABSTRACT

The state of Qatar has experienced rapid population growth over the last few years. This growth of population has caused authorities to promote the use of public transportation, by introducing new public transport systems such as transit buses and metro lines. The existing bus system was introduced in 2004 to the local community in Qatar. Despite the importance of this system, there are limited studies that are done to analyze and identify its characteristics. There is not much analysis of the stop-to-stop travel time or schedule reliability. The objective of this research is to develop a prediction model for transit route travel time. The model can predict the travel time of buses using several independent variables that are different for each transit route. The prediction model can be used as a useful tool to the decision makers and public transport officials, which can be used for planning, system reliability and quality control, and real-time advanced travelers' information systems.

The data was collected for 12 routes over a period of one year (2015-2016) within The Greater City of Doha using Automatic Vehicle Location (AVL) system. Transit travel time data was obtained from Mowasalat records, the sole operator of public transport buses in Qatar. The collected data include travel time data, route information, geometric configurations, land use, and traffic data. After systematic checking of errors in the collected data and elimination of irreverent records, more than 78,004 trips were analyzed using Artificial Neural Networks (ANN) data mining technique. Prediction model, with R^2 of 0.95 was developed. The results indicate that the developed model is accurate and reliable in predicting the travel

time. The model can be generalized as well to be applied to newly planned routes, or updating the schedules of existing routes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my thanks to my parents, wife (Sarah) and every member of my family and dear friends. They had stood by me and supported me to complete this research.

Also, I would like to express my deep appreciation to my dear friend Eng. Mazin Elamin for his guidance and support throughout the thesis development stages.

1. INTRODUCTION

Driven by a commitment to organizing exceptional and unique world cup event in 2022, Qatar population had reached 2.67 million in February 2017, with rapid population growth and growth rates as shown in Figure 1.1 (Ministry of Development Planning and Statistics, 2017). The population was a result of the booming construction industry that invited millions of workers to help Qatar deliver its vision by 2030; which includes but is not limited to developing sustainable transportation infrastructure, increasing its gas exports and developing required infrastructure to put Qatar on the map of most important event hubs worldwide.

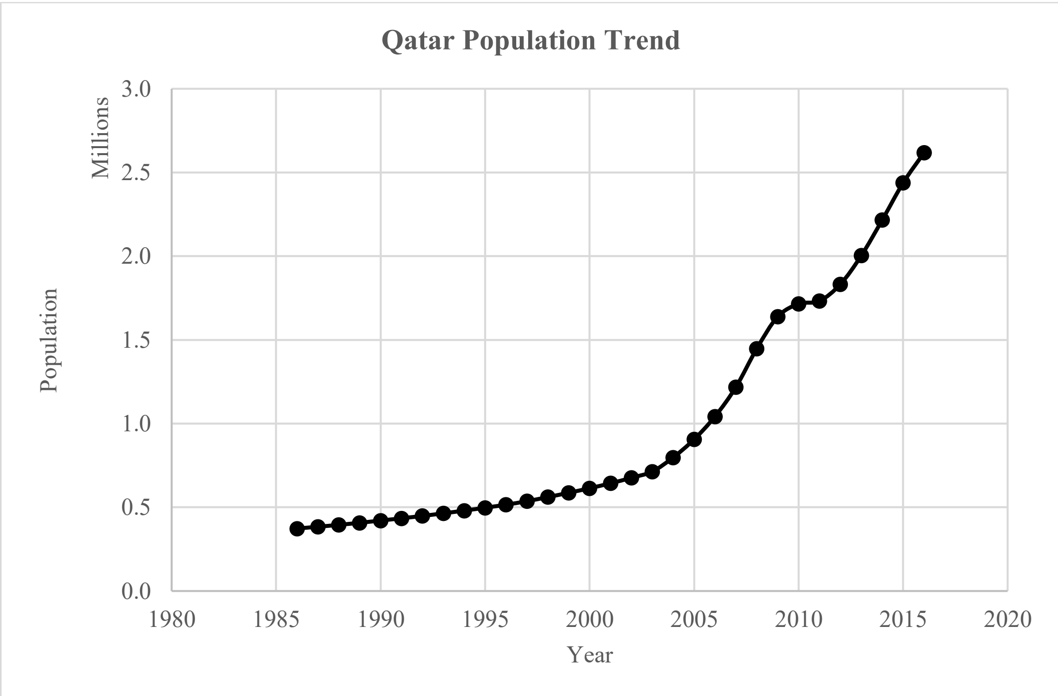


Figure 1.1: Qatar Population between 1986 and February 2017 (Ministry of Development Planning and Statistics, 2017)

This population explosion leads to significant pressure on Qatar transport system since the lifestyle of most inhabitants of Doha became car dependent (Shaaban & Khalil, 2013). Moreover, Qatar is currently proliferating in hosting events; thus, there is a need for safe, accessible and sustainable transport system (Hubschneider, 2011).

With the rapid development of the capital, Doha, and the expansion of the residential areas, the distances to be travelled and the numbers of commuters increased significantly. Public transportation in Doha, Qatar was first introduced in 2004 by Mowasalat, which initiated the national bus system for use by the public. Substantial development to the bus system to place during the planning of the Asian Games in 2006. Subsequently, the public bus system became the official transportation mode for the main events.

Mowasalat is currently operating more than 45 routes in Qatar as shown in Figure 1.2. Nine of those routes are connecting Doha to other cities in Qatar. Few number of researches were published to address this transport system. In 2015, average number of people using public transport buses in Qatar reached 60,000 (Fahmy, 2015) while this number increased by 20% to reach approximately 72,000 in 2016 (Ministry of Transport and Communication, 2017).

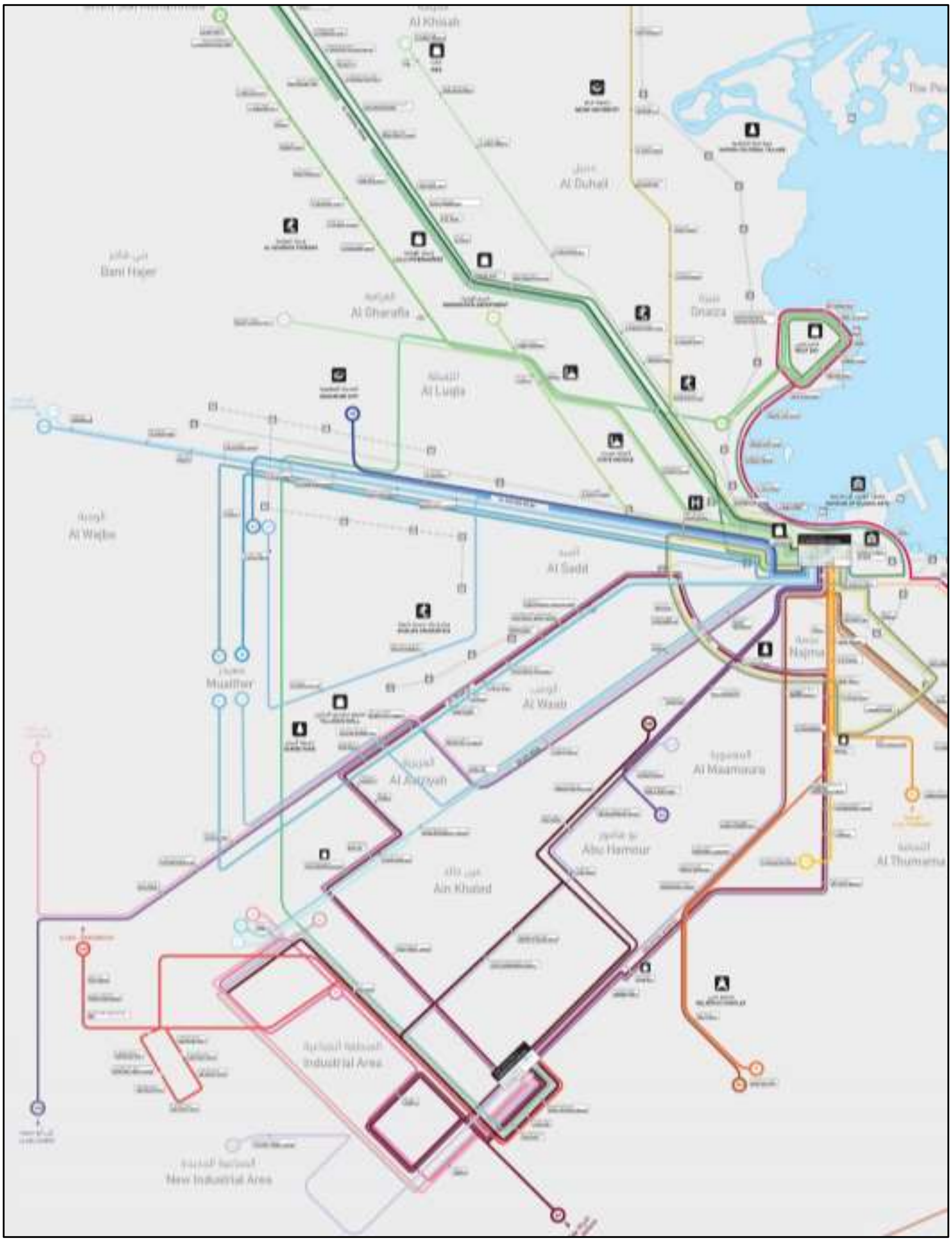


Figure 1.2 Public Transport Buses Map

With all the traffic congestion in Doha, the buses travel time, carbon footprint and fuel consumption has significantly increased while the overall mobility has decreased. This issue can be broken down into two categories:

1. Demand increase and fluctuation which was suggested to be addressed mainly by congestion pricing and traffic management
2. Supply shortage which could be addressed by constructing more roads or adding lanes

Literature had reported that public transportation is one of the most effective and promising methods to address traffic congestion issue by improving the available supply (Houghton, Reiners, & Lim, 2009) (Dewan & Ahmad, 2007). In Qatar, public transport buses service is providing an affordable and clean mode of transport to a wide range of the population (Shaaban & Khalil, 2013).

However, reliability, which is defined as adherence to bus actual arrival time to the scheduled timings is an issue. Passengers of public transport buses reported reliability of the existing system as poor (Shaaban & Khalil, 2013).

This issue could be addressed through:

1. Improving the route planning and scheduling by using more realistic and empirical models
2. Better quality control on buses using AVL systems and platforms
3. Informing passengers of actual or better estimation of arrival time through Advanced Traveler Information System (ATIS)

Accurate prediction of travel time for public transport buses is important for both the operators and users. Actual travel time can be used to inform passengers through various medium like electronic board displayed at the bus stops of bus arrival times

(Schweiger, 2003). Henceforth, the quality of public bus service is improved, which makes public transport system efficient and user-friendly. Accurate travel time information plays an important role in reducing passengers waiting time (Xinghao, 2013) and due to this the passengers can arrive at the bus stops on actual schedule time. This type of improved information will widen the commuters pool.

Bus-arrival-time information is an important component of an Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS). Predicting the bus arrival time requires development of model using appropriate tools, algorithms and techniques.

Public transport system in Qatar was addressed in only one peer-reviewed publication (Shaaban & Khalil, 2013). This research aims to be one of the foundations of knowledge based development of Qatar public transport system.

This research aims to develop a prediction model that can be a useful tool to public transport operators and concerned agencies to develop more realistic arrival time and travel time for buses. Developing a prediction model that is built based on historical data for several routes in Qatar can be useful to address the public transport travel time reliability. The model can be used in both planning and ATIS systems. Prediction model address the current gap that public transport buses operators in Qatar and the region face due to the absence of regional developed models to provide them with the proper tool to plan buses routes and travel time estimation based on empirical efforts.

In this research, Chapter 2 gives an overview of the previous efforts published in the literature and discusses theoretical and empirical models developed in this area of research. Chapter 3 states the research purpose. Overview, parameters and detailed description of the methodology is described in Chapter 4. In Chapter 5 data collected is described. Results supported by validation result are discussed in Chapter 6. Conclusion and potential future research are discussed in Chapter 07.

2. BACKGROUND AND LITERATURE SURVEY

Coming up with accurate prediction of the travel time for public transport buses is a concern to transit agencies. Providing those accurate predictions helps in improving the reliability of the bus system. Moreover, (Ghanim, Dion, & Abu-Lebdeh, 2014) (Dick Ettema, 2006) (Watkins, 2011) & (Tang, 2012) have proved that providing real-time information to passengers can improve the ridership of the system as well as its reliability from a passengers point of view.

In the past years, many theoretical and empirical methods for estimation and prediction of travel time for transit system were developed. In the below section, an overview of the published literature of the major developed methods is discussed.

2.1. THEORETICAL PUBLIC TRANSPORT TRAVEL TIME PREDICTION MODELS

Theoretical models reported in the literature are based on theoretical assumptions and data that cannot be available on a real-time basis. This forms a drawback to these models; especially considering the random factor of traffic demand, passenger demand and road incidents. Bus travel between successive stops was described using queuing models. (Vuchic, 2007) This approach was used by (Islam, 2015) to understand the reliability of buses in a public transport system. The study provided a better understanding of the public transit systems reliability, which can be used by public transport planners to analyze and evaluate different policies.

2.2. EMPIRICAL PUBLIC TRANSPORT TRAVEL TIME PERIDCITION MODELS

Considering the drawback in the practicality of theoretical models, researchers developed empirical models that adapt to real-time and historical information while considering the data availability. (Zhengyao Yu a, 2017) has classified the developed empirical models to speed-based models, classic regression models and artificial neural network (ANN) models. The published literature for most of these empirical models are reviewed and summarized in the below subsections.

2.2.1. SPEED-BASED MODELS

Kalman Filter and Particle Filter are typical methodologies that are used to develop speed-based travel time prediction models, where real-time and historical speed data are given weightage factors based on the public transport bus location in the route. In these techniques, the real-time speed gains more weight in the filter equation.

(Shalaby, 2004) uses Kalman Filter to develop a prediction model for arrival and departure time of buses. They compare the developed model and microsimulation model (VISSIM), where both models are tested against historical data. Kalman Filter model used historical data collected through Automatic Vehicle Locator (AVL) and Automatic Passenger Counter (APC).

Root Mean Square Error (RMSE), mean relative error, and maximum relative error are used to compare the results of both models. Results reported by (Shalaby, 2004) showed that Kalman Filter model leads to a lower error in prediction time compared to VISSIM modeling of public transport buses travel time.

(Etienne Hans, 2015) develops a general prediction model for bus travels along a route, using data collected in Portland, Oregon (USA) to produce and calibrate the model. (Etienne Hans, 2015) uses Particle Filter data assimilation technique to forecast bus trajectories. The proposed methodology uses historical and real-time data simultaneously. Using intelligent transport system (ITS) network infrastructure, buses send arrival and departure information from each stop to Particle Filter model. The model calibrates and produces calibrated trajectory arrival times using received data as well as historical data.

Produced information of arrival time using the model are valid within 8 minutes, which gives the agencies enough time to apply appropriate strategies to reduce bus bunching from the schedule. (Etienne Hans, 2015) reported that two stops prediction information is utterly irrelevant to represent 5% of the total generated predictions within 8 minutes' time. The advantage of this live model that real-time information is considered in the process of prediction.

(Xinghao, 2013) uses exponential smoothing algorithm to predict bus running speed based on speeds of buses and taxis within the same area. The model had considered traffic control delays, acceleration, and deceleration of buses. The predication model is based on manual data collection and historical record data of buses in Shanghai. The analyzed data is a combination of two months simulated Radio Frequency Identification (RFID) data for two bus routes and Automatic Vehicle Location (AVL) data of majority of buses in Shanghai. The factors considered in developing the model are speed, red time, green time, distance and location of the bus. The study findings where (1) taxis speed and bus speed are highly correlated within one segment (2) the performance of model used combination of the AVL and simulated RFID data is superior to models that used historical data from AVL only. This can be explained by the fact that the models which use simulated RFID consider traffic control delays within the model the proposed model.

2.2.2. PUBLIC TRANSPORT BUSES TRAVEL TIME PREDICTION REGRESSION MODELS

Regression models use factors such as route length, demand, heavy vehicle percentage, dwell time, number of stops, peak hours, etc. To build models that can predict travel time. Using regression models to build a clear relationship between factors and travel time were studied in various studies in the literature which will be discussed in this subsection. (Patnaik, 2004) adapt information from Automatic Passenger Counters (APC) to develop a multivariate regression model to predicate arrival time of the public transport buses in New Jersey, USA. In the study factors, such as passenger demand, distance, dwell time and time of the trip were included in the model. The results showed good fit of the developed model against historical data and provided promising results of the developed regression model.

(Zhengyao Yu a, 2017) use accelerated failure time log-logistic survival models to provide simultaneous predictions of public transport buses. Data was collected from one circular 6.6 km long route. The data was collected only on weekdays for almost sixteen months; due to high variability in weekends as reported by the author. The study considers factors such as number of passengers boarding and alighting at each stop, occupancy, dwell time, temperature, participation, snow depth and headway. The study uses linear regression model results to be compared with log-logistic survival model results. Comparison of the results showed that log-logistic survival model outperformed liner regression models. The study concluded that log-logistic is a promising technique that can provide not only arrival time predictions, but it reports the uncertainties level accompanying these predictions.

2.2.3. PUBLIC TRANSPORT BUSES TRAVEL TIME PREDICTION MODEL USING ARTIFICIAL NEURAL NETWORK (ANN)

Machine learning models are beneficial data mining techniques that can relate many variables that have no apparent relationship with each other; to build a model that uses hidden layers of neurons that connect these variables to each other in one model. In literature majority of ANN models outperformed regression models (Jeong, 2004).

(Johar Amita, 2016) developed a model using ANN with GPS data input to predict travel time of public transport buses in Delhi, India. The data was collected from one route over 6 days. The study considers factors such as number of passengers boarding and alighting, bus running time and dwell times. The performance of the model is evaluated against Linear Regression Model using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error. The study results showed that ANN model outperformed regression model in travel time prediction. Also, (Chien, 2003) builds an ANN model to predict travel time of public transport buses using an adjustment factor based on real data. (Chien, 2003) uses CORSIM microsimulation software to simulate passenger demand on the system. ANN model then is compared to CORSIM outputs. ANN is found to outperform CORSIM simulation outputs. Afterwards, (Jeong, 2004) builds an ANN model to estimate the arrival time of bus using data collected from automatic vehicle location (AVL). The study uses dwell time and traffic congestion, represented by schedule adherence, related factor as variables to the model. The developed ANN model is compared to historical model and regression model. Using mean absolute percentage error (MAPE), the developed ANN model showed that it outperformed other models.

(Wei Fan, 2015) uses only global positioning system (GPS) data for one year for a route in Brazil to produce three models; Historical Average (HA), Kalman Filtering (KF) and Artificial Neural Network (ANN). The three developed models were compared and assessed in terms of overall prediction accuracy and robustness. The results show that ANN developed model outperformed the other models, even when traffic conditions are not considered in the model.

2.3. LITERATURE SUMMARY

Prediction models for public transport buses are developed using several different methodologies and techniques. Majority of travel time prediction model studies used one of these methodologies to evaluate either travel time or arrival time of public transport buses:

1. Kalman Filter
2. Particle Filter
3. Regression
4. Microsimulation Model
5. Artificial Neural Networks (ANN)

Each study verified the developed model either by comparing it to another model or by comparing it to real-data that was not included in the model development by measures of effectiveness (MOEs). Table 2.1 summarizes the reviewed literature in terms of models developed and the validation methodologies.

Table 2.1: Summary of Methodologies Used to Develop Prediction Models

			Developed Models					Validation Technique					
Paper Title	**Author**	**Country**	Kalman Filter	Particle Filter	Regression	Microsimulation Model	ANN	Kalman Filter	Particle Filter	Regression	Microsimulation Model	ANN	Real Data
Real-time bus route state forecasting using Particle Filter: An empirical data application	(Etienne Hans, 2015)	Portland, Oregon, USA		●									●
Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data	(Shalaby, 2004)	Toronto, Canada	●			●		●		●	●	●	
Estimation of Bus Arrival Times Using APC Data	(Patnaik, 2004)	New Jersey, USA			●								●
Using survival models to estimate bus travel times and associated uncertainties	(Zhengyao Yu a, 2017)	Pennsylvania, USA			●					●			●

*Continued

Paper Title	Author	Country	Developed Models					Validation Technique					
			Kalman Filter	Particle Filter	Regression	Microsimulation Model	ANN	Kalman Filter	Particle Filter	Regression	Microsimulation Model	ANN	Real Data
Predicting bus real-time travel time basing on both GPS and RFID data	(Xinghao, 2013)	Shanghai, China			●								●
Prediction of Bus Travel Time using ANN: A Case Study in Delhi	(Johar Amita, 2016)	Delhi, India					●			●			
Dynamic Bus Arrival Time Prediction with Artificial Neural Networks	(Chien, 2002)	New Jersey, USA					●				●		
Bus Arrival Time Prediction Using Artificial Neural Network Model	(Jeong, 2004)	Houston, TX, USA					●			●			●
Dynamic Travel Time Prediction Models for Buses Using Only GPS Data	(Wei Fan, 2015)	Macae, Brazil	●		●		●	●		●		●	●

Parameters or variables used in model development vary significantly from one model to another; with a high dependency on data availability. Some studies include passenger demand and passenger related variables while others use traffic condition-related factors. However, the significance of the impact of these variables are reported to be inconstant from one model to another. Interestingly, most of the reviewed studies showed promising results and model accuracy. Table 2.2 summarized the variables used in each study as an input to the model.

Also, sample size and selection process are a very critical process in developing any model. Unfortunately, empirical models developed and reviewed in literature used tiny samples, which could question the generalization of those models. For example, using one route that passes through residential areas only to develop the model can be questionable when it is used to predict travel time of public transport buses that passes through commercial areas. Table 2.3 summarizes the data collected in each of the published studies reviewed.

Table 2.2: Variables Used in Models Developed

Paper Title	Author	Country	Speed	Distance	Arrival & Departure Time	Passenger Demand	Peak	Dwell Time	No. of Stops	Weather	Average link traffic volume	Average link speed	Average link delay	Average link queue time	Number of intersections	Signal Red and Green Timings
Real-time bus route state forecasting using Particle Filter: An empirical data application	(Etienne Hans, 2015)	Portland, Oregon, USA	●	●	●											
Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data	(Shalaby, 2004)	Toronto, Canada	●	●	●											
Estimation of Bus Arrival Times Using APC Data	(Patnaik, 2004)	New Jersey, USA	●	●	●	●	●	●	●							
Using survival models to estimate bus travel times and associated uncertainties	(Zhengyao Yu a, 2017)	Pennsylvania, USA	●	●	●	●	●	●		●						

*Continued

Paper Title	Author	Country	Speed	Distance	Arrival & Departure Time	Passenger Demand	Peak	Dwell Time	No. of Stops	Weather	Average link traffic volume	Average link speed	Average link delay	Average link queue time	Number of intersections	Signal Red and Green Timings
Predicting bus real-time travel time basing on both GPS and RFID data	(Xinghao, 2013)	Shanghai, China		●	●		●		●							●
Prediction of Bus Travel Time using ANN: A Case Study in Delhi	(Johar Amita, 2016)	Delhi, India		●		●	●	●								
Dynamic Bus Arrival Time Prediction with Artificial Neural Networks	(Chien, 2002)	New Jersey, USA	●	●					●		●	●	●	●	●	
Bus Arrival Time Prediction Using Artificial Neural Network Model	(Jeong, 2004)	Houston, TX, USA	●	●			●	●	●							
Dynamic Travel Time Prediction Models for Buses Using Only GPS Data	(Wei Fan, 2015)	Macae, Brazil		●	●		●		●							

Table 2.3: Literature Summary – Data Used for Model Development

Paper Title	Author	Country	Total Number of Routes	Total Duration (Days)	Total No. of Records Analyzed
Real-time bus route state forecasting using Particle Filter: An empirical data application	(Etienne Hans, 2015)	Portland, Oregon, USA	1	44	NA
Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data	(Shalaby, 2004)	Toronto, Canada	1	30	NA
Estimation of Bus Arrival Times Using APC Data	(Patnaik, 2004)	New Jersey, USA	1	180	311
Using survival models to estimate bus travel times and associated uncertainties	(Zhengyao Yu a, 2017)	Pennsylvania, USA	1	320	NA
Predicting bus real-time travel time basing on both GPS and RFID data	(Xinghao, 2013)	Shanghai, China	NA	30	8,743,709

*Continued

Paper Title	Author	Country	Total Number of Routes	Total Duration (Days)	Total No. of Records Analyzed
Prediction of Bus Travel Time using ANN: A Case Study in Delhi	(Johar Amita, 2016)	Delhi, India	1	6	NA
Dynamic Bus Arrival Time Prediction with Artificial Neural Networks	(Chien, 2002)	New Jersey, USA	1	NA	NA
Bus Arrival Time Prediction Using Artificial Neural Network Model	(Jeong, 2004)	Houston, TX, USA	1	180	NA
Dynamic Travel Time Prediction Models for Buses Using Only GPS Data	(Wei Fan, 2015)	Macae, Brazil	1	365	NA

3. RESEARCH OBJECTIVE

In Qatar, literature reported that customer satisfaction of public transit (bus service) is significantly impacted by punctuality, travel time, frequency, and ticket cost, with emphasis on the punctuality and the time of the trip. Poor reliability (punctuality) can be addressed on multiple levels, such as policies, planning, quality control and better communication with passengers through Intelligent Transportation Systems (ITS) (Shaaban & Khalil, 2013).

