APPLICATION OF NEURAL NETWORK FOR MODE CHOICE MODELING AND MODAL TRAFFIC FORECASTING

BY

UNEB GAZDER

A Dissertation Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY
In

CIVIL ENGINEERING

April 2014

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **UNEB GAZDER** under the supervision of his thesis advisors and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY IN CIVIL ENGINEERING**.

Thesis Committee

Prof. Nedal T. Ratrout	(Advisor)
Prof.. Hashim Al-Madani	(Member)
Prof. Muhammad A. Abido	(Member)
Dr. Shukri Hasan A. Al-Senan	(Member)
Dr. Hassan Mousaid Al-Ahmadi	(Member)

June 09/2014

Prof. Nedal T. Ratrout
Departmental Chairman

Prof. Salam A. Zummo
Dean of Graduate Studies

9/6/14

Date

عمادة الدراسات العليا
DEANSHIP OF GRADUATE STUDIES

DEDICATED

TO

MY LOVING (LATE) MOTHER

AND

MY CARING FATHER

ACKNOWLEDGEMENTS

All praise is to ALLAH (subhana wa ta'la), the creator and the sustainer of the worlds, for bestowing me the health, patience, knowledge and courage to complete this research. May His peace and blessings be upon the best of His creatures; Prophet MUHAMMAD (sallallahu alaihi wasallam).

I would like to thank King Fahd University of Petroleum and Minerals for providing excellent research facilities and financial assistance for pursuing my PhD.

It's a great privilege to have worked under the experienced supervision of my Thesis committee. I acknowledge sincere appreciation and thanks to my advisor Prof. Nedal T. Ratrout for his handy leadership throughout this research. I am also grateful to my committee members Prof. Hashim Al-Madani, Prof. Muhammad A. Abido, Dr. Shukri Al-Senan and Dr. Hassan Al-Ahmadi for their support during this research. I would also like to thank other faculty members for their concerns.

I owe ample credit to Mr. Ahmed Al-Saif for his logistic support during the research work. He took great interest in my work and was always there whenever I required him. The data collection was the most tedious part of this study in which I was greatly helped by Mr. Foud Rashed, Director of Bahrain stock Market, Ameena Mahmoud, Alaa Rashed and Shaikh Al-Thawadi, students from University of Bahrain (supervised by committee member Prof. Madani), Mr. Ibrahim Al-Qauz and Muhammad Abdullah, personnel King Fahd Airport Authority and General Directorate of traffic in Bahrain. This research would not have completed without their kind support and cooperation.

I thank Mr. Saad Muhammad Saad Khan, Mr. Imran Reza, Dr. Yazen Essa and Mr. Khaled Assi who helped in their fullest capacity in needful times. Also, I thank Umar, Nasir, Asim, Waseem and Luqman for their moral support.

I would always be thankful to my parents for their continuous love and support for me, **for their guidance and all the sacrifices they made**, **to make me what I am today**. **I would like to thank my wife for her uninterrupted support in the most difficult times**.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

THESIS ABSTRACT (ENGLISH)

NAME: UNEB GAZDER

TITLE: APPLICATION OF NEURAL NETWORK FOR MODE CHOICE MODELING AND MODAL TRAFFIC FORECASTING

MAJOR: CIVIL ENGINEERING

DATE: 16th April 2014

Border transport modeling is a unique research area which is under-explored in the present literature. Artificial Intelligence (AI) based techniques, especially Artificial Neural Networks (ANNs), have gained immense popularity among the researchers and practitioners of transportation planning. Ensemble learning is another new and promising approach for modeling traffic by integrating different methods. However, the application of these contemporary techniques for border transport modeling, in general, and mode choice modeling, in particular has been under-explored.

In this research, King Fahd causeway, which connects Kingdom of Saudi Arabia (KSA) and Bahrain has been selected to explore the use of contemporary techniques for border transport modeling. This research addresses employment of these techniques for operational use through accurate traffic forecasting. Moreover, the use of these techniques for long-term planning issues such as mode choice has also been investigated. Different variable sets including stock market indices, weather parameters and air-travel were tested for traffic forecasting. Stock indices have been used for the first time for border traffic forecasting as a surrogate measure of political stability and economic conditions of the countries connected. The effects of these variables were tested for different input time-series data (look-back period) and prediction horizons. The results of the study justify the use of stock market indices for border traffic forecasting. However, the use of other factors, as mentioned above, is also recommended for better accuracies in shorter prediction horizons of 1 and 7 days.