The absence of realistic prediction model for public transport buses travel time prediction lead to poor planning, quality control and inaccurate real-time advanced traveler information of public transport buses travel time.

The objectives of this research are to address this issue by:

1. Analyzing historical data of several public transport routes in urbanized areas in Doha. The data were collected over one year. The aim of the analysis is to understand the possible factors that may impact buses travel time using appropriate techniques and tools.
2. Developing a generalized transit route travel time prediction model using the collected historical data from different geographical areas. The model aims to predict the travel time of buses using several independent variables that are unique for each transit route.
3. Analyzing and developing a reduced prediction model of public transport travel time by omitting noisy variables to optimize the amount of data needed to predict travel time and arrival time of public transport buses.

By meeting the objectives mentioned above; public transport planning in Qatar can be improved to meet Qatar Vision 2030 objectives of sustainability, economic growth and improving safety that yields to human resources development.

Moreover, this importance of reliable, safe and efficient public transit buses will significantly increase by the time Qatar Rail will complete Doha Metro Project early 2020. The ridership and demand on public transit is expected to increase significantly due to the need of multimodal transport model that metro commuters will end up using. This research aims to improve those commuters' experience and reduce congestion and travel overall cost on the economy.

Knowledge gap lays in the absence of generalized prediction model that can provide a prediction of the travel time of public transit buses. Public transit prediction models reported in the literature were developed either in the US, China or European countries; where city planning, driver behavior, and public transit operations do not match Qatar conditions. Developing a model that is built using large database within the region is unique. In addition to addressing the current gap in knowledge, prediction model for public transit buses will provide the opportunity and tools to enhance public transit commuters experience in the region.

Furthermore, this research aims to be the stimulator and base at which public transit studies will be built upon in the region. Further research and applied engineering of the prediction model can significantly improve customers experience and widen the public transit system users pool.

4. RESEARCH METHODOLOGY

4.1. DATA COLLECTION

Data were obtained from the only operator for public transport, Mowasalat, buses in Qatar, after several months of communications and challenges with different teams in Mowasalat. The communication chain started from the Chief Executive officer (CEO) and went down to many departments within the organization; to be able to collect the data for this research.

Based on disclaimer signed with Mowasalat, the collected data shall not be shared without informing concerned parties. Moreover, it can be used only for academic and this research purposes.

An agreement was reached where Mowasalat can share with us a limited number of routes data only. Thus, the following criteria were used to filter the routes that will be included in the study and requested from Mowasalat:

1. Data availability and files free of error and missing records
2. The route should be within Doha, shown in Figure 4.1, and Al-Rayyan, shown in Figure 4.2, municipalities only; since these are the major two municipalities that are urbanized and covered with buses public transport system in Qatar.

Figure 4.1: Doha Municipality Territories

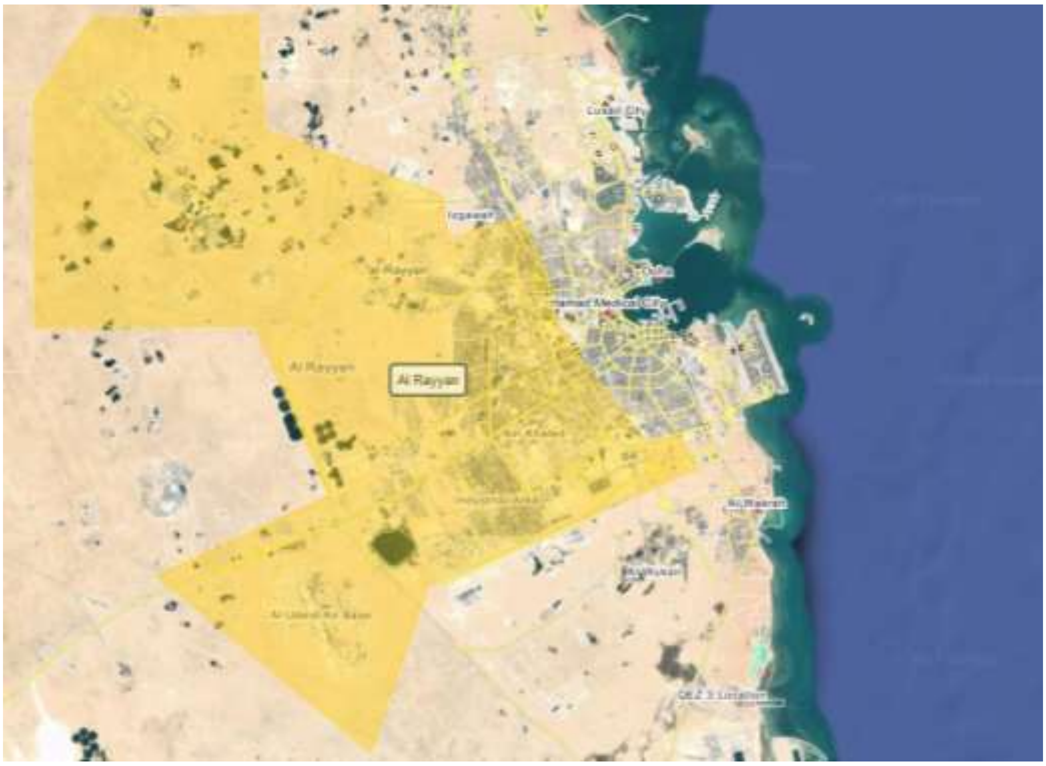


Figure 4.2: Al Rayyan Municipality Territories

3. Routes shall not be particular case routes. For example, route no. 74 which is a circular route within Westbay area was excluded. This route is exceptional route was for the route was free of charge between 15th January 2014 and 1st April 2016. Moreover, this route went through several stages of rerouting in the same period which can impact the study reliability (Marhaba, 2016)

Based on the above criteria, data for twelve routes of public transport buses were collected from Mowasalat. The shared data was generated by KentKart platform, which is "the Automated Fare Collection, Vehicle Management, Real Time Passenger Information, Planning and On-Board Video Surveillance Systems"

(Kentkart, 2017) that is being used by Mowasalat. Moreover, route map of most recent bus routes was shared with us see Figure 4.3.

Some of these routes were operated in circular bases, where the bus starts from a station and ends into it as a terminal station. While others were operated in outgoing direction bus that goes from the central station(s) towards the other bus stops and terminates its trip in the last stop of the route that is not the same station it started with. At the same time, there are in-going buses that start from a station nearby the terminal station of the outgoing station towards the central station. This type of operated bus routes had different geometric characteristics. Thus, it can be assumed as two independent bus routes that were analyzed in the study.

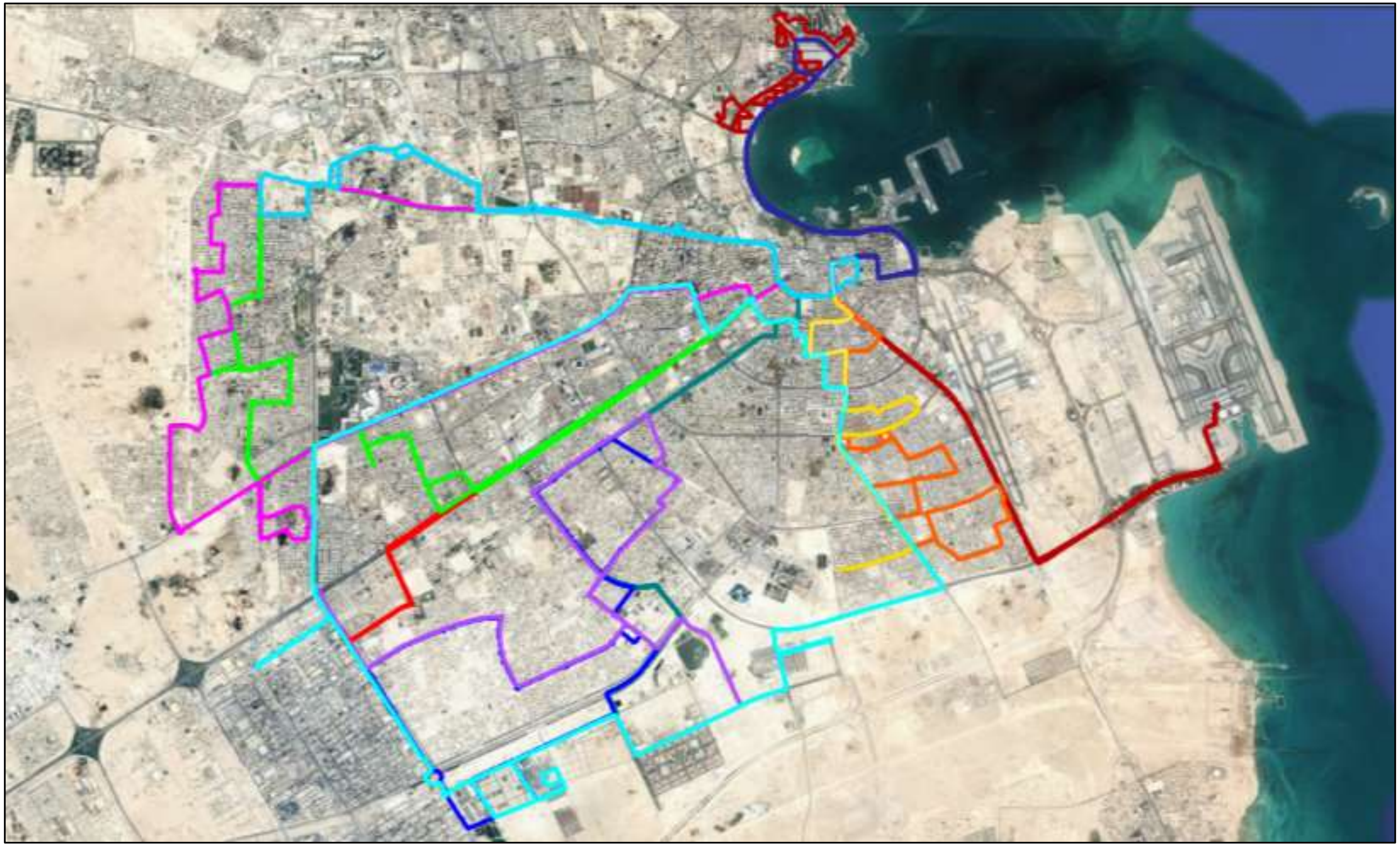


Figure 4.3: Google Earth Image Showing the 12 Routes Used in the Study

Kentkart generated data had included five parameters, Table 4.1 below provides clear description of each parameter collected. One-year data records of trips for each of the twelve routes with the below-mentioned parameters were collected and tabulated.

Table 4.1: Collected Variables Description

Parameter	Description
Route Code	A unique code for each route within the public transport network.
Direction	Description of the direction that buses are traveling on. The description is formatted as [name of trip start station/area-name of trip end station/area]
Trip Start Time	The time that trip started traveling from the starting station. The format of the trip start time is DD-MM-YYY HH:MM:SS AM/PM
Trip End Time	The time that bus stopped at last station of the trip. The format of the trip end time is DD-MM-YYY HH:MM:SS AM/PM
Distance	Distance traveled between starting station and last station in the trip in kilometers

4.2. DATA MINING TECHNIQUE

With consideration to the amount of data collected, data mining techniques were explored to select the most applicable technique; to analyze the data of more than 150,000 records. Predicting future trends that allows proactive knowledge-based decision-making is one of the main applications of data mining techniques.

Modeling techniques that were sorted under data mining category are big amount of data points. (Lee, 2001) Techniques such as Statistics, Clustering, Association Rules and Artificial Neural Networks were explored for its advantages, disadvantages, and application based on (Moawia Elfaki Yahia, 2010). Based on this publication, Artificial Neural Networks was selected as analysis techniques for collected data.

The artificial Neural network can be used for complex classification situations, clustering, feature extraction, prediction work and outlier analysis. Artificial Neural Networks generates relationships between inputs and outputs without any external guidance. However, it is complicated to understand how the input data impacts the output prediction (Moawia Elfaki Yahia, 2010).

The below section will overview the details of Artificial Neural Networks (ANN).

4.2.1. ARTIFICIAL NEURAL NETWORKS - OVERVIEW AND HISTORY

Inspired by biology, the naming of Artificial Neural Networks (ANN) was used to describe the computational neuronal structure of processing elements (neurons) connected with bound that have coefficients (weights); which is used to train and recall algorithms (Kasabov, 1995). (T. Kohonene, 1998) described Artificial Neural networks (ANN) as parallel regression, non-linear, multi-layered techniques mathematical techniques that are used for forecasting, signal processing, and clustering. ANN can be used for fitting a line through a set of data points where relationships could exist between users defined inputs and outputs as per (T. Kohonene, 1998) description.

Going back to 1943, McCulloch and Pitts were the first to introduce the concept of artificial neural networks in their publication titled “A logical calculus of the ideas immanent in nervous activity”. Four years later, McCulloch and Pitts used the term “perceptron” to describe a simple neuron model network mathematical system.

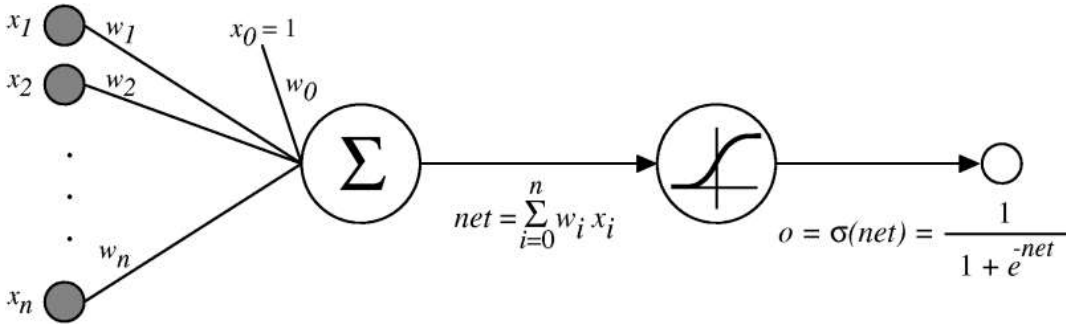


Figure 4.4: Perceptron Neuron (Mitchell, 1997)

Later, massive research efforts were made to develop algorithms such as associative memories, multi-layer perceptron (MLP), back propagation, learning algorithm, adaptive resonance theory (ART), self-organizing networks, bi-directional associative, memory introduced, radial basis function, fuzzy neurons and fuzzy networks presented, oscillatory neurons, oscillatory neural networks and many more.

4.2.2. ANN ARCHITECTURE AND METHODOLOGY

The behaviour of biological neurons was used to create a mathematical model of artificial neurons (units). The mathematical model follows the following steps:

1. Neuron receives its inputs through input connections
2. Neuron calculates a weighted sum of the inputs, called potential, donated by:

$$p_j = \sum_{i=1}^{N+1} w_{ij}\, a_i \qquad \textbf{Equation 4-1}$$

   Where;

   p_j is the potential (neuron) net,

   N is the number of input varaibles,

   w_{ij} denotes weight of the connection from unit i to unit j, and

   a_j is the activation function

3. Unit's activation is computed from the potential and sent to other units.
4. Weights of connections between units are stored in a matrix w, where w_{ij} denotes weight of the connection from unit i to unit j.
5. Bias is typically represented as an extra input unit with activation equals to one, therefore $a_{N+1} = 1$. This enables shifting the activation function along x-axis by changing the weight of connection from threshold unit.
6. Activation of the unit a_j is computed by transforming p_j by a non-linear activation function. several activation functions exist and used in ANN architectures and model development. The most common type of activation functions logistic function and the hyperbolic tangent function.

a. **Logistic Function:** The logistic is a sigmoid function with the range (0;1).

$$a_j = \frac{1}{1 + e^{-p_j}} \qquad \textbf{Equation 4-2}$$

Where;

a_j is the activation function (logistic function) and

p_j is the potential (neuron) net

The parameter affects the slope of the function which becomes a threshold function when the function apporaches infinity as shown in Figure 4.5:

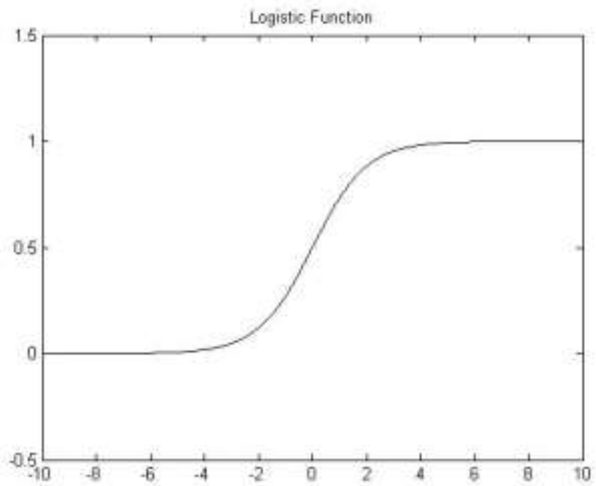


Figure 4.5 Plot of Logistic Activation Function Where a Equals 1

b. **Hyperbolic Tangent Function:** This function is another sigmoid function, with the range (-1;1).

$$a_j = \frac{e^{p_j} - e^{-p_j}}{e^{p_j} + e^{-p_j}} \qquad \textbf{Equation 4-3}$$

Where;

a_j is the activation function (hyperbolic tangent function) and

p_j is the potential (neuron) net,

c. Hyperbolic tangent function has the ability to reduce the effect of outliers in the data since; large input values are squeezed to the limits of the hyperbolic tangent function as shown in Figure 4.6:

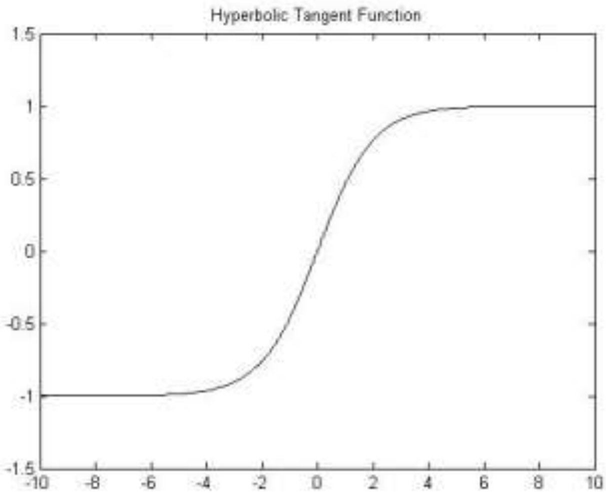


Figure 4.6 Plot of Hyperbolic Tangent Activation Function Where a Equals 1

The results of the hidden layer are transferred to output layer using linear transfer function shown in and denoted by:

$$purelin\ (x) = x \qquad \textbf{Equation 4-4}$$

Where;

x is the net of output layer matrix

4.2.3. NETWORK DEFINITION

Four distinct parameters usually define ANN. The parameters are as follows:

1. The neuron type, which is the parameter that represents each of the network's nodes. It was referred to by (W.S. McCulloch, 1947) as a form of Perceptron and by (Yamakawa, 1990) as a form of the Fuzzy neuron.
2. The architecture of the network connections, which can be presented as one of following three methods:
    a. Depending on the connection topology between the neurons (the neurons can be fully or partially connected)
    b. Depending on the number of neurons for input and output and the layers of neurons used, which is classified as:
        (i) Auto-associative: where neurons of inputs and outputs are the same neurons
        (ii) Hetero-associative: where there are different output and input neurons
    c. Depending connections on existence back of output to the inputs:
        (i) Feedforward architecture, illustrated in Figure 4.7, where no connections from output back to input; thus, the network does not remember the output of the previous neuron or input

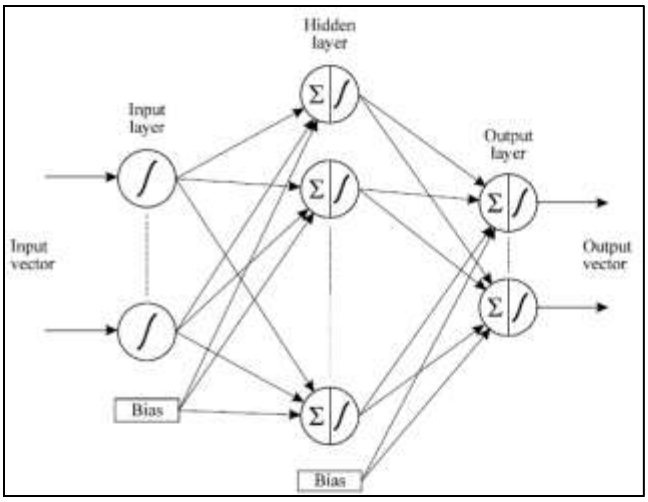


Figure 4.7 Typical One Hidden Layer Feedforward ANN Architecture (Teodorović & Vukadinovic, 1998)

(ii) Feedback architecture at which connection between outputs results are reported back to the neuron and impact the results of the next output

3. The learning algorithm, which is the algorithm in which the network is trained.

4.2.4. LEARNING ALGORITHM

The learning algorithm can be summarized in three groups; supervised learning, unsupervised learning, and reinforced learning. The three groups are defined as follows:

1. Supervised learning: This type of algorithm training uses to target and input vectors to train the network. These vectors are required by the network to build the possible relationships between the inputs and the target or the output.

2. Unsupervised learning: This type of algorithm training does not include any target vectors or desired results. Instead, the network tries to build connections between the inputs that can be generalized to the training set.
3. Reinforcement learning: In this type of algorithm learning, the neural network is trained to develop the relationships based on the feedback of the generated output.

4.2.5. USED ARTIFICIAL NEURAL NETWORKS PARAMETERS

Data analysis and model development using ANN shall go through systematic processes. Figure 4.8 below illustrates these processes flow diagram.

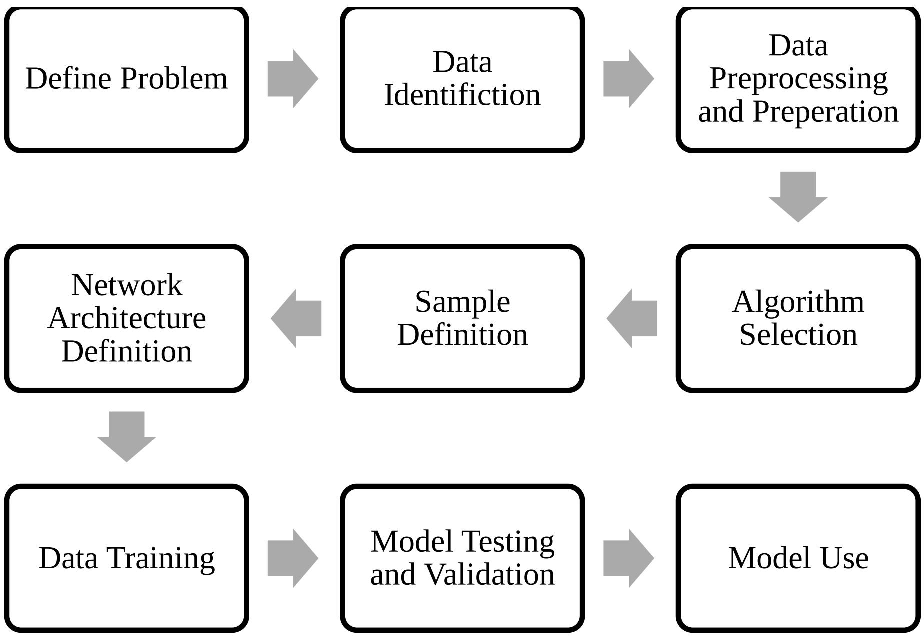


Figure 4.8: Model Development Process

Previous Chapters had discussed the first three processes of the model development process is documented. Each of the remaining processes is described in the below paragraphs.

4.2.5.1. Algorithm Selection

(Wilamowski, 2011) & (O. Kisi, 2005) recommend several learning algorithms for regression applications. However, Levenberg–Marquardt algorithm was recommended in the majority of cases where big amount of data is present since it converges in a fast and stable manner (Wilamowski, 2011).

The Levenberg–Marquardt (LM) algorithm combines the steepest descent method and the Gauss-Newton algorithm in one algorithm. LM takes advantage of Gauss-Newton algorithm speed and steepest descent method stability. Moreover, LM results in more robust algorithm than the Gauss-Newton algorithm. While Gauss-Newton algorithm converges faster than, Levenberg–Marquardt algorithm; however, LM converges much faster than steepest descent method (Wilamowski, 2011). This leads to lower computational cost and period, which is essential when big data sets as well as large number of inputs is used. Section 7.4 discuss further benefits to the use of Levenberg–Marquardt algorithm in training travel time data sets to predict public transit travel time.

An overview of the training process using Levenberg–Marquardt algorithm is described in systematic representation in Figure 4.9.

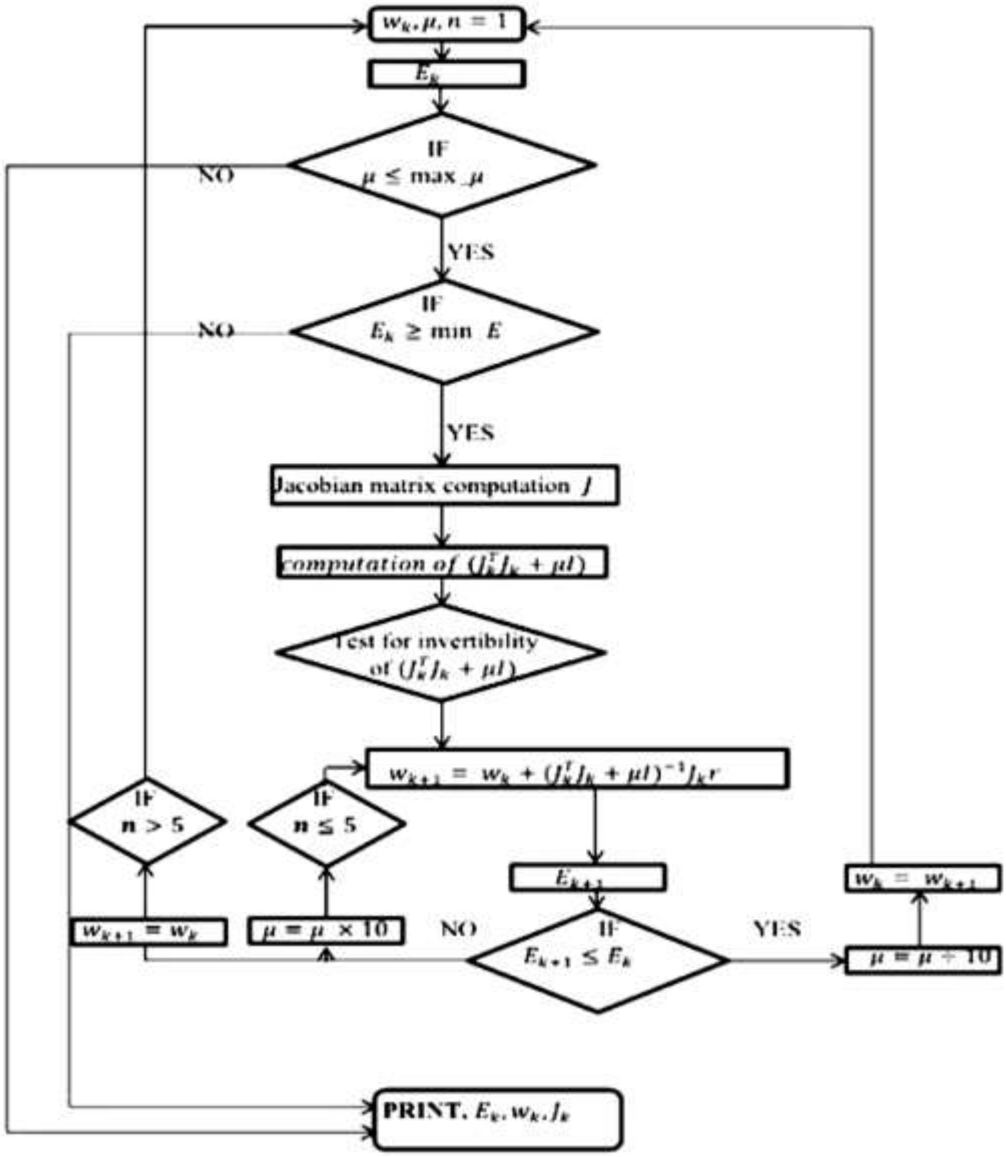


Figure 4.9: Levenberg-Marquardt Algorithm Flowchart (Ayegba & Abdulkadir, 2016)

The training process diagram follows the following steps:

(a) Sum square error e_k is estimated by starting with randomly initial weights w_k

(b) Weights are adjusted using the update rule of:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - \left(\boldsymbol{J}_k^T \boldsymbol{J}_k + \mu \boldsymbol{I}\right)^{-1} \boldsymbol{J}_k \boldsymbol{e}_k \qquad \textbf{Equation 4-5}$$

Where,

$\boldsymbol{w_k}$ is the initial weights

μ is the combination coefficient is always positive,

I is the identity matrix

n is the conditions for stopping the training, such as number of failed iterations or time limits. In the figure above, 5 donates a stopping criteria for the training which is the iterations limit.

and $\boldsymbol{J_k}$ is the Jacobian matrix given by;

$$J_k = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial w_1} & \frac{\partial e_{1,1}}{\partial w_2} & \cdots & \frac{\partial e_{1,1}}{\partial w_N} \\ \frac{\partial e_{1,2}}{\partial w_1} & \frac{\partial e_{1,2}}{\partial w_2} & & \frac{\partial e_{1,2}}{\partial w_N} \\ \cdots & \cdots & & \cdots \\ \frac{\partial e_{1,M}}{\partial w_1} & \frac{\partial e_{1,M}}{\partial w_2} & \cdots & \frac{\partial e_{1,M}}{\partial w_N} \\ \frac{\partial e_{P,1}}{\partial w_1} & \frac{\partial e_{P,1}}{\partial w_2} & \cdots & \frac{\partial e_{1,1}}{\partial w_N} \\ \frac{\partial e_{P,2}}{\partial w_1} & \frac{\partial e_{P,2}}{\partial w_2} & \cdots & \frac{\partial e_{P,2}}{\partial w_N} \\ \cdots & \cdots & & \cdots \\ \frac{\partial e_{P,M}}{\partial w_1} & \frac{\partial e_{P,M}}{\partial w_2} & & \frac{\partial e_{P,M}}{\partial w_N} \end{bmatrix} \text{Where e is given by } e = \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ \cdots \\ e_{1,M} \\ \cdots \\ e_{P,1} \\ e_{P,2} \\ \cdots \\ e_{P,M} \end{bmatrix}$$

(c) Using w_{k+1} sum square error is computed again

(d) If, e_{k+1} is increased, then reset the weight vector to the previous value and increase combination coefficient μ by certain factor. Go to step (b) and repeat

(e) If, e_{k+1} is decreased, then keep the weight vector to the current value and decrease combination coefficient μ by certain factor. Go to step (b) and repeat

(f) Repeat steps (b) to (e) until $e_{k+1} \leq e_{Target}$

4.2.5.2. Routes Selection Criteria

Sample definition or selection process that was followed took into consideration the several factors to ensure reliability and model generalization. Selection of routes for training was based on the below criteria:

1. Selection of verification routes/data randomly before selection of training routes/data
2. Located in different geographical areas of Doha City
3. Covers all operational scenarios of buses; where some buses were operated in circular operations at which bus starts and ends in the same station. While other routes were operated in a unidirectional fashion.

Based on the above criteria, nine routes were selected for data training, while remaining three routes that were selected randomly were not included in the model development and used for validation and verification.

4.3. DEFINITIONS OF INPUT VARIABLES

Travel period, length, number of passengers, surrounding land use, rainy weather and number of signalized intersections were found to be significant variables on travel time (ENGELSTEIN, 1982), (Verbich, Diab2, & El-Geneidy, 2016) & (Mazloumi, Currie, & Rose, 2010). During development of the study methodologies, published findings of these researchers were taken into consideration. However, the applicability of these factors on Qatari case as well as data availability played a significant role in identifying the variables that can be included in the study.

Factors such as rainfall cannot be applicable in Qatar since highest monthly average rainfall between 1962 and 2013 was 18.1mm (Qatar Meteorology Department, 2017); which is negligible compared to North America, Australia, and Europe.

Other variables were not included in this research due to lack of information such as boarding and alighting passenger numbers.

However, other factors that literature had not investigated, such as the impact of turning movement of buses on the travel time of public transport buses. Signalized roundabouts are an exceptional condition that Qatar transport network had used widely.

Based on the above conditions and limitations, a variable included in this research are defined and described in the below subsections based on data collection methodology as follows:

1. Subsection 4.3.1 describes trip related parameters. These parameters were collected from Mowasalat per trip. Parameters, as described in this subsection, vary from trip to trip.
2. Subsection 4.3.2 describes route related parameters. These parameters were extracted manually from public transit buses map received from Mowasalat. Parameters, as described in this subsection, are fixed for each route.

4.3.1. TRIP RELATED PARAMETERS

4.3.1.1. Distance

Distance traveled between starting station and last station in the trip in kilometers that were collected using AVL system in buses. Earlier researchers emphasized and agreed that this variable has the most significant impact on the travel time of public transport buses. (ENGELSTEIN, 1982), (Verbich, Diab2, & El-Geneidy, 2016) & (Mazloumi, Currie, & Rose, 2010).

4.3.1.2. Weekday

This variable shows the day of the week (Sunday, Monday…etc.). Its significance may have an impact on travel time based on the variation of traffic conditions daily traffic of the week.

4.3.1.3. Weekend

Due to the variation in traffic condition between working weekdays and weekend, this variable allows the model to visualize this difference. The model had two input value either 1 if the day of the trip is weekend (Friday or Saturday) or 0 otherwise.

4.3.1.4. Peak Hour

A variable that identifies which peak the trip started on. Per Qatar Ministry of Municipality and Urban Planning guidelines, AM peak starts from 6:00 to 9:00, Mid-day peak starts from 11:00 to 14:00 and PM peak starts from 17:00 to 20:00. While a trip that started during the inter-peak time were considered as off-peak trips. Literature had reported that this variable has a significant impact on travel times of public transport buses.

4.3.1.5. Month of the year

To account for seasonal traffic conditions fluctuations, the month that trips started on starting from January to December.

4.3.1.6. Travel time

The travel time of buses from the time the trip started at the first station to last station. This variable is the target data the prediction model was trained to predict; using historical data that include travel time within its records.

4.3.2. ROUTE RELATED PARAMETERS

4.3.2.1. Number of Through Movements

This parameter defines the number of through turning movement on major junctions (traffic signals and roundabouts).

4.3.2.2. Number of Left-turn Movements

This parameter defines the number of left-turn turning movement on major junctions (traffic signals and roundabouts).

4.3.2.3. Number of Right-turn Movements

This parameter defines the number of right-turn turning movement on major junctions (traffic signals and roundabouts).

4.3.2.4. Number of U-Turns Movements

This parameter defines the number of U-turns turning movement on major junctions (traffic signals).

4.3.2.5. Number of Signalized Intersections

Number of signalized intersections that the bus went through within the route. This factor was found to be significantly impacting the travel time and travel time variability of public transport routes (ENGELSTEIN, 1982), (Verbich, Diab2, & El-Geneidy, 2016) & (Mazloumi, Currie, & Rose, 2010).

4.3.2.6. Number of Roundabouts

Number of roundabouts that the bus went through within the route.

4.3.2.7. Number of Stops

Number of bus stops that bus went through, counting first and last stations.

4.3.2.8. Central Business District (CBD)

Per Qatar Ministry of Municipality and Urban Planning guidelines, the location of a bus route in relation to the central business district (CBD). Three categories were used inner-CBD, outer-CBD and non-CBD areas.

4.3.2.9. Land use

Surrounding land use of areas that bus route is going through was found to be significantly affecting travel time of public transport buses (ENGELSTEIN, 1982). The areas that routes went through were categorized to: commercial, industrial, residential and mixed land uses.

4.3.2.10. Road Functional Classification

There are several methods to classify the roads can be used. However, since route can go through vast number of roads with different classifications. Thus, two major classifications were used in this research, which are major roads and local roads. Major roads were assumed to be arterial, major collectors, expressway, and highways, while local roads were assumed to be any lower classification roads.

4.4. VALIDATION METHODOLOGY

The validation process is considered the final gateway before approving the developed prediction model release. The prediction model will be validated using routes that were randomly chosen from the data collected. Those routes will not be included in the prediction model development. The validation routes are used to compare the travel time recorded and collected from historical data versus predicted travel time generated by the developed model.

It is important to emphasis on the randomness of the validation routes selection process; to assure the generalization of the model. It is important to mention that validation routes or data shall go in the same data preparation process that prediction model inputs went through.

4.4.1. MEASURES OF EFFECTIVENESS (MOE)

Assessing the applicability of the model in quantitative measures is very curial part of the model validation. In literature, the following measures of effectiveness (MOE) were used for measuring model applicability (Ghanim, Abu-Lebdeh, & Ahmed, 2013):

1. The coefficient of Determination (R^2): Which measures the variation in model results compared to targeted or measured results.

$$r = \left(\frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}}\right)^2$$

   Mean of absolute error (MAE): which measures the error of predicted travel time to collect travel time using the following equation

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |x_t - \hat{x}_t|$$

Where:

n *is the number of points*

x_t *is the travel time collected*

$\hat{x}_t$ *is the predicted travel time*

2. Standard error (SE): which represents the residuals between the actual and the predicted travel time calculated using the following equation:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Where:

n is the number of points

σ_x is the standard devation

4.5. DATA PROCESSING

Input data used for both travel time prediction model development and validation routes went through several data processing stages. The following stages improved model processing efficiency and resulted in higher accuracy. Multiple of following steps are recommended by the MATLAB software development team (MathWorks, 2017). The below paragraphs details each step of the pre-processing process:

4.5.1. DATA EXPLORATION

Collected data were explored to understand the content of the data files and its encoding. In this stage data received in multiple files per route was merged into one database file. The data received included parameters described in Table 4.1. Using the received data the following information was extracted regarding the following parameters:

1. Distance Traveled in km
2. Month the trip took place in
3. Weekday (Sunday, Monday…etc.)
4. Peak the trip started in (Morning, Midday, Evening or off-Peak)
5. Travel Time of the bus

Using the above parameters, the data was explored more to make sense of the received information.

4.5.2. DATA ENCODING

Using both received data extracts from the software and the routes drawings, model parameters we encoded as described in the below table. Binary encoding provides advantages to the model prediction development in terms of processing time and data interpretation processing. In other words, defining peak hours as 1, 2, 3 and 4 for AM, MD, PM, and off-Peak cannot be interoperated with fractions.

Table 4.2 Data Encoding

Parameter	**Data Encoding Type**	**Encoding Description**
Distance	Numerical	Distance as recorded by AVL; rounded to nearest 1 km
Month	Binary	12 Columns (one for each month) were defined at which each column accepts either 0 or 1 as value
Weekday	Binary	7 Columns (one for each weekday) were defined at which each column accepts either 0 or 1 as value
Weekend	Binary	1 = Weekend (Friday or Saturday) 0 = All other days
Peak	Binary	4 Columns (AM, MD, PM, and off-Peak) were defined at which each column accepts either 0 or 1 as value

Parameter	Data Encoding Type	Encoding Description
Starting Time	Numerical	The timing bus started the trip, rounded to nearest 30mins for ease of analysis and interpretation
Through	Numerical	Number of Movements per trip/route
Left	Numerical	Number of Movements per trip/route
Right	Numerical	Number of Movements per trip/route
U-Turn	Numerical	Number of Movements per trip/route
CBD	Binary	3 Columns (Inner-CBD, Outer-CBD, and Non-CBD) were defined at which each column accepts either 0 or 1 as value
Road Class	Binary	1= Major roads (arterials, expressways...etc.) 0=Local Roads
Land use	Binary	4 Columns (Residential, Commercial, Industrial, and Mix) were defined at which each column accepts either 0 or 1 as value
Stops	Numerical	Number of Stops per Route/Trip
Signals	Numerical	Number of Traffic Signals per Route/Trip
Roundabouts	Numerical	Number of Roundabouts per Route/Trip
Travel Time	Numerical	Trip duration in minutes rounded to 1 min

4.5.3. OUTLIER DETECTION AND REMOVAL

The second step of pre-processing data stage is to eliminate outlier records. Using the above-mentioned parameters, outlier records were described in one of the following situations:

1. **Long or short travel time:** This case was noticed to appear in the data set due to software or hardware error. For example, the travel time of 7392187 hours or less than 2 minutes' travel time.
2. **High-speed records:** Those records reflected travel speed higher than 100 km/h; which implies that those records are not accurate
3. **Distances with more than +/- 10% of actual route length:** This error could be a result of many factors that cannot be considered in the prediction model and could jeopardize the accuracy of the prediction model.

After elimination of the above outliers, more than 137,000 records were available for model development stage. However, very high number of repeated records were noticed in the data; which could lead to confusion and inaccuracy in model development stages. Thus, the average travel time of trip was calculated for each repetitive case where the remaining 16 parameters are identical. This process leads to a reduction in data from 137,000 records to approximately 78,004 unique records.

5. DATA COLLECTION

The data were collected from Mowasalat. The public transit operator can be categorized into two ways:

1. Travel time data for each route per trip as discussed in section 4.1
2. Public transit routes map, shown in Figure 4.3

Based on section 4.3 classification the section 5.1 of this chapter describes route related parameters. While trip related parameters are discussed in section 5.2 of this chapter.

5.1. ROUTE RELATED DATA

5.1.1. ROUTE 10 – (AL-GHANIM BUS STATION - AL THUMAMA)

Route 10 shown in Figure 5.1 is approximately 13.4km on the direction going out of the central bus station Al Ghanim Bus Station toward Al Thumama Bus Stop No. 5 and passes through twenty-four bus stops in total. This route passes mainly highly populated residential areas such as Al Hilal and Al Mansour; which are an outer-CBD area. Thus, to serve as much as possible passengers, buses on both its in-going and outgoing routes used local roads. Table 5.1 summarizes the primary characteristics of this route.

From Main Station	To Main Station
	

Figure 5.1: Google Earth Image of Route 10

Table 5.1: Main Characteristics of Route 10

Category	Parameter	Value	
		Out	In
Geometrical Characteristics	Length (km)	13.4	11
	No. of Right-turns at junctions	8	7
	No. of Left-turns at junctions	6	5
	No. of through movements at junctions	7	5
	No. of U-Turns	0	0
	No. of Roundabouts	6	3
	No. of Traffic Signals	12	10
Stops	No. of Bus Stops in total	24	20
Surrounding Area and Land-use Characteristic	Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)	Residential	
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Outer-CBD	
Road Type	Category of Roads bus went through (Major Roads or Local Roads)	Local Roads	

5.1.2. ROUTE 11 – (AL-GHANIM BUS STATION – AL THUMAMA AREA)

Route 11 shown in Figure 5.2 is approximately 28.9 km the circular route that starts from Al-Ghanim Bus Station to the same point and passes through 47 bus stops in total. This route passes mainly highly congested commercial areas such as Al Mansoura, Mall, Lulu Hypermarket and old airport street; which are outer-CBD areas. Thus, to serve as much as possible passengers, buses on the circulatory routes used local roads. Table 5.2 summarizes the primary characteristics of this route.

Circular Route

Figure 5.2: Google Earth Image of Route 11

Table 5.2: Main Characteristics of Route 11

Category	Parameter	Value Circular
Geometrical Characteristics	**Length (km)**	28.9
	No. of Right-turns at junctions	23
	No. of Left-turns at junctions	19
	No. of through movements at junctions	20
	No. of U-Turns	0
	No. of Roundabouts	10
	No. of Traffic Signals	28
Stops	**No. of Bus Stops in total**	47
Surrounding Area and Land-use Characteristic	**Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)**	Commercial
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD
Road Type	**Category of Roads bus went through (Major Roads or Local Roads)**	Local Roads

5.1.3. ROUTE 21 – (MOWASLAT CITY MAIN BUS STOP - AL-GHANIM BUS STATION)

Route 21 shown in Figure 5.3 have two types, the first one is about 42.2 km circular route that starts from Mowaslat City main bus stop to the same point and passes through 54 bus stops in total. While the second type is a partial or unidirectional that supports the service of the main line which does not have a specific length. This route passes mainly highly congested commercial areas such as wholesale market, Al Maamoura, and Al Muntazah; which are outer-CBD areas. Thus, to serve as much as possible passengers, buses on the circulatory routes used major roads. Table 5.3 summarizes the primary characteristics of this route.

Circular Route

Figure 5.3: Google Earth Image of Route 21

Table 5.3: Main Characteristics of Route 21

Category		Parameter	Value	
			Circ.*	**Part. ***
Geometrical Characteristics		**Length (km)**	42.2	NA**
		No. of Right-turns at junctions	17	8
		No. of Left-turns at junctions	16	10
		No. of through movements at junctions	24	9
		No. of U-Turns	0	0
		No. of Roundabouts	16	8
		No. of Traffic Signals	30	14
Stops		**No. of Bus Stops in total**	54	29
Surrounding Area and Land-use	**Characteristics**	**Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)**	Commercial	
		Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Outer-CBD	
Road Type		**Category of Roads bus went through (Major Roads or Local Roads)**	Major Roads	

* Circular full route starts from Mowasalat City main bus stop and goes back to it. However, partial route starts from Al Ghanim Bus Station toward Mowasalat City main bus stop.
** Exact Data is not available.

5.1.4. ROUTE 33A – (AL-GHANIM BUS STATION - INDUSTRIAL AREA STREET NO.1)

Route 33A shown in Figure 5.4 is approximately 18.4 km on the direction going out of the central bus station Al Ghanim Bus Station toward Industrial Area Street No.1 Bus Stop No. 1800 and passes through 20 bus stops in total. Route 33A is 20.3 km long on its way toward Al Ghanim Bus Station from Al Thumama Bus Stop No.

1800 and passes through 17 bus stops in total. This route passes mainly highly congested commercial areas such as Industrial Area and Salwa Road; which are non-CBD areas. Thus, to serve as much as possible passengers, buses on both its in-going and outgoing routes used main roads. Table 5.4 summarizes the primary characteristics of this route.

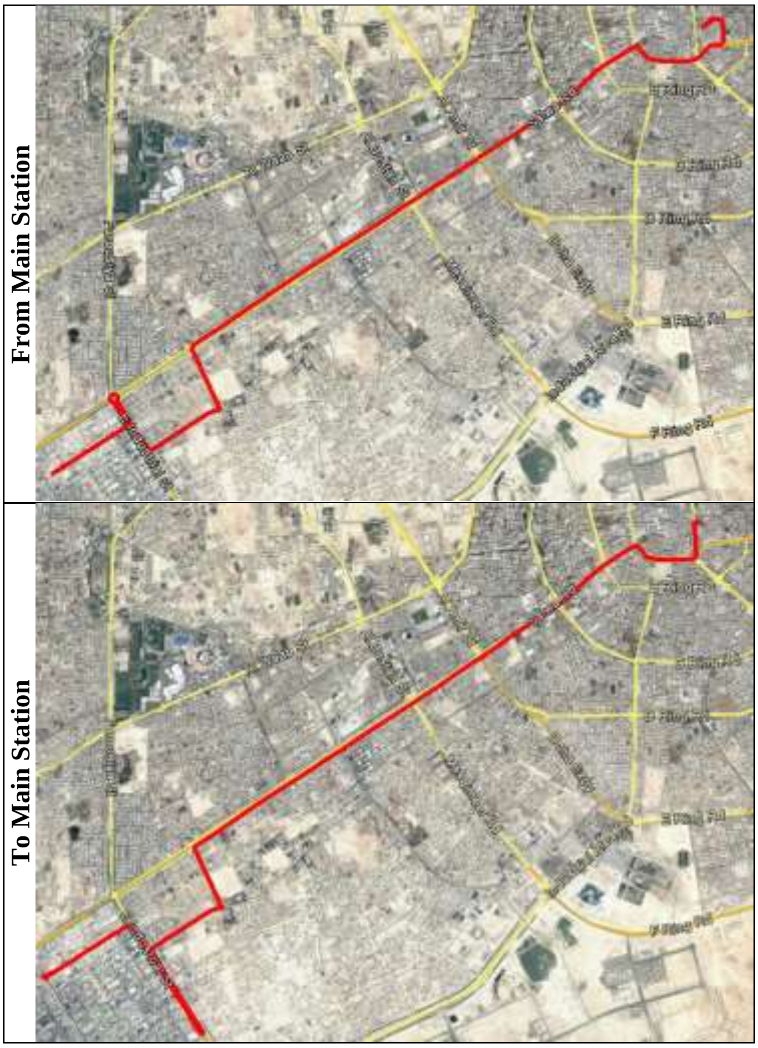


Figure 5.4: Google Earth Image of Route 33A

Table 5.4: Main Characteristics of Route 33A

Category	Parameter	Value Out	Value In
Geometrical Characteristics	Length (km)	18.4	20.3
	No. of Right-turns at junctions	5	5
	No. of Left-turns at junctions	2	2
	No. of through movements at junctions	10	11
	No. of U-Turns	1	2
	No. of Roundabouts	7	5
	No. of Traffic Signals	9	12
Stops	No. of Bus Stops in total	20	17
Surrounding Area and Land-use Characteristic	Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)	Commercial	
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD	
Road Type	Category of Roads bus went through (Major Roads or Local Roads)	Major Roads	

5.1.5. ROUTE 34 – (AL-GHANIM BUS STATION – AL AZIZIA – MUAITHER)

Route 34 shown in Figure 5.5 have two types, the first one is about 38.2 km the circular route that starts from Al-Ghanim Bus Station to the same point and passes through 46 bus stops in total. This route passes mainly highly congested commercial areas such as Al Shafi Street, Muaither Street, and Al Azizia Street; which are non-CBD areas. Thus, to serve as much as possible passengers, buses on the circulatory routes used major roads. Table 5.5 summarizes the primary characteristics of this route.