This research also compares the accuracies of logit model, ANN and ensemble learning. A heterogeneous ensemble is proposed, with the combination of logit model and ANN, for mode choice modeling. These models give better performance than single logit models as well as ANN models for multinomial mode choice problems. Thus the use of these ensembles is recommended

for these problems. Moreover, number of modes and input variables also affect the accuracy of all types of models.

THESIS ABSTRACT (ARABIC)

ملخص الرسالة (عربي)

الإسم : انيب غازدير

: تطبيق الشبكات العصبية الاصطناعية لنمذجة اختيار وسيلة النقل وللتنبؤ بالحركة المرورية عنوان البحث

التخصص : هندسة مدنية

تاريخ التخرج :2014/04/16

درجة الدكتوراه في العلوم الهندسية

جامعة الملك فهد للبترول والمعادن

الظهران – ٣١٢٦١

المملكة العربية السعودية

تعتبر نمذجة النقل الحدودي (النقل بين الدول) من المجالات الفريدة من نوعها والتي هي تحت الاستكشاف من قبل الباحثين والدارسين. وقد اكتسبت تقنيات الذكاء الاصطناعي (AI) وخاصة الشبكات العصبية الاصطناعية (ANNs) شهرة كبيرة بين الباحثين والعاملين في تخطيط النقل. وتعتبر النماذج المدمجة نهج جديد وواعد لنمذجة حركة النقل من خلال دمج أساليب وطرق مختلفة. ومع ذلك، فإن تطبيق هذه التقنيات المعاصرة لنمذجة النقل الحدودي بشكل عام ونمذجة اختيار وسيلة النقل بشكل خاص ما يزال تحت الاستكشاف.

في هذا البحث تم اختيار جسر الملك فهد الذي يربط المملكة العربية السعودية مع مملكة البحرين لاستكشاف استخدام التقنيات المعاصرة لنمذجة النقل الحدودي. حيث يتناول هذا البحث توظيف هذه التقنيات للاستخدام في التنبؤ الدقيق لحركة المرور. إضافة إلى التحقق من فعالية استخدام هذه التقنيات في قضايا التخطيط على المدى الطويل مثل نمذجة اختيار وسيلة النقل. وقد تم اختبار مجموعات مختلفة من المتغيرات و التي تشتمل على مؤشرات أسواق الاسهم ، الاحوال الجوية و السفر الجوي للتنبؤ بحركة المرور. وقد استخدمت مؤشرات الاسهم للمرة الاولى للتنبؤ بحركة المرور عبر الحدود كمقياس بديل للاستقرار السياسي و الظروف الاقتصادية في البلدين. كما تم اختبار تأثير هذه المتغيرات على فترات زمنية سابقة مختلفة وعلى زمن التنبؤ المستقبلي. وقد اتضح من النتائج امكانية استخدام مؤشرات سوق الاسهم للتنبؤ بحركة المرور بين الحدود. ومع

ذلك، فإنه ينصح باستخدام العوامل الاخرى والمذكورة أعلاه للحصول على دقة أفضل في فترات زمنية اقصر تتراوح بين يوم و 7 أيام.

ويقارن هذا البحث بين دقة نتائج النموذج اللوغاريتمي (logit model) المستخدم في الشبكات العصبية الاصطناعية(ANN) والنماذج المدمجة. وقد تم اقتراح نموذج متنوع (غير متجانس)عن طريق دمج النموذج اللوغاريتمي مع الشبكات العصبية الاصطناعية لنمذجة اختيار وسيلة النقل. وقد اعطت هذه النماذج مجتمعة أداء أفضل من النتائج التي يعطيها كل نموذج على حدة. ولذلك فإنه ينصح باستخدام هذه النماذج المدمجة لنمذجة اختيار وسيلة النقل. وزيادة على ذلك، فإن عدد وسائل النقل المتوفرة والمتغيرات المدخلة يؤثر على دقة نتائج جميع أنواع النماذج.

CHAPTER 1 INTRODUCTION

Planning is the first step in any engineering work. It is especially important in transportation because transportation systems require large investments to build and/or maintain. Realistic and accurate forecasts of the future demand will decide the amount and method of investment in transportation improvements [1].

The characteristics that are found to be significant in modeling the transport demand can be classified into three groups; characteristics of trip makers, journey and transport facility [2]. The first group includes car ownership, household structure, income and employment status and familiarity with network conditions. Journey characteristics most often included in modeling are trip purpose and time of day. While characteristics of the transport facility include, but not limited to, relative times and costs for each mode of travel, level of service, safety, and parking regulations and availability which affect the transport demand [3]–[6].

Transport demand models are not easily transferable from one region to another. This can be attributed to the characteristics that are pertinent to specific regions, including, but not limited to; change in household trip rates and lengths, area characteristics, and level of service offered by the system. This fact has been established after examination of results in different areas. For example studies by Haire and Machemehl (2010) [6] and Wang (2009) [7]; on public transport mode choice, can be reviewed. Characteristics of trip-makers that determine their choices may also differ from region to region; these characteristics include cultural, socioeconomic and safety.

Generally, increase in automobile ownership results in increasing percentage of car trips. Households/persons who do not own a car rely more heavily on the public transit services such as buses or rails or para-transit services like taxi, as a means of transport. On the contrary, people with limited or no access to public transport are ‘auto-dependent’ which have to rely on private vehicles to reach their destinations. However, in such situations where travelers have more than one choice for mode of their transport, certain exogenous factors, such as the fuel prices and income, also influence their decision [7], [8].

Two different approaches are commonly used for modeling transportation data: statistical and Computational Intelligence (CI) which are also referred to as Artificial Intelligence (AI). The former one deals with making inference for a population by mathematically collecting, organizing and interpreting numerical data of a representative sample. Statistics, can be useful in providing insights on the mechanisms that created the data through its solid and widely accepted mathematical principles. The latter approach is a combination of learning, adaptation, evolution and fuzzy logic which are employed to create “intelligent” models. These models have the ability to create structure from an unstructured beginning (the data) [9].

With the growth of AI techniques, their applications have also become very popular in various fields of engineering. Artificial Neural Networks (ANNs), which is a popular branch of AI, are termed as universal approximators due to their generalization capabilities [10]. Recently ANNs are applied with ensemble learning methods to give higher accuracies than single ANN models [11], [12].

Considering the popularity of AI techniques in general and ANNs in particular, this research has investigated the use of this technique for traffic forecasting and mode choice modeling of border transport. For both types of models, ensemble learning approach has been used to improve their accuracy.

1.1 Study Area

The study area comprises of Dammam-Khobar metropolitan (KSA) and Bahrain. King Fahd Causeway provides the only land link for travelers from Dammam-Khobar (KSA) and Bahrain, which is a series of islands. The latest available statistics for traffic on the causeway indicates that the average daily traffic on the causeway in 2013 is about 10,000 in one direction. There is increasing trend in traffic demand on the causeway. This causeway not only provides connection for travelers but it is also the only road network connecting Bahrain to other Gulf countries. Since there is no railway connection between Bahrain and KSA, the road and air transport become very important for travelers to and from Bahrain. Many travelers from Gulf countries go to Bahrain through this causeway passing from Saudi Arabia.

Bahrain Airport also acts as a regional hub for connecting international flights to many countries of the world. So the travelers from the Dammam-Khobar region go to Bahrain by the causeway or airplane for their connecting flights. This is one of the reasons that Gulf airlines (national airline of Bahrain), operates its own bus service, in addition to frequent flights, between Dammam (KSA) and Bahrain. However, the bus service is not a popular mode for travelers of this route because it is slower and uncomfortable than the other two modes. The same has been observed during the data collection for this study.

1.2 Significance of Research

Border transport modeling is different from intra-region travel because of the inclusion of travelers from two countries which may have different characteristics and priorities. Moreover, consideration has to be given to the ties between the countries and economic and political conditions of each country which makes it more complicated. An indicator of political stability and economic conditions of the countries will be helpful to include in border traffic forecasting models. These conditions affect the trips attraction and generation of the region. Some studies have used gross domestic product (GDP) of the origin destination countries for modeling border travel demand [13], [14]. Similarly, other factors can also affect border transport whose relationship with border travel demand is not clear. In spite of these issues, border transport modeling is scarcely found in the present literature.