Circular Route

Figure 5.5: Google Earth Image of Route 34

Table 5.5: Main Characteristics of Route 34

Category	Parameter	Value (Circular)
Geometrical Characteristics	**Length (km)**	28.9
	No. of Right-turns at junctions	23
	No. of Left-turns at junctions	19
	No. of through movements at junctions	20
	No. of U-Turns	0
	No. of Roundabouts	10
	No. of Traffic Signals	28
Stops	**No. of Bus Stops in total**	47
Surrounding Area and Land-use Characteristic	**Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)**	Commercial
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD
Road Type	**Category of Roads bus went through (Major Roads or Local Roads)**	Major Roads

5.1.6. ROUTE 41 – (AL-GHANIM BUS STATION – AL MANASEER)

Route 41 shown in Figure 5.5 is approximately 47.2 km the circular route that starts from Al-Ghanim Bus Station to the same point and passes through 61 bus stops in total. This route passes mainly highly populated residential areas such as Al Rayyan Al Jadeed, Muaither, and Al Manaseer; which are non-CBD areas. Thus, to serve as much as possible passengers, buses on the circulatory routes used local roads. Table 5.6 summarizes the primary characteristics of this route.

Circular Route

Figure 5.6: Google Earth Image of Route 41

Table 5.6: Main Characteristics of Route 41

Category	Parameter	Value (Circular)
Geometrical Characteristics	Length (km)	47.2
	No. of Right-turns at junctions	26
	No. of Left-turns at junctions	19
	No. of through movements at junctions	28
	No. of U-Turns	0
	No. of Roundabouts	22
	No. of Traffic Signals	61
Stops	No. of Bus Stops in total	47
Surrounding Area and Land-use Characteristic	Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)	Residential
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD
Road Type	Category of Roads bus went through (Major Roads or Local Roads)	Local Roads

5.1.7. ROUTE 42 – (AL-GHANIM BUS STATION - AL RAYYAN AL JADEED)

Route 42 shown in Figure 5.7 is approximately 16.5 km on the direction going out of the central bus station Al Ghanim Bus Station toward Al Rayyan Al Jadeed Bus Stop No. 5371 and passes through 16 bus stops in total. Route 42 is 15 km long on its way toward Al Ghanim Bus Station from Al Thumama Bus Stop No. 5371 and passes through 14 bus stops in total. This route passes mainly mixed a variety of land uses areas; which are outer-CBD areas. Thus, to serve as much as possible passengers, buses on both its in-going and outgoing routes used major roads. Table 5.7 summarizes the primary characteristics of this route.

Figure 5.7: Google Earth Image of Route 42

Table 5.7: Main Characteristics of Route 42

Category	Parameter	Value	
		Out	**In**
Geometrical Characteristics	**Length (km)**	16.5	15
	No. of Right-turns at junctions	8	5
	No. of Left-turns at junctions	4	4
	No. of through movements at junctions	13	11
	No. of U-Turns	1	0
	No. of Roundabouts	11	9
	No. of Traffic Signals	9	9
Stops	**No. of Bus Stops in total**	16	14
Surrounding Area and Land-use Characteristic	**Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)**	Mixed	
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Outer-CBD	
Road Type	**Category of Roads bus went through (Major Roads or Local Roads)**	Major Roads	

5.1.8. ROUTE 43 – (AL-GHANIM BUS STATION – AL MANASEER)

Route 43 shown in Figure 5.8 is approximately 40.2 km the circular route that starts from Al-Ghanim Bus Station to the same point and passes through 43 bus stops in total. This route passes mainly highly congested commercial areas such as Al Rayyan, Muaither, and Al Azizia; which are non-CBD areas. Thus, to serve as much as possible passengers, buses on the circulatory routes used local roads. Table 5.8 summarizes the primary characteristics of this route.

Circular Route

Figure 5.8: Google Earth Image of Route 43

Table 5.8: Main Characteristics of Route 43

Category	Parameter	Value Circular
Geometrical Characteristics	**Length (km)**	40.2
	No. of Right-turns at junctions	16
	No. of Left-turns at junctions	9
	No. of through movements at junctions	24
	No. of U-Turns	3
	No. of Roundabouts	24
	No. of Traffic Signals	20
Stops	**No. of Bus Stops in total**	43
Surrounding Area and Land-use Characteristic	**Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)**	Mixed
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD
Road Type	**Category of Roads bus went through (Major Roads or Local Roads)**	Major Roads

5.1.9. ROUTE 76 – (AL-GHANIM BUS STATION – CITY CENTER DOHA)

Route 76 shown in Figure 5.9 is approximately 22 km the circular route that starts from Al-Ghanim Bus Station to the same point and passes through 29 bus stops in total. This route passes mainly through Al Corniche; which is an inner-CBD area. Thus, to serve as much as possible passengers, buses on the circulatory routes used major roads. Table 5.9 summarizes the primary characteristics of this route.

Circular Route

Figure 5.9: Google Earth Image of Route 76

Table 5.9: Main Characteristics of Route 76

Category	Parameter	Value Circular
Geometrical Characteristics	Length (km)	22
	No. of Right-turns at junctions	10
	No. of Left-turns at junctions	5
	No. of through movements at junctions	13
	No. of U-Turns	0
	No. of Roundabouts	3
	No. of Traffic Signals	23
Stops	No. of Bus Stops in total	29
Surrounding Area and Land-use Characteristic	Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)	Commercial
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Inner-CBD
Road Type	Category of Roads bus went through (Major Roads or Local Roads)	Major Roads

5.1.10. ROUTE 301 – (MOWASLAT CITY MAIN BUS STOP- AL-GHANIM BUS STATION)

Route 301 shown in Figure 5.10 have two types, the first one is about 52 km circular route that starts from Mowaslat City main bus stop to the same point and passes through 46 bus stops in total. This route passes mainly highly congested areas of mixed land-uses such as Mesaimeer, Al Thumama, Al Mansoura, Al Sadd, Al

Waab, Muaither, Al Manaseer and Industrial Area; which are non-CBD areas. Thus, to serve as much as possible passengers, buses on the circulatory routes used major roads. Table 5.10 summarizes the major characteristics of this route.

Circular Route

Figure 5.10: Google Earth Image of Route 301

Table 5.10: Main Characteristics of Route 301

Category		Parameter	Value
			Circular
Geometrical Characteristics		Length (km)	52
		No. of Right-turns at junctions	13
		No. of Left-turns at junctions	17
		No. of through movements at junctions	26
		No. of U-Turns	2
		No. of Roundabouts	14
		No. of Traffic Signals	30
Stops		No. of Bus Stops in total	46
Surrounding Area and Land-use Characteristic		Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)	Mixed
		Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD
Road Type		Category of Roads bus went through (Major Roads or Local Roads)	Major Roads

5.1.11. ROUTE 302 – (MOWASLAT CITY MAIN BUS STOP- AL-GHANIM BUS STATION)

Route 302 shown in Figure 5.10 have two types, the first one is about 52 km circular route that starts from Mowaslat City main bus stop to the same point and passes through 45 bus stops in total. This route passes mainly highly congested areas of mixed land-uses such as Mesaimeer, Abu Hamour and Industrial Area; which are non-CBD areas. Thus, to serve as much as possible passengers, buses on the circulatory routes used major roads. Table 5.10 summarizes the major characteristics of this route.

Circular Route

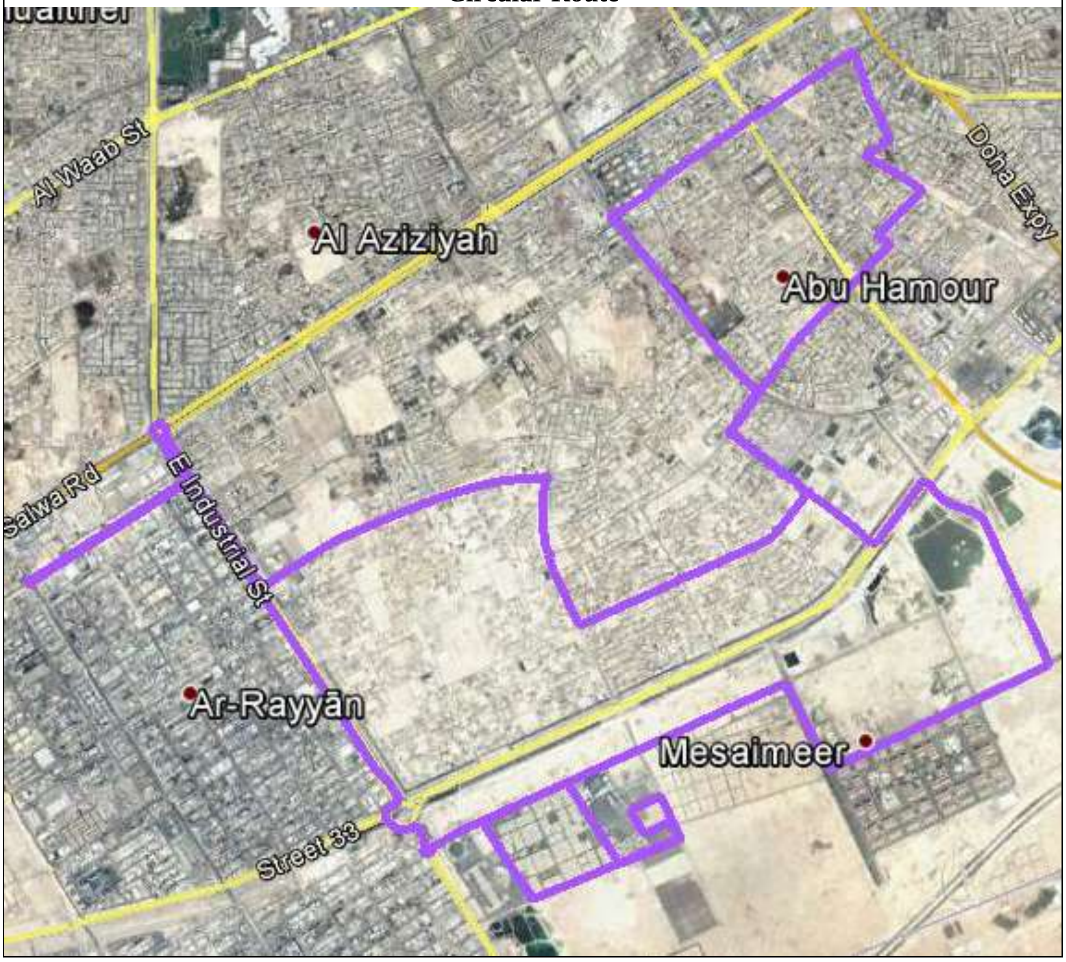


Figure 5.11: Google Earth Image of Route 302

Table 5.11: Main Characteristics of Route 302

Category	Parameter	Value Circular
Geometrical Characteristics	Length (km)	47.2
	No. of Right-turns at junctions	15
	No. of Left-turns at junctions	13
	No. of through movements at junctions	25
	No. of U-Turns	2
	No. of Roundabouts	26
	No. of Traffic Signals	16
Stops	No. of Bus Stops in total	45
Surrounding Area and Land-use Characteristic	Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)	Mixed
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD
Road Type	Category of Roads bus went through (Major Roads or Local Roads)	Major Roads

5.1.12. ROUTE 303 – (MOWASLAT CITY MAIN BUS STOP- AL-GHANIM BUS STATION)

Route 303 shown in Figure 5.12 have two types, the first one is about 44.6 km circular route that starts from Mowaslat City main bus stop to the same point and passes through 42 bus stops in total. This route passes mainly areas of residential land-uses such as Mesaimeer, Abu Hamour and Industrial Area; which are non-CBD areas. Thus, in order to serve as much as possible passengers, buses on the circulatory routes used local roads. Table 5.12 summarizes the major characteristics of this route.

Circular Route

Figure 5.12: Google Earth Image of Route 303

Table 5.12: Main Characteristics of Route 303

Category	Parameter	Value Circular
Geometrical Characteristics	Length (km)	44.6
	No. of Right-turns at junctions	17
	No. of Left-turns at junctions	13
	No. of through movements at junctions	19
	No. of U-Turns	3
	No. of Roundabouts	25
	No. of Traffic Signals	8
Stops	No. of Bus Stops in total	42
Surrounding Area and Land-use Characteristic	Land-use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)	Residential
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Non-CBD
Road Type	Category of Roads bus went through (Major Roads or Local Roads)	Local Roads

5.1.13. SUMMARY OF ROUTES CHARACTERISTICS

Below table summarizes manually extracted data for each route. These inputs were the basis at which the model was developed.

Table 5.13:Summary of Routes Main Characteristics

Route No.		10		11	21		33A		34	41	42		43	76	301	303	302
Category	**Parameter**	**Out**	**In**	**Circ.***	**Circ.***	**Part. ***	**Out**	**In**	**Circular**	**Circular**	**Out**	**In**	**Circular**	**Circular**	**Circular**	**Circular**	**Circular**
Geometrical Characteristics	**Length (km)**	13.4	11	28.9	42.2	NA**	18.4	20.3	38.2	47.2	16.5	15	40.2	22	52	44.6	47.2
	No. of Right-turns at junctions	8	7	23	17	8	5	5	13	26	8	5	16	10	13	17	15
	No. of Left-turns at junctions	6	5	19	16	10	2	2	7	19	4	4	9	5	17	13	13
	No. of through movements at junctions	7	5	20	24	9	10	11	31	28	13	11	24	13	26	19	25
	No. of U-Turns	0	0	0	0	0	1	2	0	0	1	0	3	0	2	3	2
	No. of Roundabouts	6	3	10	16	8	7	5	22	22	11	9	24	3	14	25	26
	No. of Traffic Signals	12	10	28	30	14	9	12	23	26	9	9	20	23	30	8	16

***Continued**

Route No.		10		11	21		33A		34	41	42		43	76	301	303	302
Category	**Parameter**	**Out**	**In**	**Circ.***	**Circ.***	**Part. ***	**Out**	**In**	**Circular**	**Circular**	**Out**	**In**	**Circular**	**Circular**	**Circular**	**Circular**	**Circular**
Stops	**No. of Bus Stops in total**	24	20	47	54	29	20	17	46	61	16	14	43	29	46	42	45
Surrounding Area and Land Use Characteristics	**Land use category of areas surrounding majority of bus route (Commercial, Industrial, Residential or Mixed)**	Residential		Commercial	Commercial		Commercial		Commercial	Residential	Mixed		Commercial	Commercial	Mixed	Residential	Mixed
	Bus route location in reference to Central Business District (CBD) (Inner CBD, Outer-CBD or non-CBD)	Outer-CBD		Outer-CBD	Outer-CBD		Non-CBD		Non-CBD	Non-CBD	Outer-CBD		Non-CBD	Inner-CBD	Non-CBD	Non-CBD	Non-CBD
Road Type	**Category of Roads bus went through (Major Roads or Local Roads)**	Local Roads		Local Roads	Major Roads		Major Roads		Major Roads	Local Roads	Major Roads		Major Roads	Major Roads	Major Roads	Local Roads	Major Roads

Nine out of twelve routes operated in a circular pattern, where the bus starts from the central station of Al Ghanim and ends to the same station. The other three routes operated in the unidirectional pattern. Which means that the bus starts from a station and ends in another station.

Interestingly it was noticed that 7 out of 12 of the routes which data was collected for falls in non-CBD regions. While only 1 falls within inner-CBD and 4 within outer-CBD.The below table describes the numerical data values of the collected data statistically.

Table 5.14 Numerical Data Descriptive Statistics

Statistical Measure	**Distance (km)**	**Stops**	***Through***	***L-Turn***	***R-Turn***	***U-Turn***	***Signals***	***Roundabouts***
Mean	31	39	18.9	12.02	14.05	0.78	22.25	12
Median	29	46	20	16	13	0	26	10
Mode	11	46	26	19	13	0	30	14
Standard Deviation	15.69	12.9	7.76	6.32	6.28	0.94	7.74	7.04
Range	51	44	26	17	21	3	21	23
Minimum	9	17	5	2	5	0	9	3
Maximum	60	61	31	19	26	3	30	26

The lengths of routes used to build the model ranged from 9km to 60km with an average of 31km. Although the mode resulted in 11km routes, it can be explained that Mowasalat discharges more often short duration buses than long duration buses.

5.2. TRIP-RELATED DATA

Travel time data for each route per trip was received from Mowasalat, as described in section 4.1. The number of records used in model development is 78,004 unique record that combines total of nine routes. Table 5.14 statistically describes the travel time data used in developing the model.

Table 5.15 Statistical Description Summary

Statistical Measure	Travel time (min)
Mean	88
Median	89
Mode	53
Standard Deviation	34.26
Range	162
Minimum	26
Maximum	188

Travel time ranged from 26 to 188 minutes with an average of 88 minutes. The average travel time of buses seems to be relatively long duration with 88 minutes. This can be explained by the operational strategy of circular routes that the service provider operates in.

Being interested in understanding the travel time behavior, peak to peak comparison per route had been conducted. The below chart shows the average travel time per peak for each route.

As an overall trend, it can be seen from the below figure that the average duration of the trips is the highest in all routes in the PM time. There is also a clear trend when it comes to the comparison between AM and MD, as the AM appears to be higher in the average trip duration. Such patterns might be translated since the majority of the working class in Qatar work in the private sector, which usually work from 8 to 5. This causes the MD to be slightly less than the PM and AM in all routes.

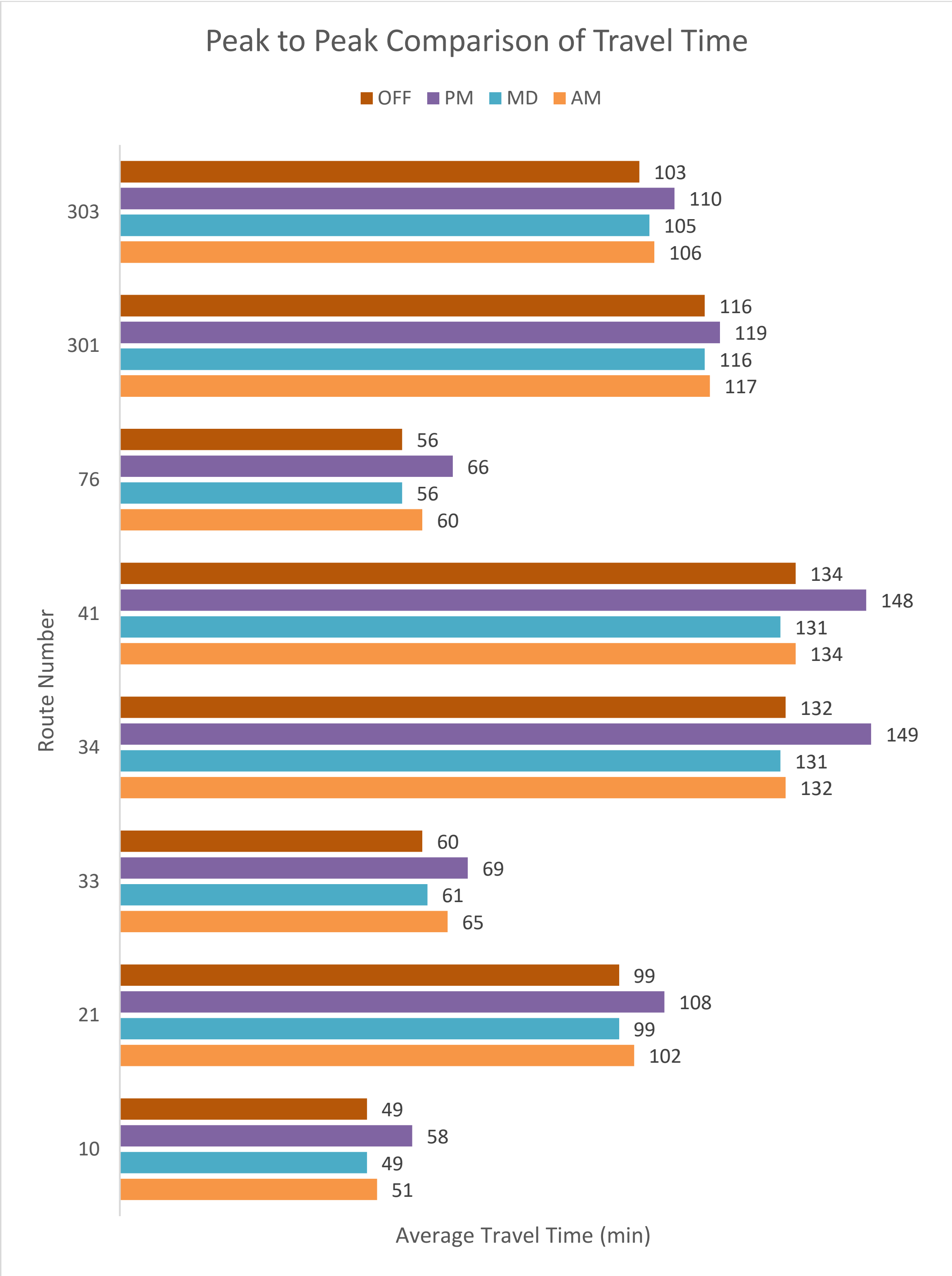


Figure 5.13 Peak to Peak Comparison of Travel Time

Moreover, correlation analysis to determine the significance of each parameter impact on other parameters was conducted. In the below, table correlation analysis of numerical parameters is shown. The gray highlighted correlation coefficients show a relatively strong correlation between parameters.

Table 5.16 Correlation Analysis

	Distance (km)	Start Time	Through Movements	L-Turn Movements	R-Turn Movements	U-Turn Movements	Traffic Signals	Roundabouts	Stops	Duration (min)
Distance (km)	1.00									
Start Time	-.03	1.00								
Through Movements	0.82	-.02	1.00							
L-Turn Movements	0.53	-.03	0.75	1.00						
R-Turn Movements	0.30	-.02	0.63	0.86	1.00					
U-Turn Movements	0.67	-.02	0.42	0.13	-.14	1.00				
Traffic Signals	0.47	-.02	0.75	0.82	0.68	0.04	1.00			
Roundabouts	0.78	-.02	0.83	0.51	0.48	0.48	0.31	1.00		
Stops	0.65	-.02	0.90	0.90	0.85	0.13	0.82	0.74	1.00	
Duration (min)	0.88	-.01	0.79	0.54	0.43	0.43	0.51	0.71	0.69	1.00

Multiple strong correlations that range between 0.90 and 0.74 appeared between numbers of through, left and right turns movements as well as number of signals and roundabouts in relation with a number of stops on the route. This can be explained by the fact that the longer the route, the higher are the numbers of stops and movements. Moreover, this explanation can apply to many of the above-highlighted correlations. Although it does not add significant value either useful information; however, it emphasizes on the accuracy and creditability of the analyzed data.

An interesting correlation can be seen in duration or travel time with the number of roundabouts. Although this correlation could be explained by the same reasoning mentioned earlier; however, the correlation between the number of traffic signals and travel time is much lower significant. This may imply that roundabouts generally lead to higher travel time. Nevertheless, this reasoning shall be supported by further investigation and analysis.

6. ANN MODEL DEVELOPMENT AND RESULTS ANALYSIS

Using methodology and parameters defined in chapter 4 of this report, more than 78,004 data points, collected from nine routes over a year. The prediction model was trained to forecast travel time of bus routes using MATLAB Neural Network Toolbox.

The total of 78,004 data points is used in ANN model development stage. To ensure the validity of the ANN training process, three independent groups of data points were created, and the 78,004 data points were randomly assigned to one of the three groups. These three groups are:

- **Training data points:** 54,602 data points, representing 70% of overall data points dedicated to ANN model development, used to train the ANN public transport travel time prediction model
- **Validation data points:** 11,701 records, representing 15% of overall data points dedicated for ANN model development, used to monitor the performance of the ANN during the training process and to keep the network from over-fitting; which results in a more generalized model that can predict independent inputs that were not included in the model.
- **Testing data points:** 11,701 records, representing 15% of overall data points dedicated to ANN model development, used to validate the capability of the network to predict the targeted results.

6.1. ANN ARCHITECTURE

Several network architectures were experimented with to achieve the model which resulted in the best prediction accuracy. A description of the different network architecture is tabulated in Table 6.1.