King Fahd causeway is a very important link across the border of KSA and Bahrain as explained in the previous sub-section. Currently this link faces problem of congestion on routine basis. In order to cater this problem, short as well as long-term strategies are needed. This study contributes to the short-term improvement of the causeway congestion by proposing appropriate variables and models for traffic forecasting. Travel demand can be better managed with accurate forecasts by dissipating demand. This study also contributes to the long-term travel demand management on the causeway by investigating the effects of two potential modes on the traveler mode choice behavior.

ANN and ensemble learning are two contemporary techniques which are becoming popular for research in different arenas including transportation modeling. However, the use of these techniques learning for modeling border transport is still under-explored, especially for mode choice modeling. So this research explores the ANN-based ensembles for traffic prediction and mode choice modeling.

In summary, it can be said that this study will be helpful in improving the quality of travel between KSA and Bahrain. Moreover, it also points out research horizons to be explored by other researchers for border transport modeling with contemporary techniques.

1.3 Objectives

The overall objective of this study was to develop traffic forecasting and mode choice models for modeling border transport by using appropriate variables, models and methodology. The specific objectives were as follows:

1. To develop models for short term (daily) as well as medium and long term (weekly, monthly and yearly) traffic forecasting using ANN ensembles and study the effects of different variable sets on model accuracy
2. To develop artificial neural network based and logit mode choice model for border mode choice modeling
3. To develop ensemble for border mode choice modeling and compare the performance of single models and proposed ensemble
4. To determine feasibility for potential new modes by calculating the realistic market shares including induced demand

1.4 Research Contribution

This study contributes in two important aspects of transportation planning; border traffic forecasting and mode choice modeling. Stock market indices are proposed for the first time to be used in traffic forecasting across borders. These indices will be used as a surrogate measure of the political and economic situation of the country which in turn affects its travel demand. Ensemble learning has been used for predictive traffic modeling which resulted in improvement in accuracy as compared to single models.

The use of ANN in general and ANN-based ensemble, in particular, is under-explored for border mode choice modeling. This study proposes a logit-ANN ensemble for mode choice modeling of border transport. This ensemble combines the inherent explanatory nature and generalization capabilities of ANN.

Regional parameters such as the factors affecting travel mode choice decisions on the area under study are also analyzed through travelers' survey and logit-regression models. Moreover, the effects of potential modes on the current market shares have been investigated.

This research inspires the use of ITS techniques such as traveler information systems, variable message signs in optimizing traffic flow and enhancing traffic management through accurate traffic predictions with online applications.

1.5 Research Methodology

The research is conducted in two dimensions addressing the short and long-term travel demand management of the study area. Modeling has been carried out for traffic forecasting and mode choice modeling using ANN-based ensembles. Figure 1-1 presents the complete description of the work done in this study.

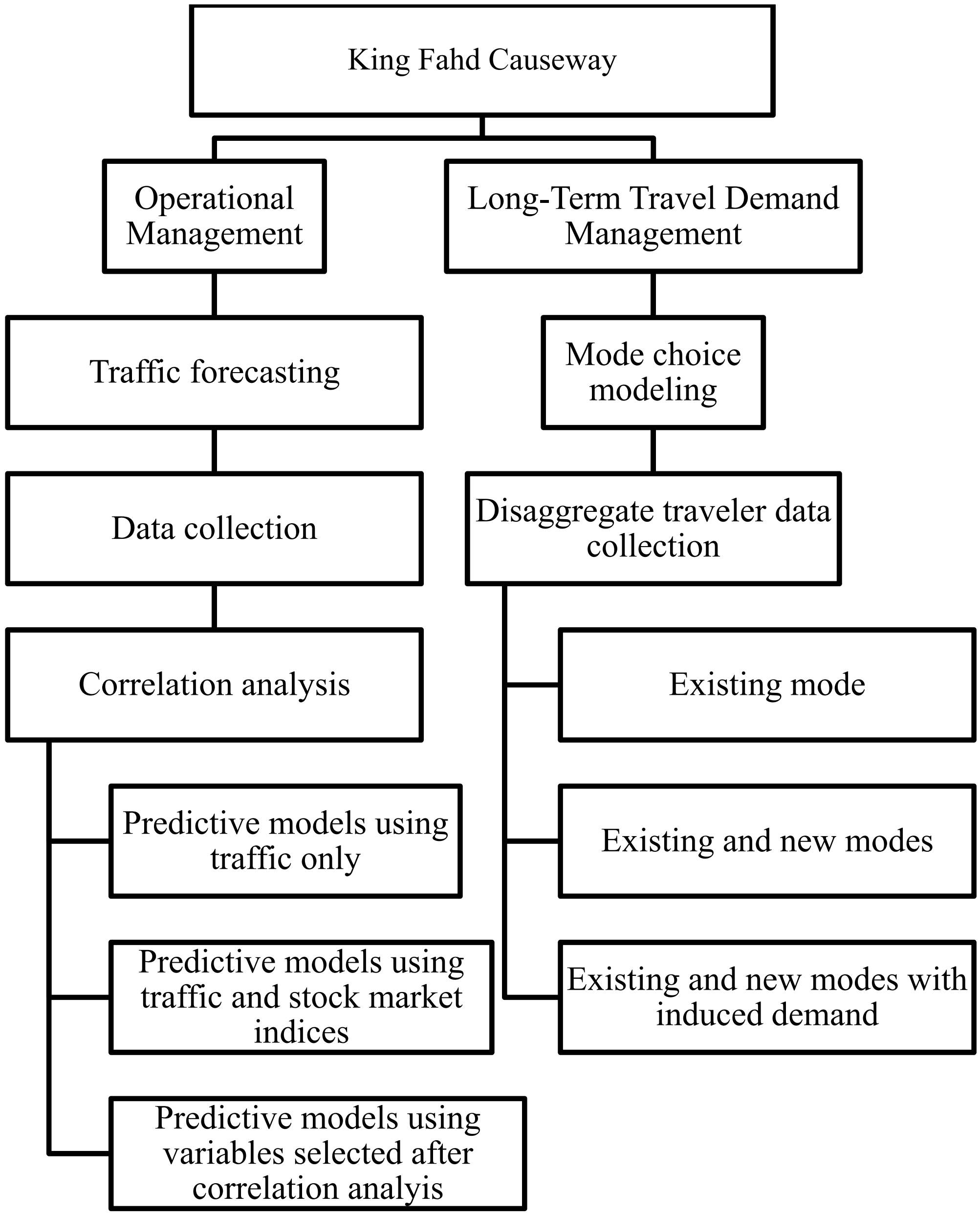

Figure 1-1: Research Methodology

Traffic forecasting was done using homogenous ANN ensembles. The traffic data was only available for daily basis so the forecasting models were developed accordingly. Mode choice models were developed using logit model, ANN and ensemble for comparison their accuracies.

A more detailed description of this methodology is provided in the subsequent chapters which are organized as follows. Chapter 2 presents a detailed review of the techniques used for traffic forecasting and mode choice modeling. Chapter 3 presents the analysis and discussion of traffic forecasting models. Chapter 4 includes mode choice modeling methodology and results. Lastly, chapter 5 summarizes conclusions of this research and also presents recommendations made on its basis.

CHAPTER 2 LITERATURE REVIEW

This research focuses on two aspects of transportation planning; traffic forecasting and mode choice modeling. In the proceeding sections, a review of the relevant literature in both areas is provided.

2.1 Traffic Forecasting

Traffic prediction involves forecasting traffic in terms of Annual Average Daily Traffic (AADT), Design Hour Volumes (DHV) and/or Directional Design Hour Volumes (DDHV). These traffic forecasts are used at the design stage to determine the geometric and pavement specifications. It is also required in the operation stage to investigate the efficiency of the roadway in terms of delay or level of service. The following factors must be addressed when traffic forecasts are to be made:

- Forecasting model
- Variable set and data collection
- Analysis year: both present and future
- Analysis period: hourly, daily, weekly, monthly, yearly

The data used for traffic forecasts may include; classified traffic counts, and traffic related factors such as directional, seasonal and peak hour factors and heavy traffic fractions.

Artificial Neural Networks (ANNs) is one of the contemporary methods used for traffic forecasting. ANNs have good generalization capabilities and work well with noisy data.

Because of these benefits ANNs are also becoming popular for traffic modeling. Their accuracy is better than statistical models. This has been proved by Ahmed and Gazder (2013) in which they compared performance of ANN with regression analysis [15].