Table 6.1 ANN Architecture Tested

Model ID	Number of Hidden Layers	Transfer Function	Training Algorithm	Learning Algorithm	Number of Neurons	Performance Function
1	1	Hyperbolic Tangent then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	5	MSE
2	1	Hyperbolic Tangent then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	10	MSE
3	1	Logistic then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	5	MSE
4	1	Logistic then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	10	MSE
5	2	Hyperbolic Tangent then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	5	MSE
6	2	Hyperbolic Tangent then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	10	MSE
7	2	Logistic then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	5	MSE
8	2	Logistic then Linear	Levenberg–Marquardt	Gradient Descent with Momentum	10	MSE

Levenberg–Marquardt algorithm and Gradient Descent with Momentum were selected as training and learning algorithms, respectively. Levenberg–Marquardt algorithm coverages faster than any other algorithm appropriate for curve fitting purpose, without higher computational cost. While Gradient Descent with Momentum was selected as the learning algorithm since it accelerates the learning process using momentum coefficient. Error in learning and training process is measured using Mean Square Error (MSE), donated by:

$$MSE = \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{N}, n = 1,2,3 \ldots . N$$ **Equation 6-1**

Where;

N is number of data points

y_n is the historical recorded travel time for n^{th} data point

$\hat{y}_n$ is the predicted travel time by the model for n^{th} data point

In addition to the above ANN architecture the following training parameters were used:

Table 6.2 ANN Training Parameters

Parameter Description	Value
Maximum Number of Epochs The maximum allowed number of iterations at which the training process will be stopped after	1000
Performance Goal The targeted MSE error that ANN model is driven to	0
Maximum Validation Failures The maximum number of failed iterations that at which the model will stop the training process after	6
Minimum Performance Gradient The slope of performance curve is targeted; which means that the model learning curve had reduced significantly and further training will not improve the model significantly	1E-5
Initial Learning Momentum Factor used in Gradient Descent with Momentum method to expedite the learning process	0.001
Learning Momentum Decrease Factor	0.1
Learning Momentum Increase Factor	10
Maximum Learning Momentum	1E10
Maximum Time to Train in (Seconds)	Infinity

6.2. ANN ARCHITECTURE EVALUATION

MATLAB Artificial Neural Networks Toolbox is used to train, test and validated the model. Total of eight ANN architectures described in previous section were tested. The aim of this exercise is to identify the optimum ANN architecture to predict public transport buses travel time. MSE, MAE, coefficient of determination, correlation factor and SE are tabulated in Table 6.3 for each ANN architecture.

Table 6.3 ANN Architecture Evaluation Results

Model ID	Number of Iterations	R	R^2	MSE (min)	MAE (min)	SE (min)
1	42	0.9732	0.947	62.07	6.3	0.13
2	117	0.9762	0.953	57.05	5.94	0.13
3	86	0.9726	0.946	63.37	6.37	0.13
4	138	0.9737	0.948	60.84	6.23	0.13
5	96	0.9732	0.947	62.2	6.31	0.13
6	94	0.9736	0.948	61.26	6.25	0.13
7	69	0.9639	0.929	83.17	7.18	0.13
8	172	0.9746	0.950	59.00	6.16	0.13

All tested ANN architectures showed promising prediction results with minimal error. Among the eight tested ANN architecture, model number 2 had the highest determination factor and lowest error per MOE shown above. This model architecture consists of one input layer, 10 neurons in one hidden layer and one output layer was used in prediction model development through MATLAB Neural Network Toolbox.

The training process was relatively short and converging; due to the use of Levenberg–Marquardt (LM) algorithm. Figure 6.1 shows the performance curve of the model training, which shows convergence and no errors. Moreover, it can be noticed that goal was met within 117 epochs (iterations).

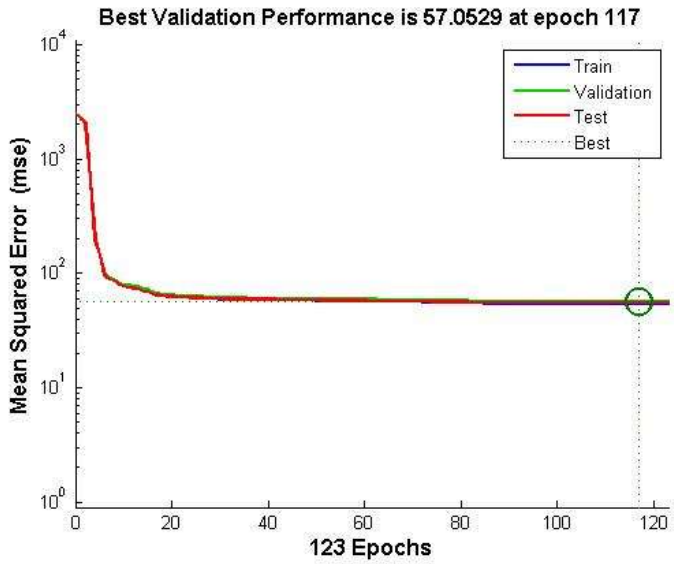


Figure 6.1: Performance Curve of ANN Prediction Model

Training, validation, and testing are the stages at which the ANN goes through to build the prediction model. The following sections of this chapter discusses analyze the results for each of this stage; toward further understanding of the ANN travel time prediction model development stages.

6.3. ANN TRAINING STAGE

In the training stage, prediction results are compared to the targeted results using Levenberg–Marquardt (LM) algorithm. Weights are adjusted then based on the comparison result, and iterations continue until error threshold is met. During this phase of model development, 70% of the data dedicated to ANN prediction model development is used. The regression plot of the training stage is shown in Figure 6.2. ANN showed high correlation, R^2=0.9533, between targeted and predicted travel time by the model.

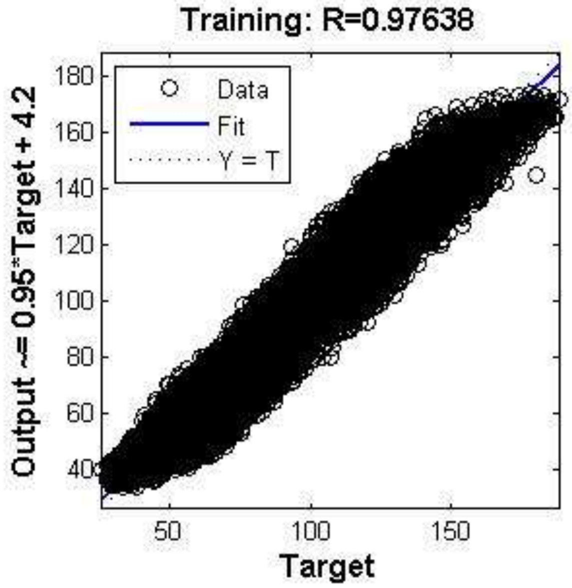


Figure 6.2: Training Stage Regression Plot Generated by Neural Network Toolbox In MATLAB

Using methodology described in Chapter 4; the below measures of effectiveness (MOE) were calculated for prediction model outputs versus targeted historical travel time values during training stage of model development and summarized in Table 6.4

Table 6.4 ANN Training Stage Measures of Effectiveness

MAE (min)	Coefficient of Determination (R^2)	Standard Error (SE) (min)
6.22	0.95	0.14

The performance of ANN prediction model in training stage is auspicious with R^2=0.95 and MAE=6.22 minutes. Further analysis of prediction results was conducted. The analysis shows in that more than 48% of predictions had an error less than 5 minutes. While, 80% of predictions had an error less than 10 minutes per route.

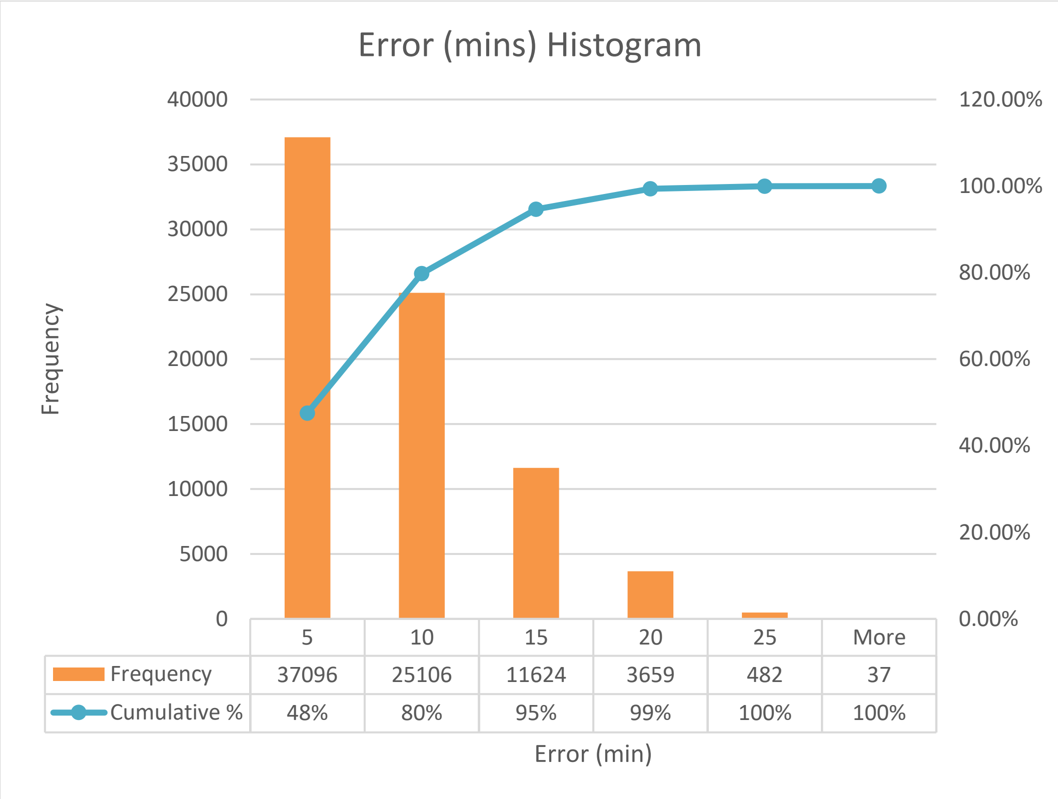
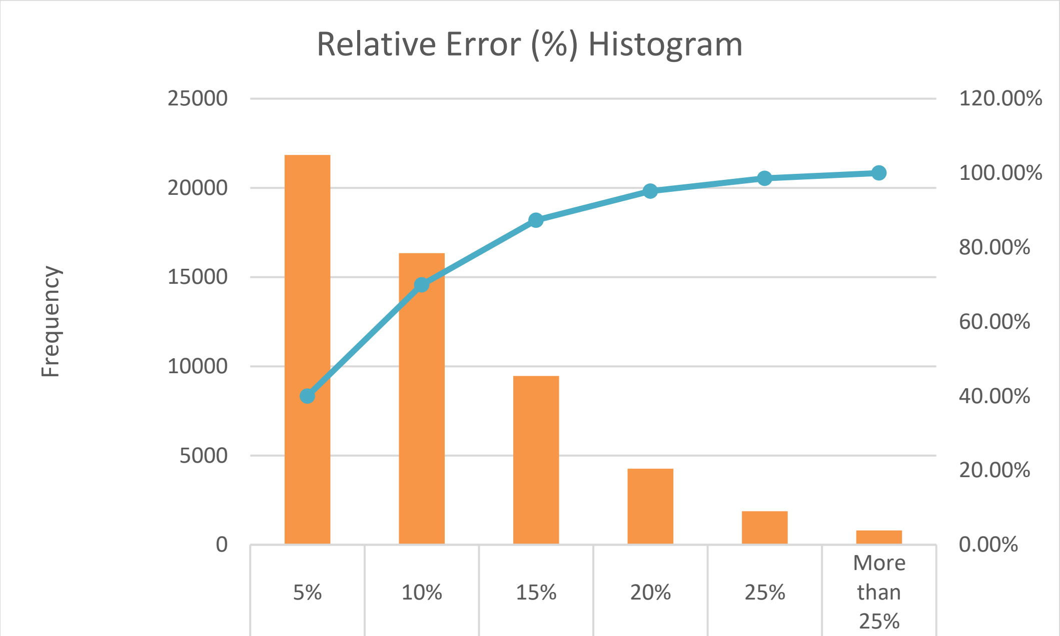


Figure 6.3 ANN Training Stage Prediction Model Error Analysis – in Minutes

Although the error analysis in minutes shows promising results however, 10 minutes of error in prediction raises concerns of model applicability to short trips. Thus, further analysis of training stage predictions error as a percentage of overall trip duration was conducted and illustrated in Figure 6.4.

Analysis of prediction error as a percentage of total trip duration showed that 70% of predictions had an error less than 10% of trip duration. While less than 1% of prediction had error representation more than 25% of total trip duration.



	5%	10%	15%	20%	25%	More than 25%
Frequency	21849	16349	9457	4270	1880	797
Cumulative %	40%	70%	87%	95%	99%	100%

Figure 6.4 Ann Training Stage Prediction Model Error Analysis – Percentage

These results in training stage should be validated and tested using randomly selected data points. Validation and testing stage of model development are detailed in the following section.

6.4. ANN VALIDATION STAGE

Validation stage takes place during ANN prediction model development stage to monitor the performance of the ANN during the training process and to keep the network from over-fitting.

Overfitting which refers to model behavior at which model learns the detail and noise in the training data to the extent that it cannot generalize its learning to data points that were not included in the model training. Moreover, the noise in the training data is picked up and learned as concepts by the model; which negatively impact the model outputs. Therefore, validation stage is essential to assure more generalized model development; that can predict independent inputs that were not included in the model.

For validation stage, 15% of the data points dedicated to model development is used. The total number of random data points used for model validation is 11,701.

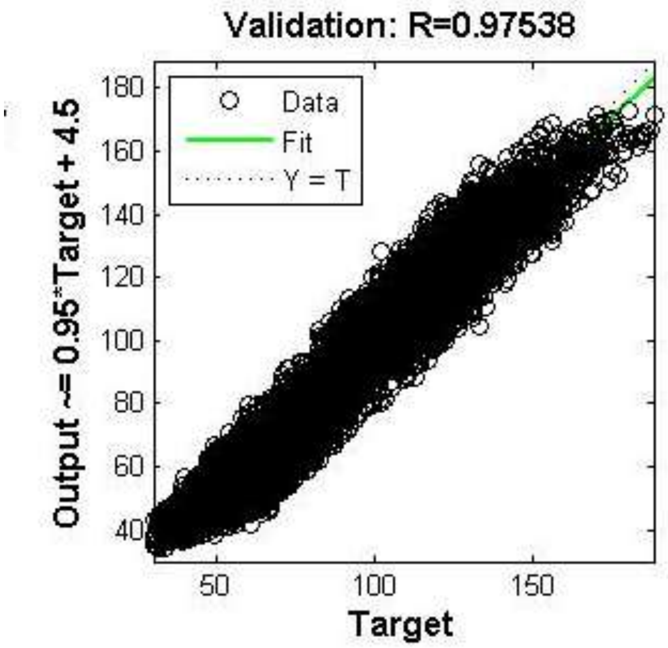


Figure 6.5: Validation Stage Regression Plot Generated by MATLAB

Validation regression shows high correlation with R^2=0.9514. This indicates that trained model is capable of predicting the travel time of public transit buses with high accuracy. Moreover, the results prove that the trained model has not over-fitted to training data points.

ANN validation stage has the MOE tabulated in Table 6.5.

Table 6.5 ANN Validation Stage Measures of Effectiveness

MAE (min)	Coefficient of Determination (R^2)	Standard Error (SE) (min)
6.29	0.95	0.31

The ANN validation stage had MAE of 6.30 minutes and coefficient of determination of 0.95. Analysis of error represented by error histogram is shown in Figure 6.6.

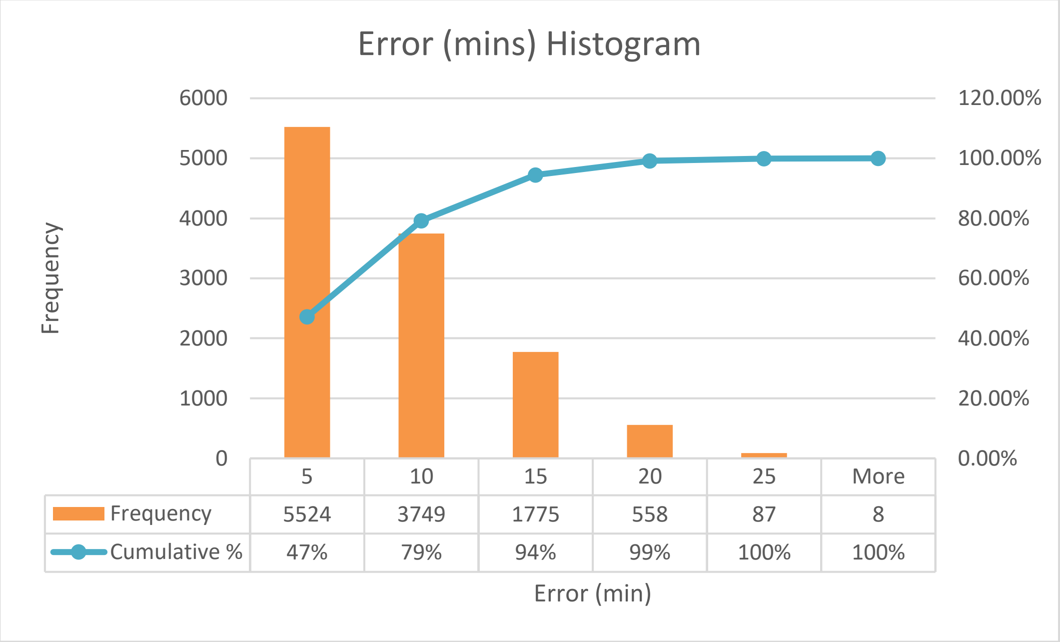


Figure 6.6 ANN Validation Stage Prediction Model Error Analysis – in Minutes

With very close results to training stage error analysis results, less than 6% of data points used in validation stage had to error more than 15 minutes. While, 79% of the data points had error below 10 minutes. In terms of percentage of prediction error in relation to total trip duration, Figure 6.7 summarizes the analysis results. Less than 13% of predicted travel time error is more than 15% of total trip duration, with only 2% of predictions that had error more than 25% of total trip duration.

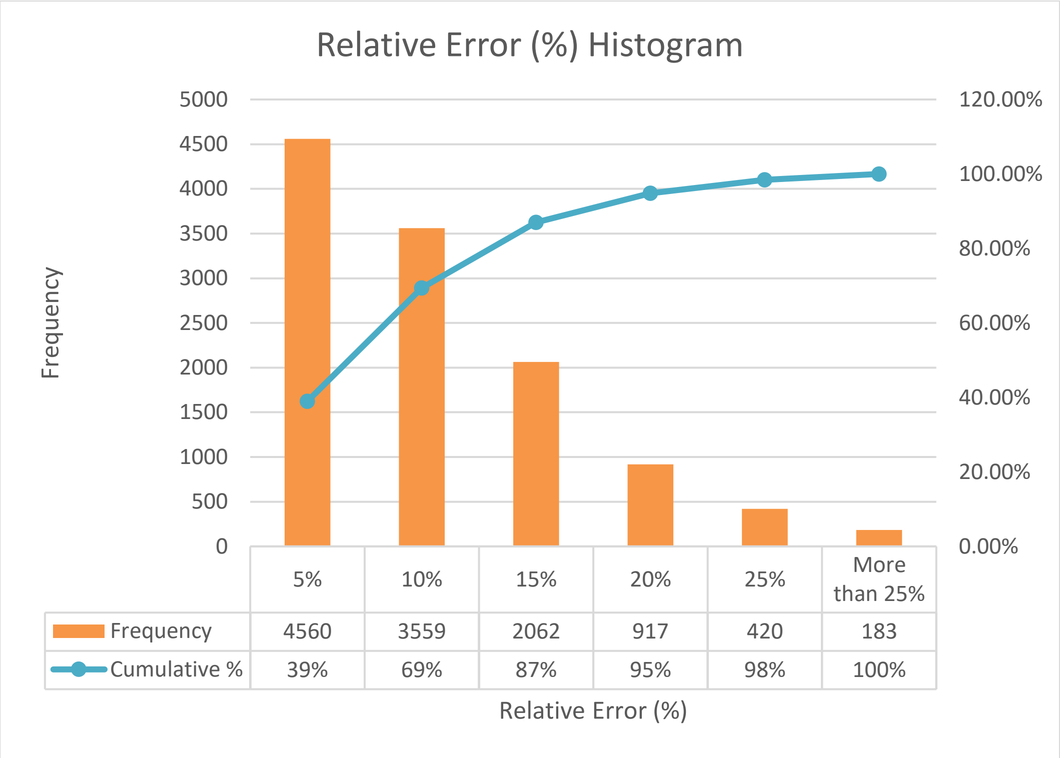


Figure 6.7 Ann Validation Stage Prediction Model Error Analysis – Percentage

6.5. ANN TESTING STAGE

Validation of ANN model capability to predict the targeted results is tested at this stage of model development. Total of 11,701 data points, representing 15% of overall data points dedicated to ANN model development is used at the testing stage of model development.

Figure 6.8 illustrates the high correlation in regression between predicted and targeted travel time results of public transit buses. MOE at testing stage reported in Table 6.6. The coefficient of determination R^2=0.95, MAE=6.23 minutes and SE=0.31 together show that trained and validated ANN prediction model has the capability to predict public transit travel time with high accuracy.

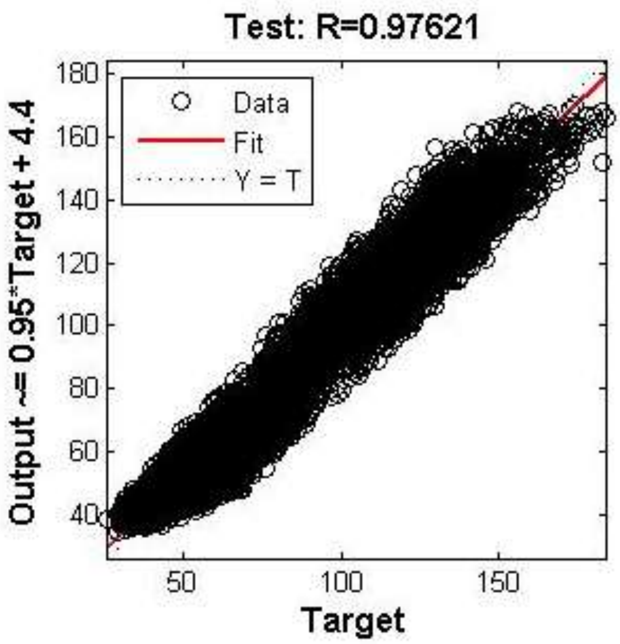


Figure 6.8: Testing Stage Regression Plot Generated by Neural Network Toolbox in MATLAB

Table 6.6 ANN Testing Stage Measures of Effectiveness

MAE (min)	Coefficient of Determination (R^2)	Standard Error (SE) (min)
6.23	0.95	0.31

Further analysis of error was conducted at this stage as well to ensure alignment of all model development stages results. Figure 6.9 shows that less than 5% of travel time predictions have error more than 15 minutes. While the majority, more than 80%, of predictions had an error less than 10 minutes.

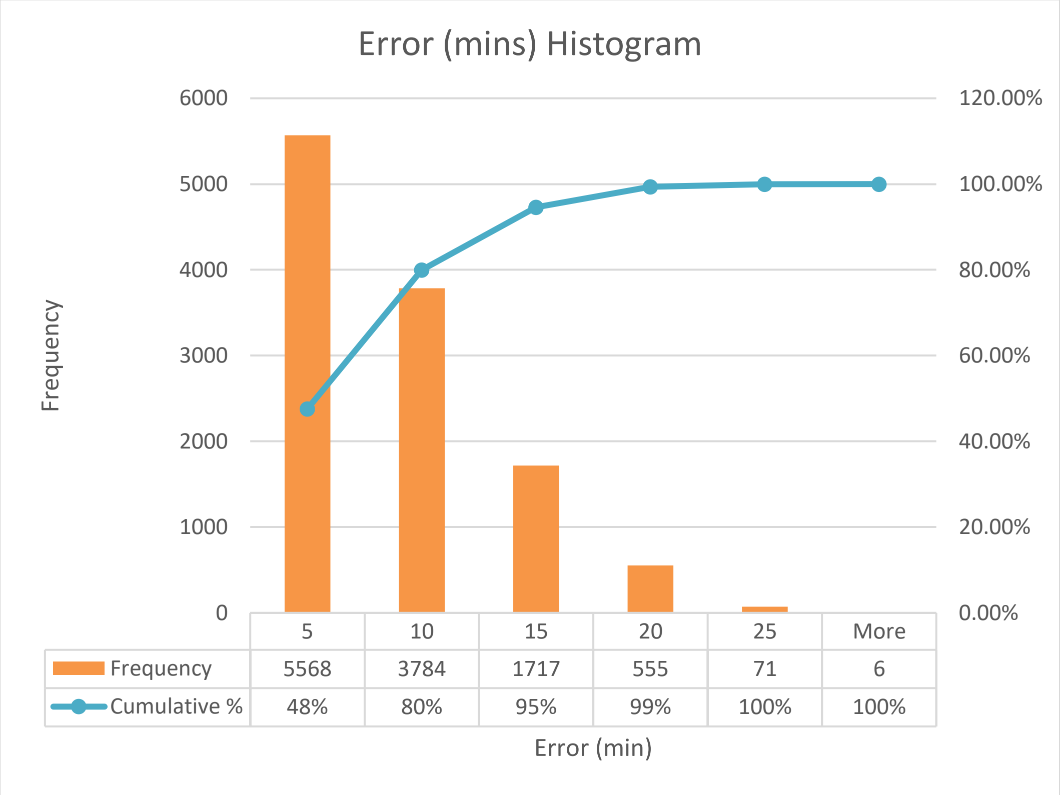
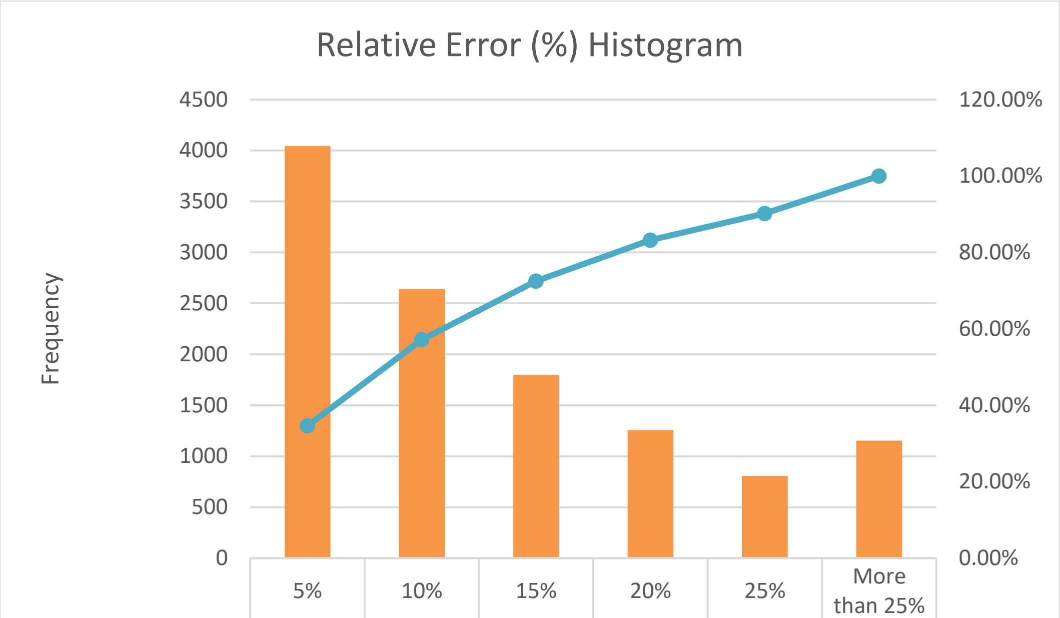


Figure 6.9 ANN Testing Stage Prediction Model Error Analysis – in Minutes

Interestingly, analysis of prediction errors in terms of percentage of total trip time shows lower prediction accuracy, refer to Figure 6.10. The majority, more than 70%, of predictions were associated with an error up to 15%. On the other hand, 10% of predictions had an error of 25% or more compared to targeted travel times. Thus, further analysis to understand these interesting results was conducted. The analysis was conducted by dividing data points into two sets:

a) Data points associated with prediction error more than 25% of total trip time
b) Data points associated with prediction error less than 25% of total trip time

The analysis showed that more than 75% of data points with error more than 25% trip travel time is less than 50 minutes. While 75% of data points with error less than 25% trip travel time is more than 50 minutes. Thus, even if the prediction error is low in terms of minutes, it could reflect a significant percentage of error in prediction. This can explain the significant increase in high error values compared to previous model development stages.



	5%	10%	15%	20%	25%	More than 25%
Frequency	4046	2639	1798	1257	807	1154
Cumulative %	35%	57%	72%	83%	90%	100%

Figure 6.10 Ann Testing Stage Prediction Model Error Analysis – Percentage

6.6. ANN PREDICTION MODEL

Going through the training, validation and testing stages discussed in the previous section, an ANN prediction model is developed. In the development of the ANN model, a total of 78,004 data point is used. Regression plot of correlation between targeted and predicted travel time results is high with R^2= 0.9530, as shown in Figure 6.11.

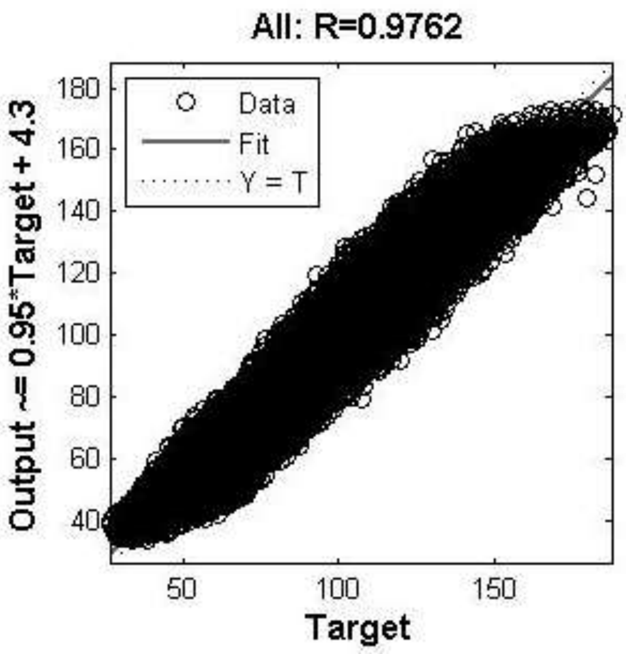


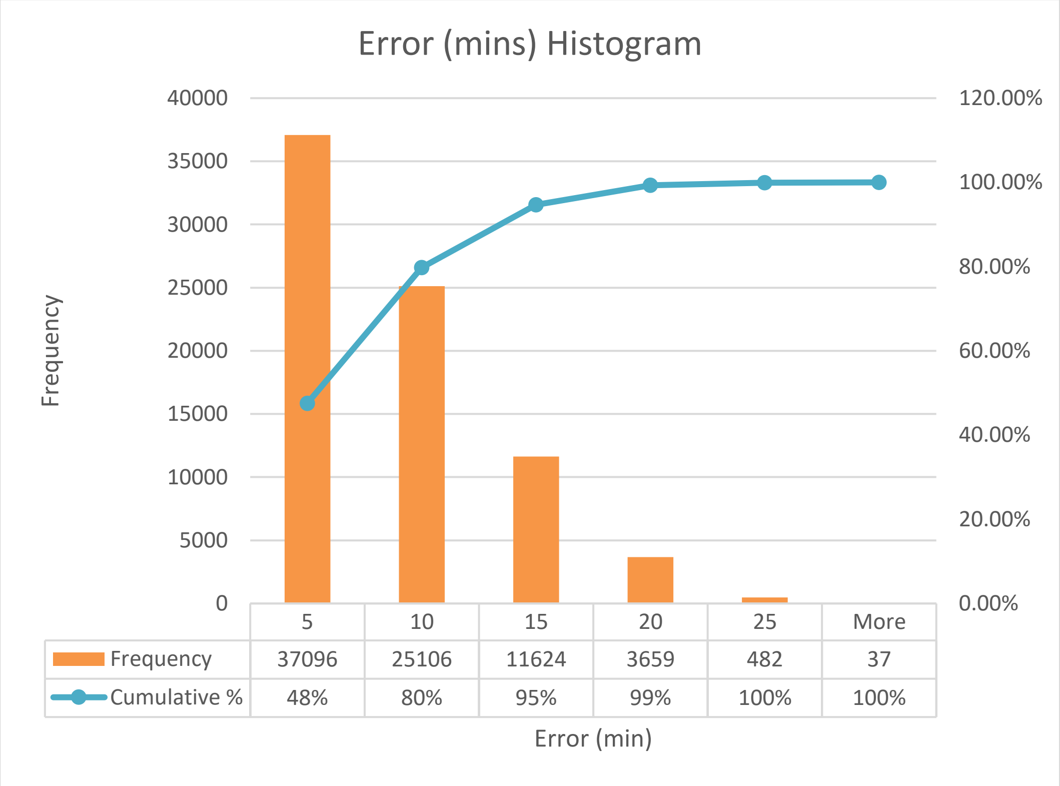
Figure 6.11: ANN Regression Plot Generated by Neural Network Toolbox in MATLAB

MOE for ANN prediction model summarized in Table 6.7 below shows the very high coefficient of determination with R^2=0.95 and MAE of 6.23 minutes.

Table 6.7 ANN Model Measures of Effectiveness

MAE (min)	Coefficient of Determination (R^2)	Standard Error (SE) (min)
5.94	0.95	0.13

Error histogram is shown in Figure 6.12. The error analysis revealed that 95% of predictions have an error of 15 minutes or less. While 80% of total predictions have an error of 10 minutes or less. Moreover, only 1% of predictions has error more than 20 minutes.



	5	10	15	20	25	More
Frequency	37096	25106	11624	3659	482	37
Cumulative %	48%	80%	95%	99%	100%	100%

Figure 6.12 ANN Prediction Model Error Analysis – in Minutes

In addition to the above analysis of error in minutes, relative error analysis was conducted and shown in Figure 6.13. Majority of predictions error represents 10% of total trip duration. While 1% of predictions resulted in an error that is more than 25% of total trip travel time.

The analysis and results of ANN model development stage show that developed model is promising. The developed model had successfully predicted travel time for public transit buses with minimal error. Moreover, MOE reflected consistency throughout model development stage.

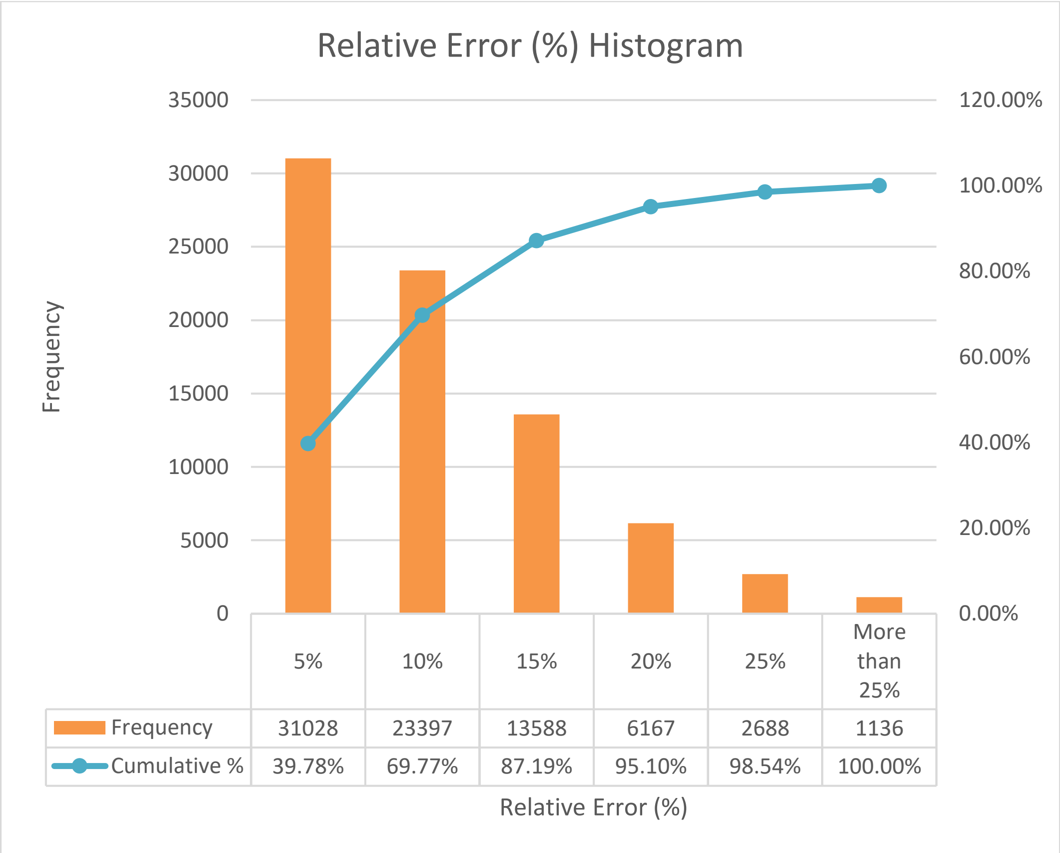


Figure 6.13 Ann Prediction Model Error Analysis – Percentage

6.7. MODEL VALIDATION

The methodology of model validation mentioned in Section 4.4 was used to validate the developed prediction model. Three independent route data was randomly selected for the validation process. The routes that were not included in the prediction model are the routes numbered 11, 43 and 302. Using pre-processing data methodology mentioned in section 4.5 data points for each route was pre-processed independently, then the prediction model was used to evaluate the travel time for each route. The below results represent the validation results for each of the mentioned routes.

6.7.1. ROUTE 11 – VALIDATION RESULTS

Using a total of 13,521 data points for this route that was collected over a year. An evaluation of travel time using developed model was generated. Figure 6.14 below shows the regression plot between targeted and predicted travel time. The validation results for Route 11 using developed prediction MOE are tabulated in Table 6.8 below.

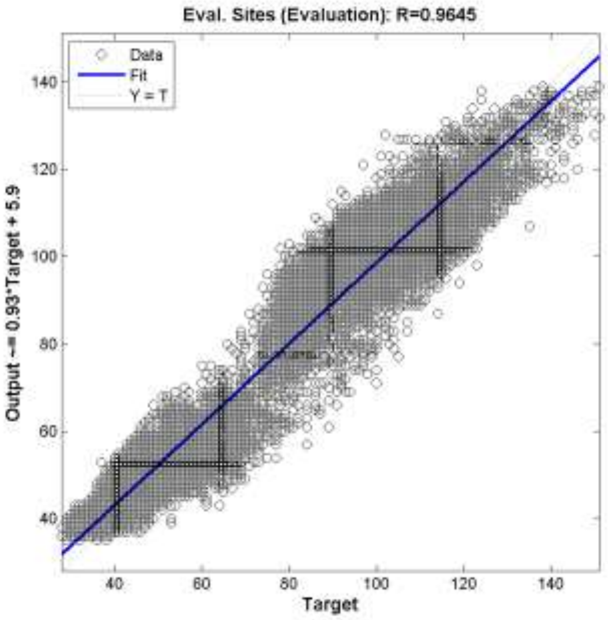


Figure 6.14 Regression Plot of ANN Prediction Results of Route 11

MOE reflects high correlation and low prediction error among model predictions for route 11 data points. These results emphasize on the effectiveness of the model in predicting public transit travel time.

Table 6.8 Route 11 Measures of Effectiveness

MAE (min)	Coefficient of Determination (R^2)	Standard Error (SE) (min)
5.91	0.93	0.23

6.7.2. ROUTE 43 – VALIDATION RESULTS

Using a total of 7,751 data points for this route that was collected over a year an evaluation of travel time using developed model was generated. Figure 6.15 below shows the regression plot between targeted and predicted travel time. The validation results for Route 43 using developed prediction MOE are tabulated in Table 6.9 below.

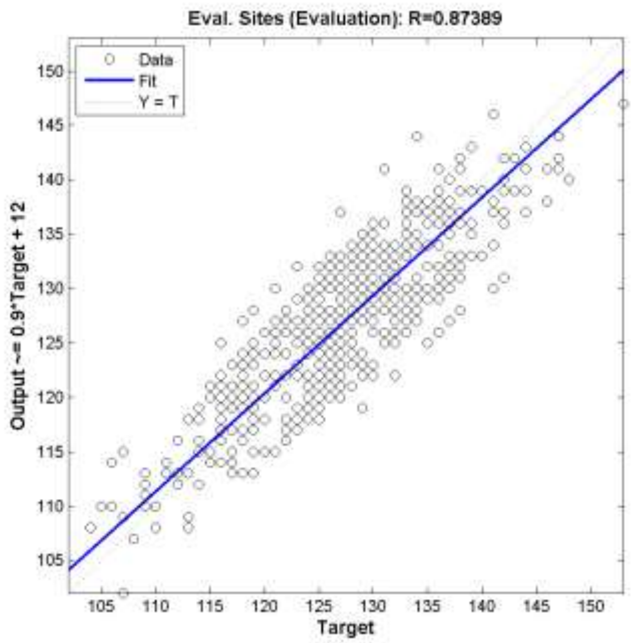


Figure 6.15 Regression Plot of ANN Prediction Results of Route 43

Although the correlation is lower than route 11, however, it still can be interpreted as very high.

Table 6.9 Route 43 Measures of Effectiveness

MAE (min)	Coefficient of Determination (R^2)	Standard Error (SE) (min)
3.04	0.76	0.08

6.7.3. ROUTE 302 – VALIDATION RESULTS

Using a total of 6,250 data points for this route that was collected over a year an evaluation of travel time using developed model was generated. Figure 6.16 below shows the regression plot between targeted and predicted travel time. The validation results for Route 302 using developed prediction MOE are tabulated in Table 6.10 below.

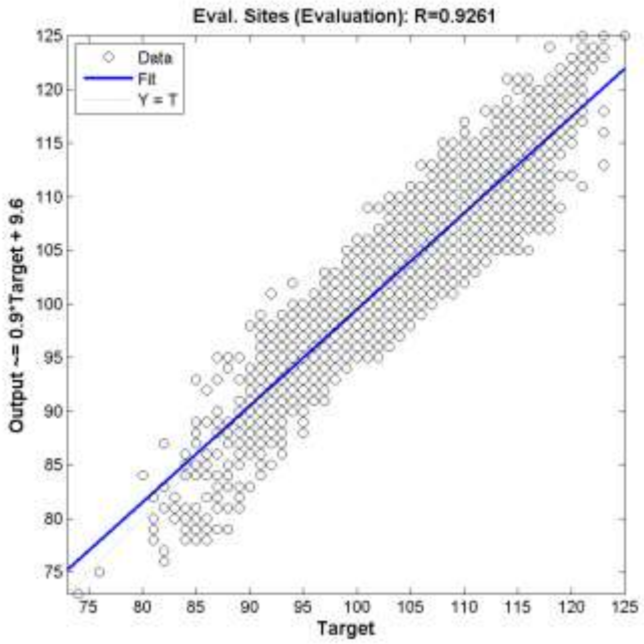


Figure 6.16 Regression Plot of ANN Prediction Results of Route 302

Minimal MAE of 2.24 minutes is found for route 302 prediction results in association with very high coefficients of correlation and determination.

Table 6.10 Route 302 Measures of Effectiveness

MAE (min)	Coefficient of Determination (R^2)	Standard Error (SE) (min)
2.24	0.86	0.09

6.8. TRAVEL TIME ANALYSIS

Analysis of prediction model results and behavior is discussed in the section. IBM SPSS Software that is specialized in data analysis is used to develop the below analysis of travel time behavior. Data points and prediction model results were analyzed based on each of the categorical factors used during the model development stage. The below subsections will discuss the findings of this analysis and travel time behavior in comparison to these categorical parameters.

6.8.1. WEEKDAY – WEEKEND

During weekday, busy Doha City roads are usually congested. Travel time would sometimes significantly vary between weekday and weekend. To understand the behavior of public transport buses travel time during weekday compared to weekends, an analysis was conducted.

The average travel time of public transit buses over the nine routes used in model development was found to be approximately 89.5 minutes during weekdays. While weekends average travel time was found to be approximately 85 minutes. Average travel time captured via AVL and model prediction results were found to be almost identical as shown in Figure 6.17.

MAE in prediction results was found to be higher in weekdays compared to weekends prediction errors as illustrated in Figure 6.18.

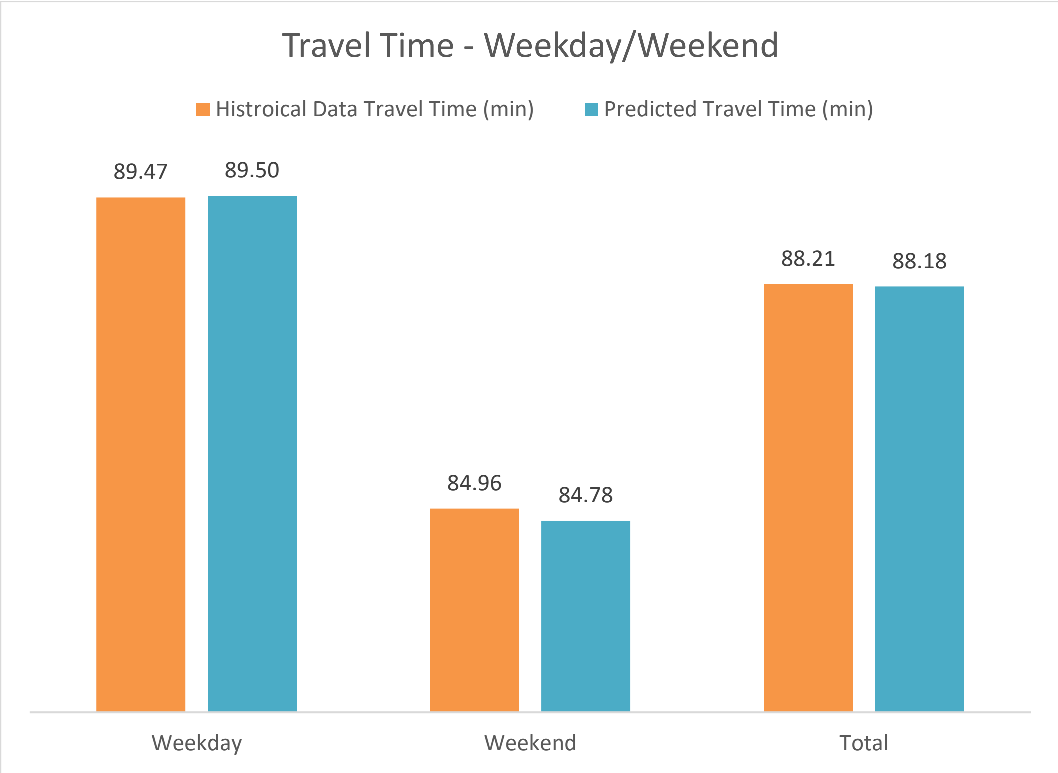


Figure 6.17 Travel Time Analysis for Weekdays vs Weekends

However, relative error was found to be higher during weekends than weekdays. The variation in prediction results error is insignificant, refer to Figure 6.18. The difference between predictions of weekdays compared to weekends in terms of MAE is less than 4 seconds.

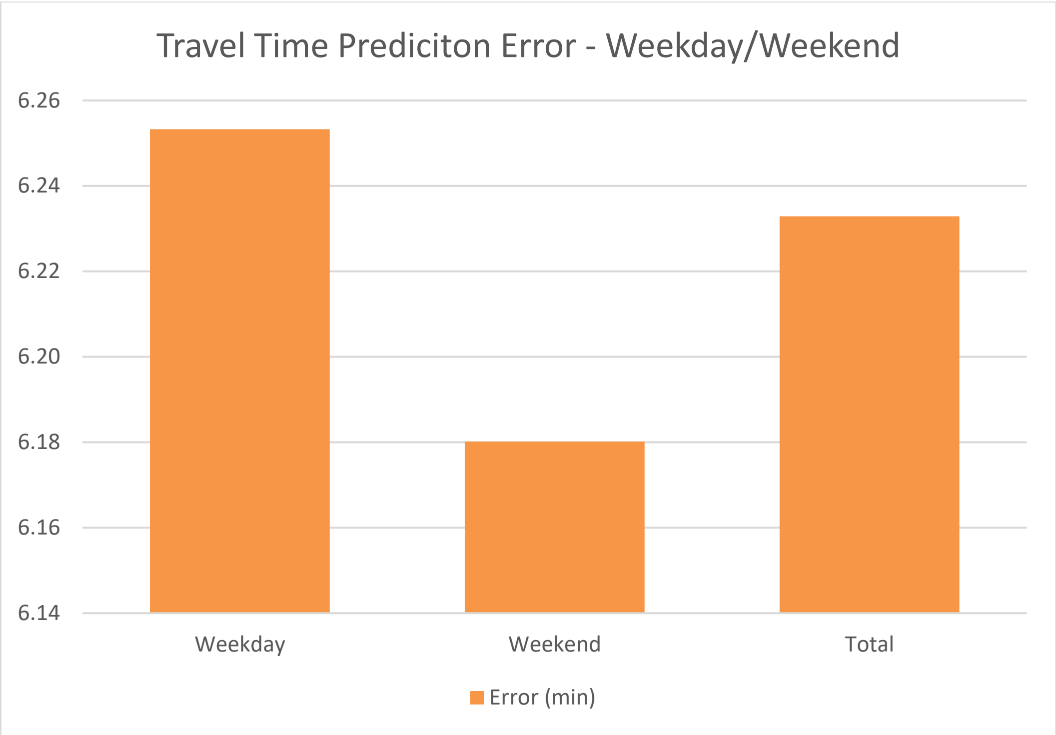


Figure 6.18 Error in ANN Predictions Analysis for Weekdays vs Weekends

6.8.2. PEAK HOURS

Variation in travel time does not only happen from day to day; it does from hour to another. Collected and predicted travel times of public transit buses was segregated per peak; to understand the variation that is happening.

It can be noticed that evening peak (PM) have the highest travel time with 96.5 minutes on average. While off-peak travel time shows an average of 84.5 minutes. Both prediction model and collected data had shown almost identical average travel time per peak as shown in Figure 6.19.

Although both relative and absolute error showed some variation between peak hours, the error difference is insignificant as illustrated in Figure 6.20.