2.1.1 Selection of Model

Application of a particular model in real time projects is sometimes dependent upon the adoptability of that model by the local organizations. In most of the cases, locally recognized models are preferred by the transportation authorities for planning purposes. However, the accuracy of the model is also an important issue and requires the planners and researchers to explore the forefronts of modern modeling techniques [16]. The subsequent subsections present an overview of the traditional and contemporary techniques used for traffic forecasting.

2.1.2 Traffic Forecasting Using Growth Trends

Traffic forecasts made using growth trends provide simplified, quick and fairly good estimate of long-term traffic demand (yearly or beyond). The historical trends of traffic demand can be established from one of the following resources:

- Historical Data
- Land-Use Management Systems (LUMS)
- Gas Sales Record

The data is plotted on a graph with AADT on y-axis and year of count on the x-axis. A least square regression analysis is done to establish the growth trends. Either the empirically developed regression equations are used or the graph can be extended manually to the future year for traffic forecasts [16] . This approach does not take in to account the change in land-use pattern or similar factors which would change the travel pattern.

2.1.3 Site Specific Models

Site-specific models are developed based upon the characteristics of the traffic generators and attractors. National Cooperative Highway Research Program (NCHRP) developed regression models for trip generation and attraction as a function of land-use, intensity and location. The factors included in these models are:

- Automobile ownership
- Income
- Household size
- Density
- Type of development
- Employment Density

One of the examples of the regression model given by NCHRP is given in equation 1 [18].

$$\text{HBW Attr.}_i = 1.45(\text{Total Employment}) \tag{1}$$

Where; HBW Attr. = Trips attracted towards zone 'i' for work from residents of other zones

Institute of Transport Engineers (ITE) developed their own rates for trip generation based upon the land-use and the number of units as per type of land-use such as residential, commercial [16].

All of the above models are linear and use a constant weight or coefficient for all the predictors. These coefficients may change based upon change in behavior of the people, government interventions or from one location to another. Therefore the spatial and temporal transferability of these models is highly questionable. Furthermore, these models do not take in to consideration the present traffic growth trend or seasonal variations.

2.1.4 Parametric and Non-Parametric Techniques

For accurate traffic forecasting parametric techniques such as linear and nonlinear regression, seasonal time series methods including autoregressive moving average (ARMA) have been employed since 1980s [19]. These techniques mainly assume the data to be part of a distribution or curve that can be represented by a fixed set of parameters such as mean, standard deviation, intercept, etc.

Some of these techniques, like Holt-Winters and ARMA, mainly present the future conditions as the repetition of present trends. Holt-Winters is a forecasting technique based on trended and seasonable patterns and the noisy data is distinguished by averaging the historical values. Some modifications have been made to these techniques to broaden their range by applying smoothening and filtering to the data. However, they are still based upon a linear combination of past values [20], [21].

With the growth in development of computational intelligence techniques such as fuzzy logics, machine learning and neural networks; non-parametric techniques are also being employed in this area by many researchers.

Non-parametric techniques mainly focus on identifying the past conditions which are similar to the conditions at the prediction time. They work on the principle of dynamic clustering unlike parametric techniques which try to fit the functions according to pre-defined structures. The application of these intelligent techniques has shown satisfactory results in traffic forecasting. But these techniques require rich quality of data for development of accurate forecasting models. The trade-off is made between the simplicity of traditional parametric techniques and the effectiveness of intelligent techniques for accurate predictions [22].

2.1.5 Application of Neural Networks for Forecasting

Single linear statistical algorithms may not be sufficient to deal with complexity of underlying process of inter-urban traffic flow. ANNs have received much attention in the area of traffic forecasting because of the ability to approximate any degree of complexity and work without prior knowledge of relationship between input and output variables. ANN is used to determine number of vehicles and temporal characteristics using the historical flow patterns and it is also suitable for non-linear and dynamic evolutions [23]. The basic architecture of ANN is given in figure 2-1.