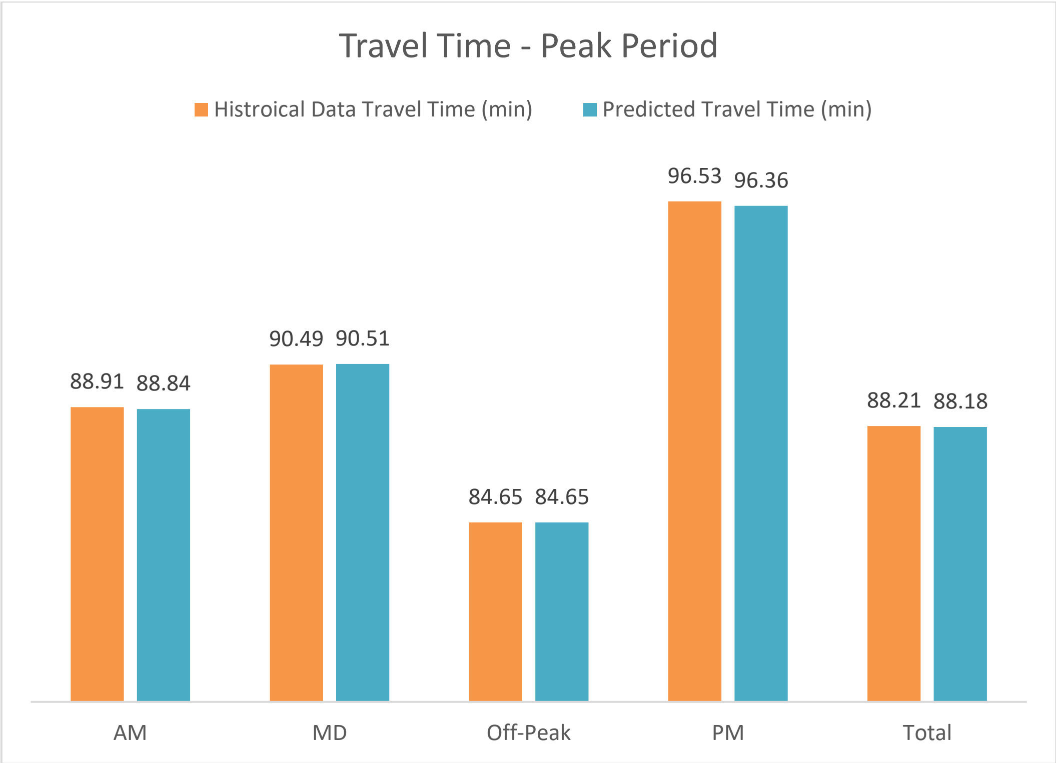


Figure 6.19 Travel Time Analysis per Peak Period

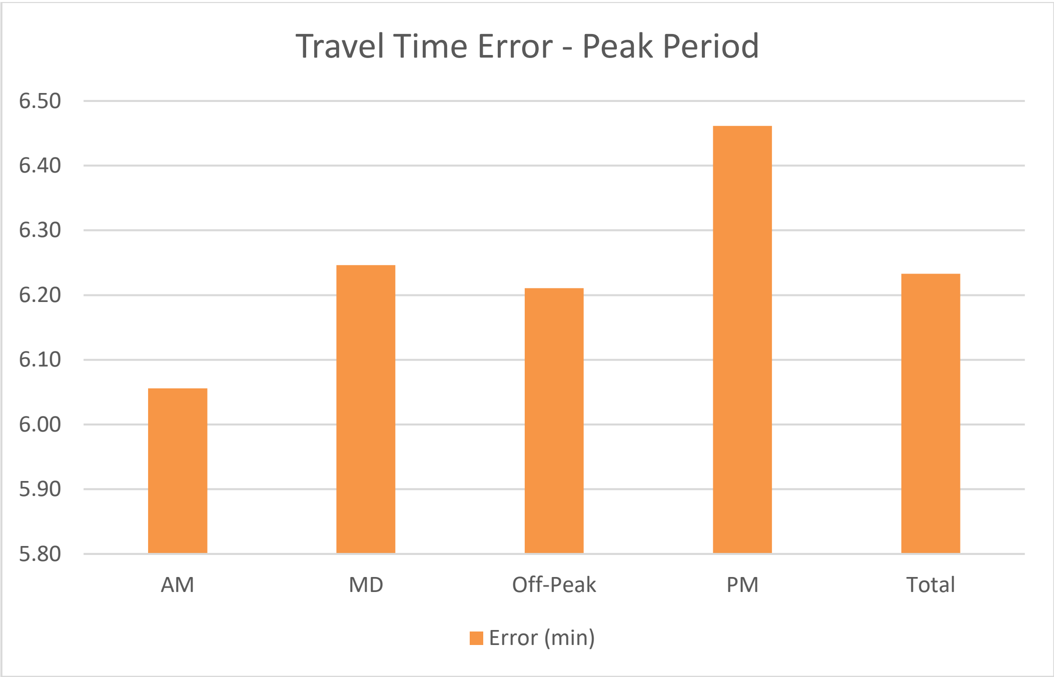


Figure 6.20 Error in ANN Predictions Analysis per Peak Period

6.8.3. MONTH

It’s interesting to see the travel time per month distribution. The behavior of traffic during summer season, months from June to September experience lower travel times. While months that falls in spring or fall seasons were noticed to experience more traffic.

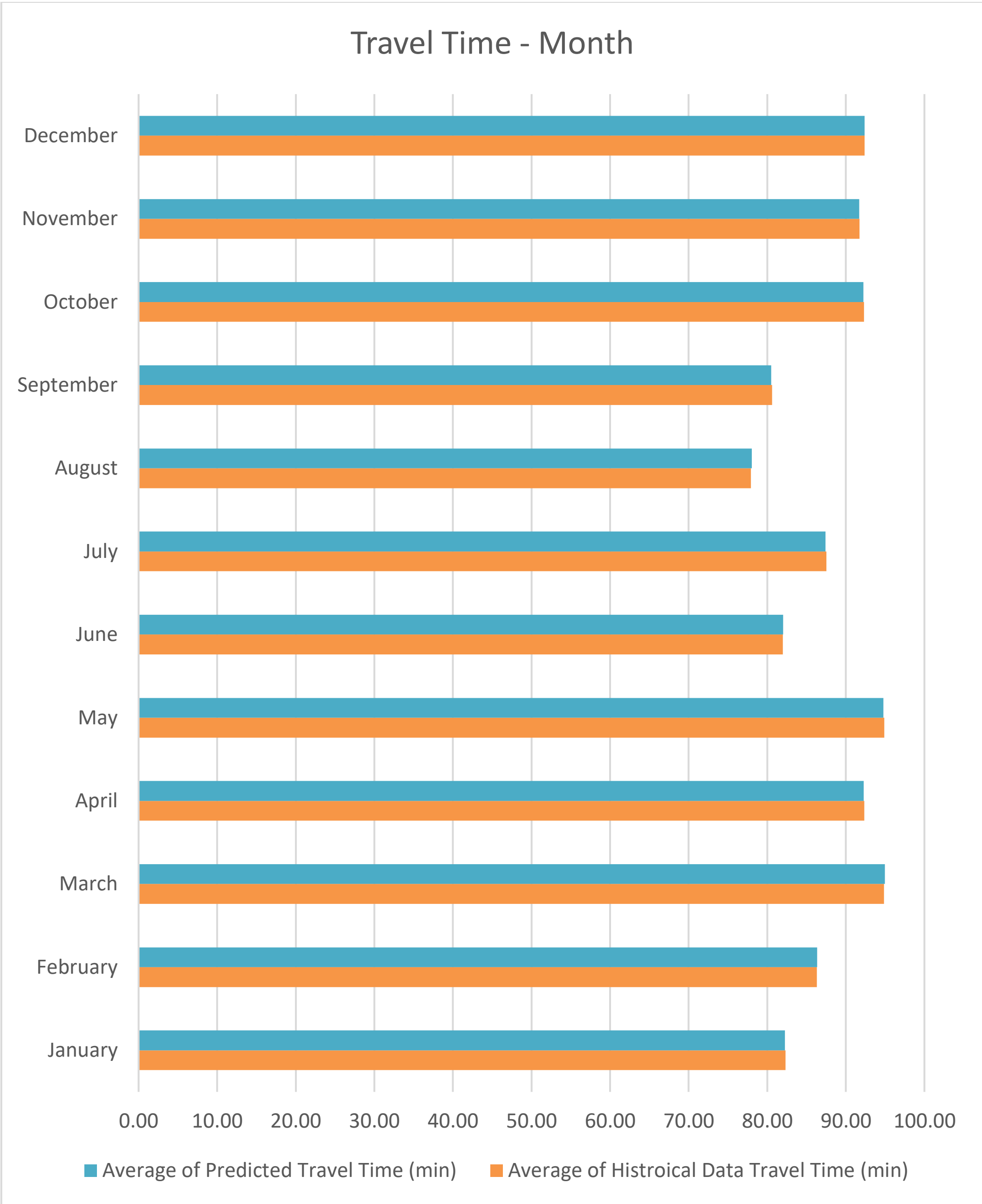


Figure 6.21 Travel Time Analysis per Month

Prediction error per month was analysed as well. Variation in absolute error were less than 1 minute between months.

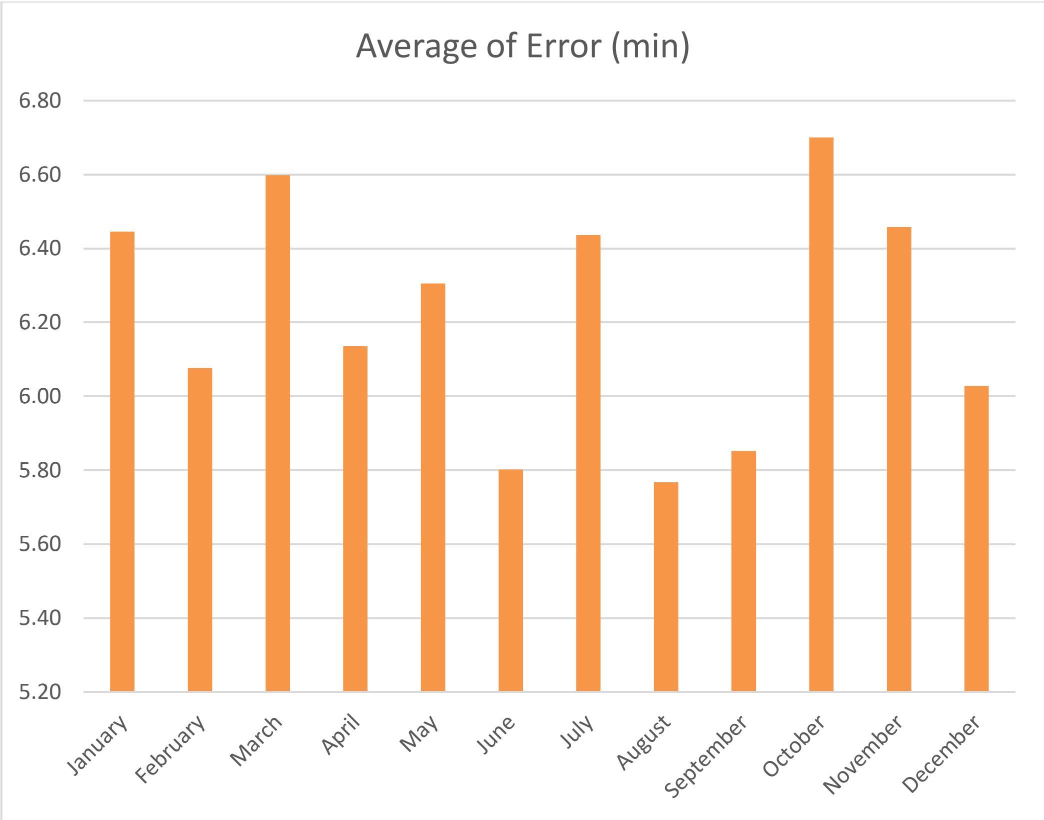


Figure 6.22 Error in ANN Predictions Analysis per Month

6.8.4. SCHOOL TIME VS. HOLIDAYS

During the holiday season (from June to August) a significant drop in traffic can be noticed. Analysis of travel time shown below summarize the behavior of public transit buses during holiday seasons versus school time.

In average 5 minutes of difference in travel time of public transport buses can be found in holidays season compared to school time.

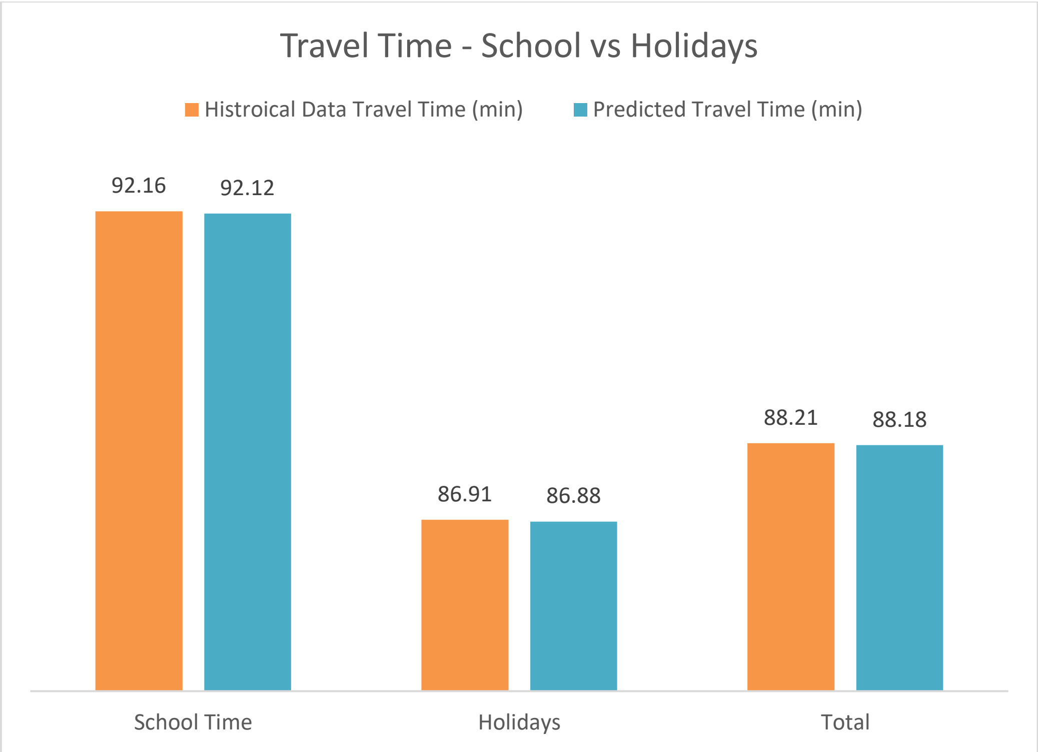


Figure 6.23 Travel Time Analysis per Peak Period

Although both relative and absolute error showed some variation between school and holiday periods, the error difference is insignificant as illustrated in Figure 6.24.

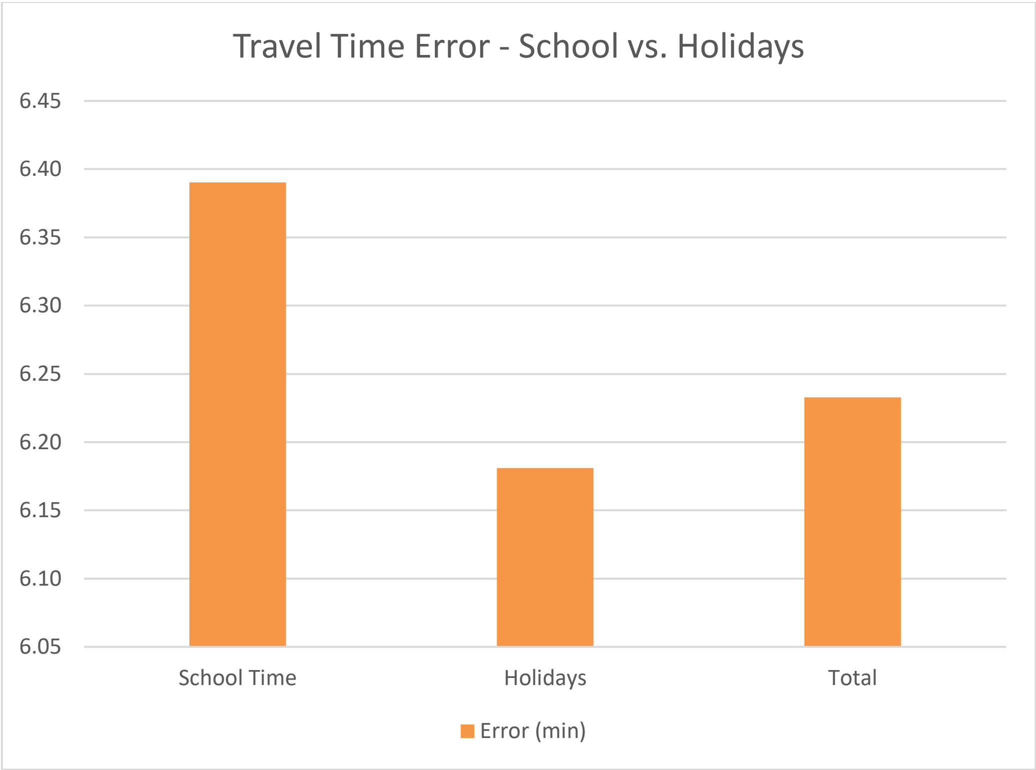


Figure 6.24 Error in ANN Predictions Analysis for School vs Holidays

6.8.5. CENTRAL BUSINESS DISTRICT

The area classification that public transit route operates within does have an impact on the travel time of the trip. The average travel time of trip based on its location is summarized in Figure 6.25.

Public transport buses operating within non-CBD areas has longer travel times significantly compared to buses operating within inner-CBD areas as shown in Figure 6.25. This can be explained by the fact that the central station for public transit buses is located within the inner-CBD area. Majority of public transit buses starts at this station. Thus, it is expected for inner-CBD buses to have shorter durations.

Absolute error showed some variation between different CBD areas, the error difference is insignificant as illustrated in Figure 6.26

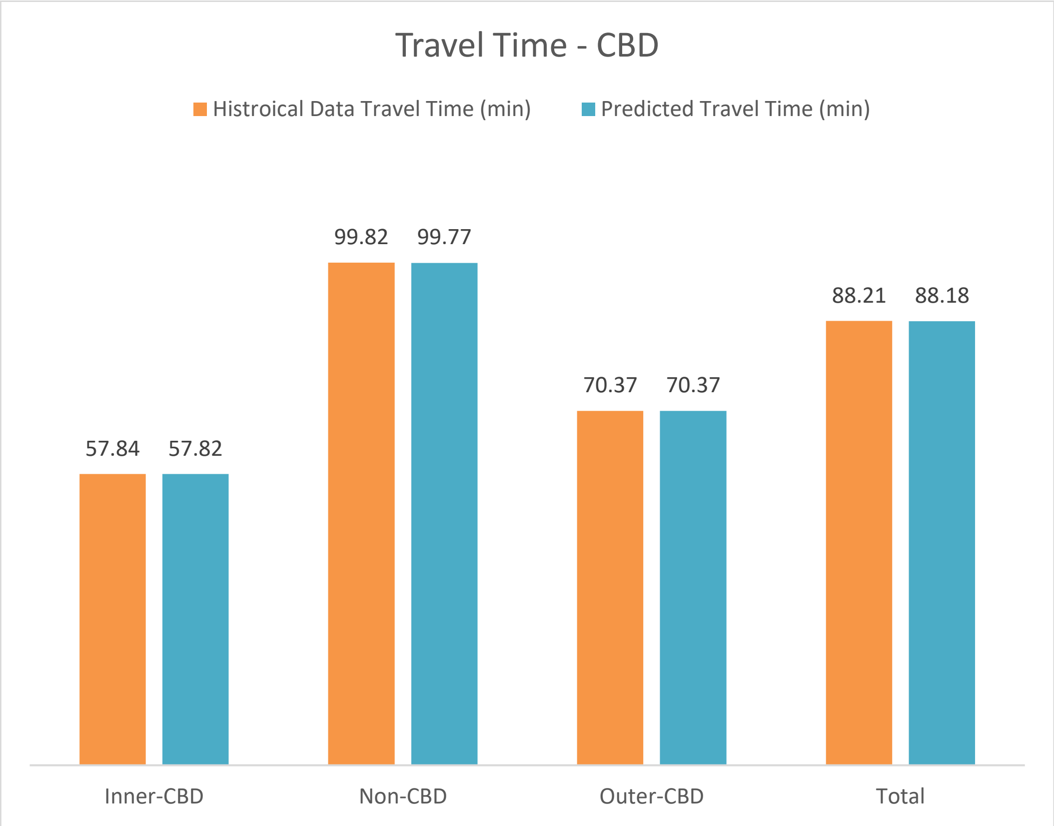


Figure 6.25 Travel Time Analysis per CBD

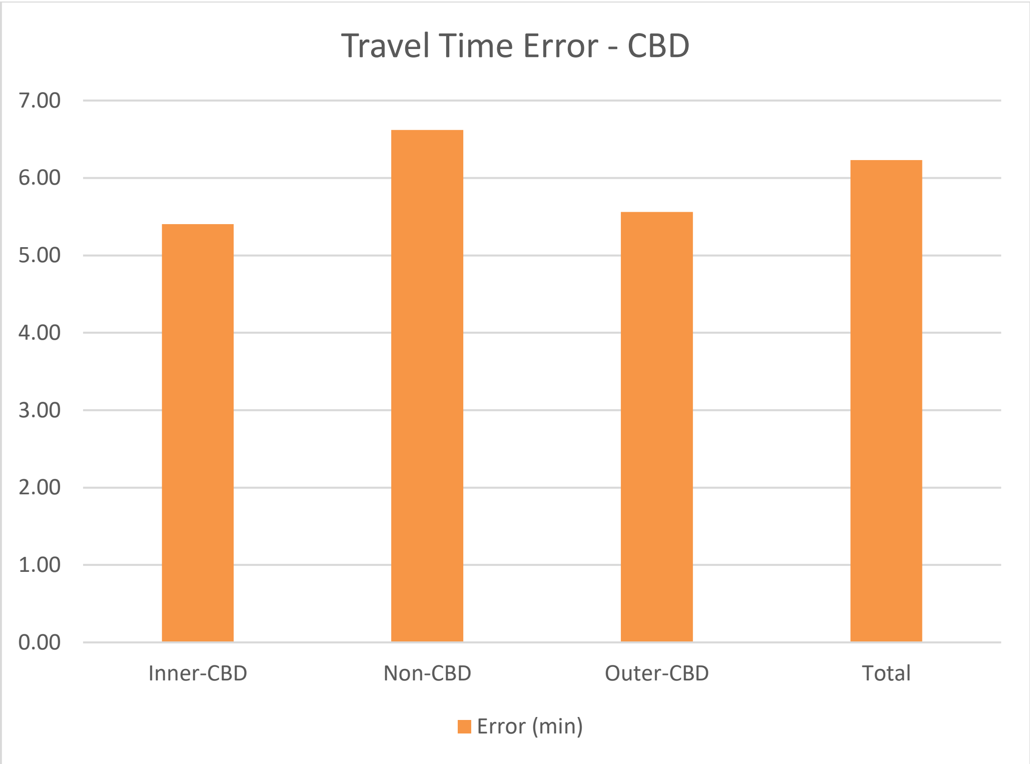


Figure 6.26 Error in ANN Predictions Analysis per CBD

6.8.6. LAND USE

The land use of the area that public transit route operates within proved to have an impact on the travel time of the trip. The average travel time of trip based on its surrounding area land use is shown in Figure 6.27.

Mixed land use was found to have significantly higher travel times than any other land use as shown in Figure 6.27. Moreover, mixed land use had shown lower standard deviation than other land uses. This could indicate lower variability in travel time of mixed land use trips compared to the travel time of trips surrounded by other land uses.

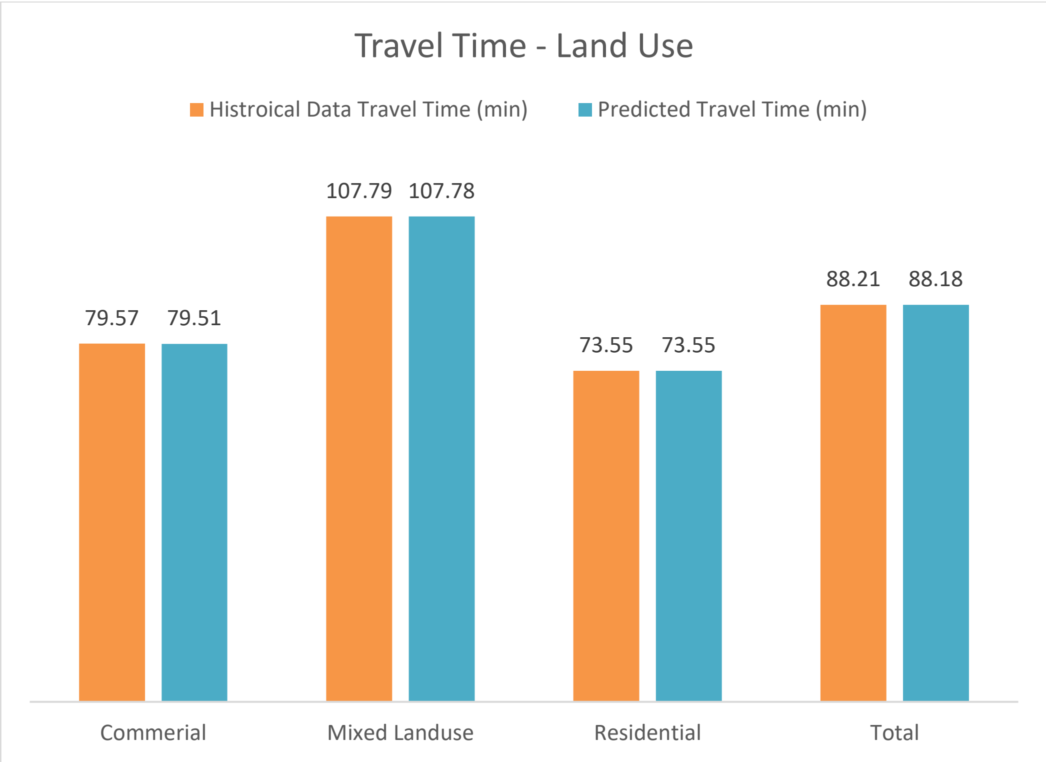


Figure 6.27 Travel Time Analysis per Land Use

Absolute error showed some variation between different land use areas, the error difference is insignificant as illustrated in Figure 6.28.

Figure 6.28 Error in ANN Predictions Analysis per Land Use

6.8.7. ROAD FUNCTIONAL CLASSIFICATION

Road functional classification and its impact on the travel time of public transport buses were analyzed and illustrated in Figure 6.29

Buses that run on major roads tend to have longer travel time compared to buses running on local roads as shown in Figure 6.29. That is mainly due to route planning of buses, at which longer trips tend to cover major roads. While short trips are intended to address the needs of users within local roads. On the other hand, buses that run through local and minor roads were associated with higher standard deviation compared to other buses.

Absolute error showed some variation between major and minor road classification, the error difference is insignificant as illustrated in Figure 6.30.

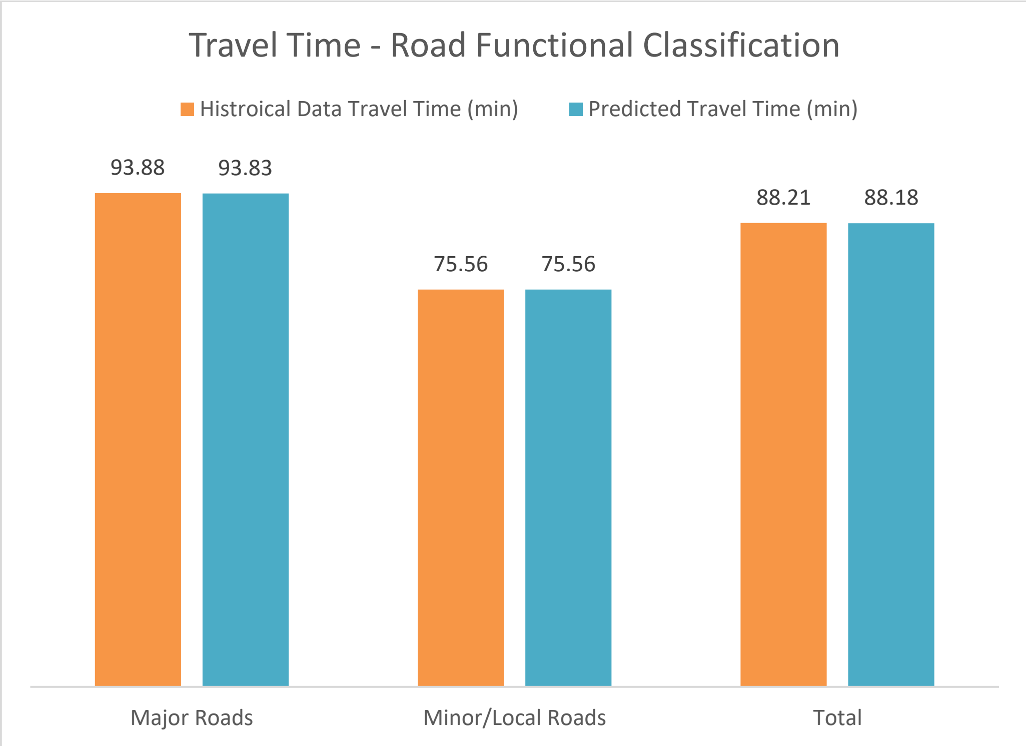


Figure 6.29 Travel Time Analysis per Road Functional Classification

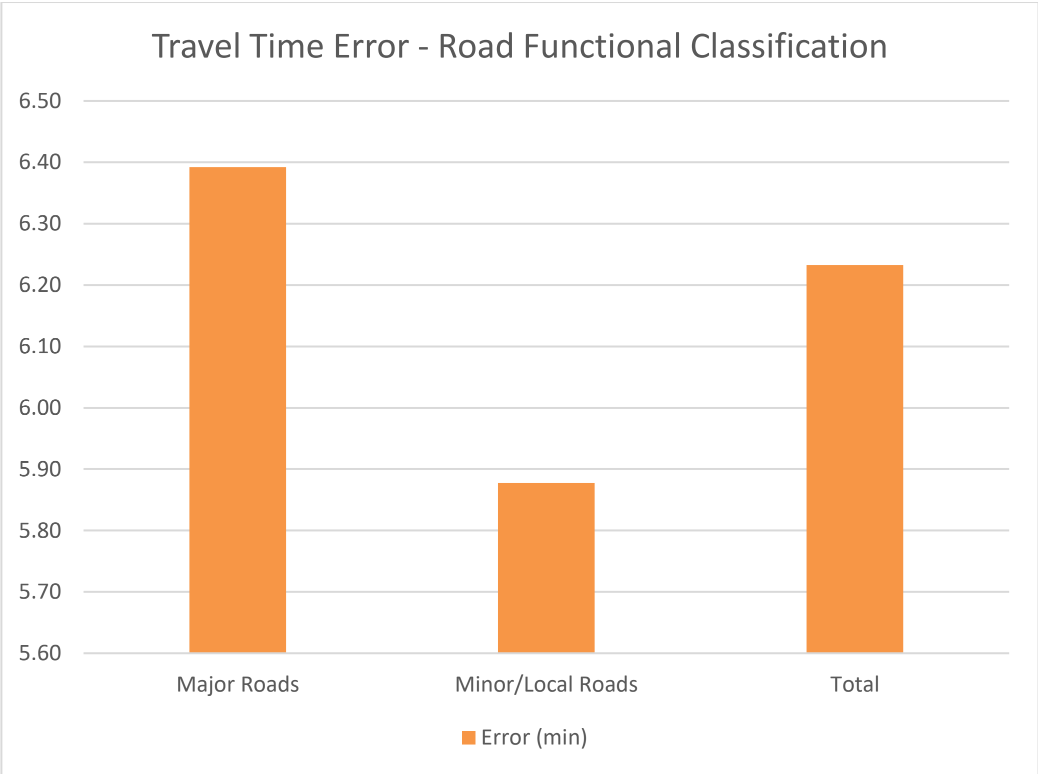


Figure 6.30 Error in ANN Predictions Analysis per Road Functional Classification

7. CONCLUSION

In Qatar, public transport buses service is providing affordable and clean mode of transport to very big category of the population (Shaaban & Khalil, 2013). However, reliability, which is defined as adherence of bus arrival to scheduled timings, is an issue that passengers of public transport buses reported as poor (Shaaban & Khalil, 2013). Literature didn't report development of any public transport travel time model that was built on big data set; which questions the generalization of the prediction models developed.

This research aims to develop a prediction model that can be useful tool to public transport operators and concerned agencies to develop more realistic arrival time and travel time of buses; using big amount of data over long period, a year, to improve generalization of the model.

Data for 12 routes of public transport buses were collected from Mowasalat. The shared data were generated by KentKart platform, which is "the Automated Fare Collection, Vehicle Management, Real Time Passenger Information, Planning and On-Board Video Surveillance Systems" (Kentkart, 2017)that is being used by Mowasalat. Total of more than 150,000 records had been obtained and analyzed.

Several predication models were developed using the parameters of distance, month, weekday, weekend, peak, starting time, through movements, left movements, right movements, U-turn movements, CBD area, road class, land use surrounding the route, number of stops, number of traffic signals and number of roundabouts to predict travel time of public transport buses.

The total of 78,004 data points is used in ANN model development stage. To ensure the validity of the ANN training process, three independent groups of data points were created, and the 78,004 data points were randomly assigned to one of the three groups. Which are training, validation and testing that were divided into 70%, 15% and 15%, respectively.

Artificial Neural Networks (ANN) was used to build the prediction model after its proven record in promising results reported in literature. Total of 8 network architectures were experimented to achieve the model which resulted in the best prediction accuracy. The develop ANN architectures were evaluated using multiple MOE such as R^2, MSE and MAE. The results of travel time prediction of all tested models were highly correlated to the targeted travel time, historical data travel time. The final ANN architecture parameters as summarized in Table 7.1.

Table 7.1 Final ANN Architecture

Model ID	Number of Hidden Layers	Transfer Function	Training Algorithm	Learning Algorithm	Number of Neurons	Performance Function
2	1	Hyperbolic Tangent	Levenberg–Marquardt	Gradient Descent with Momentum	10	MSE

The ANN prediction model predicted the travel of public transit buses with MAE of 5.94 minutes with R^2=0.95. Other MOE values of final ANN architecture are shown in Table 7.2.

Table 7.2 Final ANN Architecture MOE

Model ID	Number of Iterations	R	R^2	MSE	MAE	SE
2	117	0.9762	0.953	57.05	5.94	0.13

The results showed that more than 80%, of predictions had error less than 10 minutes. While errors more than 15 minutes represents less than 5% of the 78,004 predictions.

The prediction model was validated using three independent routes which reported determination coefficients of 0.936 and 0.81 for two routes.

7.1. TRAVEL TIME DATA ANALYSIS SUMMARY

Analysis of prediction model results and behavior using IBM SPSS Software was conducted. Data points and prediction model results were analyzed based on each of the categorical factors. The results of both ANN prediction model and historical data analysis were found to be identical. The analysis revealed that public transit buses takes an average 5 minutes more to arrive its destination during school time compared to holidays season. Moreover, public transit buses take an average of 10 minutes extra during PM peak compared to off-peak hours; which could impact short trip travel time significantly. The analysis showed that buses with residential land use surrounding its route has the least average travel time, while mixed land use was found to have the longest travel time.

7.2. PREDICTION MODEL LIMITATIONS

Although the aim of this research is to develop a generalized prediction model for public transport buses travel time, there are number of limitations to the developed model that shall be carefully considered during usage of this model:

1. Theoretical route travel time duration shall not exceed 188 minutes, while it's recommended to keep the duration to be less than 160 minutes for higher accuracy
2. Applicable to cities where city planning is like Doha City
3. There is a margin of prediction error of average 7.7% expected in the model results

7.3. FUTURE WORK AND RESEARCH AREAS

During the research work on this project, the following future work and research areas are recommended to be explored and investigated:

1. The model can be improved by connecting it to an active database which will improve the prediction results.
2. Testing the model in other countries would be a potential research area that might result in regional prediction model.
3. Investigation on the applicability of the developed model on predicting stop-to-stop travel time would be interesting and very useful to end users and commuters of public transport system in Qatar and the region, potentially.
4. The model could be improved to consider other factors that were not accessible at the time of this research; however, would be of significant impact as reported in literature such as ridership and real-time traffic conditions.

7.4. RECOMMENDATIONS

With only one peer-reviewed publication about public transport system in Qatar; this research is considered a foundation for knowledge based improvement to public transport system in Qatar. The thesis addresses the gap in knowledge about Qatar & the region public transport systems.

Based on the research work documented in this thesis further research in the area of public transport system is needed to understand the shortcomings of the public transport system and user behavior in the region.

ANN developed model developed in this research was found to be a reliable tool that is recommended to be improved via further research and application by public transport buses operators and agencies. This prediction of public transport buses travel time and arrival time can positively impact reliability of the system; which was reported to be in poor conditions in Qatar (Shaaban & Khalil, 2013).

This research highlighted the capabilities of ANN for application on public transport buses travel time prediction. Levenberg–Marquardt algorithm, which was associated with low computational cost and short converging time, is recommended to be used as the training technique to predict travel time. ANN prediction model showed high prediction accuracy with low prediction error using exceptional amount of data. The model was developed using large number of parameters and data that were collected over a year period; which improves its capabilities in predicting vast number of cases.

ANN model develop is recommended to be used in planning of future public transport routes. Moreover, the prediction model can be used as quality control and restructuring guidance of existing public transport routes. It’s recommended to use the develop model to review and validate the existing public transport buses time-tables.

Qatar is currently working on developing advanced traveler information systems (ATIS). Public transport buses travel time prediction model developed in this work can be used as knowledge based building prediction model for ATIS system in Qatar. The developed model has several advantages compared to imported and earlier developed models in terms of its generalization and capabilities in predicting travel time of buses; which gives it advantage over other developed models to be used in intelligent transport systems (ITS) platforms development, in Qatar and the region.

8. BIBLIOGRAPHY

Qatar Meteorology Department. (2017, Nov 23). *Qatar Meteorology Department*. Retrieved from http://www.qweather.gov.qa/ClimateNormals.aspx

Ayegba, P. O., & Abdulkadir, M. (2016). PREDICTION OF AVERAGE VOID FRACTION AND PDF OF VOID FRACTION IN VERTICAL 90O BEND FOR AIR–SILICONE OIL FLOW USING MULTILAYER PERCEPTRON (MLP) CODES. *International Journal of Lean Thinking*, 80-105.

Chien, S. I. (2003). Dynamic Travel Time Prediction with Real-Time and Historic Data. *Journal of Transportation Engineering*, 608-616.

Dewan, K., & Ahmad, I. (2007, February). Carpooling: a step to reduce congestion (A Case Study of Delhi). *Engineering Letters*, *14*(1), 61-66.

Dick Ettema, H. T. (2006). Costs of travel time uncertainty and benefits of travel time information: Conceptual model and numerical examples. *Transportation Research Part C: Emerging Technologies*, *14*(5), 335-350.

ENGELSTEIN, M. D. (1982). FACTORS AFFECTING RUNNING TIME ON TRANSIT ROUTES.

Etienne Hans, N. C. (2015). Real-time bus route state forecasting using Particle Filter: An empirical data application. *Transportation Research Procedia*, 434 – 447.

Fahmy, H. (2015). *Doha News*. Retrieved from Mowasalat plans public transit expansion with more buses on Doha roads: https://dohanews.co/mowasalat-plans-public-transit-expansion-with-more-buses-on-doha-roads/

Ghanim, M. S. (2011). Florida Statewide Design-Hour Volume Prediction Model. *Transportation Research Board*.

Ghanim, M. S., Abu-Lebdeh, G., & Ahmed, K. (2013). Modeling Historical Traffic Data using Artificial Neural Networks. *5th International Conference on Modeling, Simulation & Applied Optimization (ICMSAO)* (pp. 1-4). Tunisa: IEEE.

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*, 154-163.

Houghton, J., Reiners, J., & Lim, C. (2009, June). *Intelligent transport: How cities can improve mobility*. New York: IBM. Retrieved March 23, 2017, from https://public.dhe.ibm.com/common/ssi/ecm/gb/en/gbe03232usen/global-business-services-global-business-services-gb-executive-brief-gbe03232usen-20170629.pdf

Hubschneider, I. H. (2011, May). *Transport Master Planning in the Middle East.* Retrieved March 23, 2017, from http://www.gabf.ghorfa.de/fileadmin/inhalte/wirtschaftsforum/2011/Sessions/Session%203/Hubschneider.pdf

Islam, M. V. (2015). A model to evaluate the impact of headway variation and vehicle size on the reliability of public. *IEEE Trans. Intell. Transp. Syst.*, *16*(4), 1840–1850.

Jeong, R. R. (2004). Bus arrival time prediction using artificial neural network model. *The 7th International IEEE Conference on Intelligent Transportation Systems* (pp. 988–993). IEEE.

Johar Amita, J. S. (2016). Prediction of Bus Travel Time using ANN: A Case Study in Delhi . *Transportation Research Procedia*, 263-272.

Kasabov, N. (1995). *Foundations of Neural Networks.* London: MIT.

Kentkart. (2017, November 23). *Kentkart*. Retrieved from Kentkart: kentkart.com

Lee, S. a. (2001). A review of data mining techniques,. *Journal of Industrial Management & Data Systems*, 41-46.

Marhaba. (2016, March 31). *West Bay Shuttle Services on Fare Basis Starting 1 April*. Retrieved from Marhaba: https://www.marhaba.qa/west-bay-shuttle-services-on-fare-basis-starting-1-april/

MathWorks. (2017, November 23). *Improve Neural Network Generalization and Avoid Overfitting*. Retrieved from MathWorks: https://www.mathworks.com/help/nnet/ug/improve-neural-network-generalization-and-avoid-overfitting.html?requestedDomain=www.mathworks.com#bt0cnqo-1

Mazloumi, E., Currie, G., & Rose, a. G. (2010). Using GPS Data to Gain Insight into Public Transport Travel Time Variability. *JOURNAL OF TRANSPORTATION ENGINEERING,* 623-631.

Ministry of Development Planning and Statistics. (2017, February 15). *Homepage*. Retrieved March 23, 2017, from http://www.mdps.gov.qa/en/Pages/default.aspx

Ministry of Transport and Communication. (2017, April 12). *20% Rise in Bus Ridership in 2016, Improvements Continue*. Retrieved from Ministry of Transport and Communication: http://www.motc.gov.qa/en/news-events/news/20-rise-bus-ridership-2016-improvements-continue

Mitchell, T. M. (1997). *Machine Learning.* Toronto: McGraw Hill.

Moawia Elfaki Yahia, M. E.-m.-t. (2010). A New Approach for Evaluation of Data Mining Techniques. *IJCSI International Journal of Computer Science Issues*, 1696.

O. Kisi, E. U. (2005). Comparison of the three backpropagation training algorithms for two case studies, 12 (5) . *Indian J Eng Mater Sci*, 434–442.

Patnaik, J. C. (2004). Estimation of Bus Arrival Times Using APC Data. *Journal of Public Transportation,* 1-20.

Schweiger, C. L. (2003). *TCRP Synthesis 48: Real-Time Bus Arrival Information Systems: A Synthesis of Transit Practice.* Washington, D.C: Transportation Research Board of the National Academies.

Shaaban, K., & Khalil, R. F. (2013, December 2). Investigating the Customer Satisfaction of the Bus Service in Qatar. *Procedia - Social and Behavioral Sciences, 104*, 865-874.

Shalaby, A. a. (2004). Prediction model of bus arrival and departure times using AVL and APC data. *Journal of Public Transportation*, 41-61.

Sinha, N., & Gupta, M. (1999). *Soft Computing and Intelligent Systems - Theory and Applications.* Saskatoon: Academic Press.

T. Kohonene, G. D. (1998). *Visual Explorations in Finance with Self-organizing Maps.* London: Springer.

Tang, L. T. (2012). Ridership effects of real-time bus information system: a case study in the City of Chicago. *Transp. Res. Part C: Emerg., 22*, 146-161.

Teodorović, D., & Vukadinovic, K. (1998). *Traffic Control and Transport Planning: A.* Amsterdam: Kluwer Academic Publishers.

Verbich, D., Diab2, E., & El-Geneidy, A. (2016, June 2016 23). Have they bunched yet? An exploratory study of the impacts of bus bunching on dwell and running times. Berlin , Heidelberg, Germany.

Vuchic, V. (2007). *Urban Transit Systems and Technology.* Hoboken, New Jersey: John Wiley & Sons Inc.

W.S. McCulloch, W. P. (1947). A logical calculus of the ideas immanent in nervous activity. *Math. Biophys.*, 115-133.

Watkins, K. F. (2011). Where is my bus? Impact of mobile real-time information on the perceived and actual. *Transp. Res. Part A: Policy Pract., 45*(8), 839-848.

Wei Fan, Z. G. (2015). Dynamic Travel Time Prediction Models for Buses Using Only GPS Data. *International Journal of Transportation Science and Technology*, 353 - 366.

Wilamowski, H. Y. (2011). “Levenberg-Marquardt Training,”. In *The Industrial Electronics Handbook, Vol. 5−Intelligent Systems, 2nd ed. .* Boca Raton: CRC Press.

Xinghao, S. J. (2013). Predicting bus real-time travel time basing on both GPS and RFID data. *Procedia-Social and Behavioral Sciences*, 2287-2299.

Yamakawa, T. (1990). Pattern recognition hardware system employing a fuzzy neuron. *Proceedings of the International Conference on Fuzzy Logic and Neural Networks*, (pp. 943–948). Iizuka, Japan.

Zhengyao Yu a, J. S. (2017). Using survival models to estimate bus travel times and associated uncertainties. *Transportation Research Part C, 74*, 366-382.

APPENDIX A – ARTIFICIAL NEURAL NETWORK MATLAB CODE (TYPICAL)

```
% Size of original data m is the number of cases and n is the number of attributes

clear;
clc;
load C:\Users\mothasem\Dropbox\Thesis_1\Analysis\Matlab\OVERALL\Data.txt ;

% get Network Size. m is for number of records, n is for number of inputs
% and outputs.
[m,n]=size(Data);
disp('Original Data dimensions [Attributes Cases]');
disp([n m]);

p = Data(:,1:41); p = p';
t = Data(:,42); t = t';

% Preparing data for Training, Validation, and Testing based on input
% ratios
trainRatio = .70;
valRatio = 0.15;
testRatio = 0.15;

% Setting up Lambert-Marrquardt FF Backpropagation Network.
%(More details at 'Neural Network Toolbox: Backpropagation: Levenberg-Marquardt (trainlm)')
```

```
disp('Setting up network ...');
net = feedforwardnet(10,'trainlm'); %10 is the number of neurons in hidden layer
net = configure(net,p,t);
net=init(net);
net.layers{1}.transferFcn = 'logsig';              %logsig stands for logistic algorithm
net.layers{2}.transferFcn = 'tansig';     %tansig stands for hyperbolic tangent algorithm
net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
net.outputs{2}.processFcns = {'removeconstantrows','mapminmax'};


% The code below is to randomize training, validation, and testing data points
% in MatLAB without the need to do so in Excel.Otherwise, use 'divideblock'
net.divideFcn = 'dividerand';
net.divideParam.trainRatio = trainRatio;
net.divideParam.valRatio = valRatio;
net.divideParam.testRatio = testRatio;

% Setting up the training parameters
net.trainParam.epochs = 1000;
net.trainParam.goal = 1e-4;    % MSE to reach 0.00000001
net.trainParam.max_fail = 100 ;   % Five Validations to check when error rises
net.trainParam.min_grad = 1e-5;
net.trainParam.mu = 0.0001;
net.trainParam.mu_dec = 0.1;
net.trainParam.mu_inc = 100;
net.trainParam.mu_max = 1e10;
net.trainParam.show = 25;
net.trainparam.lr = 0.05;
net.trainParam.mc = 0.1;
net.trainParam.time = inf;

% Performance = 'Mean Square Error'
net.performFcn = 'mse';
```

```
disp('Training wis starting ...');
[net,tr] = train(net,p,t);
net_progress = plotperf(tr);
saveas (plotperform(tr), 'plotperf','png');
%view(net);

y = net(p);

y_tr = y(:,tr.trainInd);
y_val = y(:,tr.valInd);
y_test = y(:,tr.testInd);
t_tr = t(:,tr.trainInd);
t_val = t(:,tr.valInd);
t_test = t(:,tr.testInd);



plotregression(t,y,'Observed vs. Predicted Value')


%It is also possible to calculate the network performance only on the test set,
%by using the testing indices, which are located in the training record.
%(See Analyze Neural Network Performance After Training for a full description of the training record.)
%testOutputs = net(inputs(tr.testInd));


% Plot the DDHV Training, validation, and Testing
ANN_Plot = plotregression(t,y,'All');
saveas(ANN_Plot,'C:\Users\mothasem\Dropbox\Thesis_1\Analysis\Matlab\OVERALL\ANN_Plot.png');

train_ID = tr.trainInd';
val_ID = tr.valInd';
tst_ID = tr.testInd';

%It is also possible to calculate the network performance only on the test set,
%by using the testing indices, which are located in the training record.
```

```
%(See Analyze Neural Network Performance After Training for a full description of the training record.)

All_Perform = perform(net,t,y);
All_Perform_tr = perform(net,t_tr,y_tr);
All_Perform_val = perform(net,t_val,y_val);
All_Perform_test = perform(net,t_test,y_test);

% Mean Standard Error Calculations
error_all = t - y;
error_tr = t_tr - y_tr;
error_val = t_val - y_val;
error_test = t_test - y_test;

std_all = std2(error_all');
std_tr = std2(error_tr');
std_val = std2(error_val');
std_test = std2(error_test');

%perf_tr = mse(error_tr);        % same as perform(net,t_tr,y_tr)
%perf_val = mse(error_val);      % same as perform(net,t_tr,y_tr)
%perf_test = mse(error_test);    % same as perform(net,t_tr,y_tr)

% Save variables.
regvar = [p', t',y'];
save('C:\Users\mothasem\Dropbox\Thesis_1\Analysis\Matlab\OVERALL\yall.txt', 'regvar',  '-ascii', '-double', '-tabs');


%%%%%%%%%% SAVE MODEL PARAMETERS


%InLim    = [minp,maxp];
Wt1      = net.IW{1};
Bias1    = net.b{1};
%OutLim = [mint,maxt];
Wt2      = net.LW{2};
Bias2    = net.b{2};
```