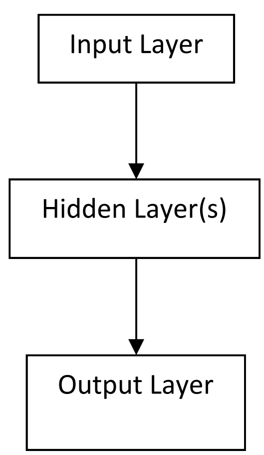

Figure 2-1: ANN Architecture

Each layer in ANN architecture consists of a number of processing units called neurons. The input layer has the same number of neurons as the input parameters. Hidden layers collect the input and feed it forward after summation and application of the chosen activation function. The output layer has the same number of units as the model outputs. The simplest type of ANN is also known as Multi-Layer perceptrons (MLPs) or feedforward backpropagation ANN because of their architecture [20], [24]. Other commonly used types of ANN are Radial Basis Function Neural Network (RBFNN) and Probabilistic Neural Network (PNN). will be discussed in the appropriate sub-sections with their applications. Few examples of the application of ANNs for traffic forecasting are described below.

RBFNNs consist of single layers of input, hidden and output. Input layer simply passes the input to the next layer, hidden layer calculates the proximity of the input vector to the centre of neuron and applies the activation function on it. While the output layer takes the weighted sum of these values coming from hidden neurons and gives the network outputs. Unlike these networks, MLP neurons apply the activation function to the weighted summation of the input vector.

PNN is specifically designed ANN for classification problems and consists of three layers: input, radial or hidden, and output layers. The radial units are equal to the number of training cases and are copied directly from the training data. Each unit in the hidden layer models a Gaussian function centered at the training case. Number of output units is equal to the number of classes to be predicted. Each output unit is connected to the radial units of its class, and have no connection with other radial units. Responses for each class are summed by the output units and normalized to give estimates of class probability [25].

Chen et al. (2001) [26] integrated the ANN and Autoregressive Integrated Moving Average (ARIMA) models. As a conclusion they found out that hybrid ANN approach gives more satisfactory results than all individual ARIMA models. They also checked the accuracy of RBFNN with backpropagation ANN and found out that the former approach gives slightly better results for short term traffic forecasting.

An ARIMA model can be represented by equation 2.

$$\phi(B)\Delta^d z_t = \theta(B)a_t \qquad (2)$$

Where $\phi(B)$, $\Theta(B)$ and a_t is a sequence of independent normal deviates with common variance, and $\Delta^d z_t$ is the d^{th} difference of the time series z_t [27].

A traffic dispersion phenomenon is very important for traffic forecasting, especially for signalized networks. The probabilistic approach to simulate dispersion can be very useful for traffic flows under ideal conditions. This approach may not work well in complex cases such saturated flow conditions because it is based upon strict mathematical rules. Qiao et al. (2001) [28] developed the ANNs to simulate traffic flows and their application includes queue prediction and dispersion, classification of highway traffic states and ramp metering. They found that ANNs have the ability to emulate all kinds of traffic variations provided that an appropriate structure is chosen. The observation of this study points to the fact that ANNs have good generalization capabilities which encourages its use for traffic forecasting and mode choice modeling.

Structure of ANN model in terms of number of neurons and other parameters affects its accuracy. Vlahogianni et al. (2005) [30] employed a genetic algorithm (GA) based optimization strategy to determine the appropriate neural network structure. GAs minimize a misfit function to achieve a combinatorial optimization of parameters of the target system [31]. Their study showed that combining the capabilities of a simple static neural network, with genetically optimized step size, momentum and number of hidden units, produces satisfactory results for modeling both univariate and multivariate traffic data.

A similar kind of study is carried out by Nagare et al. (2012) [32]. They used the Simulated Annealing (SA) algorithm with back propagation NN. They used the GA to elect, cross and mutate the synaptic weights and thresholds of ANNs. SA algorithm also decides the structure of ANN by identifying the number of input, output and hidden neurons. SA is a stochastic search tool which applies the same principle as annealing process of material physics. The error values for each ANN architecture are calculated and compared with the error value for previous architecture, the new architecture is accepted with a probability function if the error for the previous architecture of ANN model is greater than that of the new generated architecture [33]. The results from the above mentioned study show that this integration gives lesser Mean Square Error than the normal Back propagation NN. Hence it can be concluded that the accuracies of simple ANNs can be made better by integrating them with other techniques.

As an effort to reduce the iterations required by a normal backpropagation ANN algorithm, Tang et al. (2010) [32] used the Extended Kalman Filter (EKF) algorithm in combination with ANN. This algorithm gauges the weights as per the principle of minimum root mean squared covariance, which reduces the number of iterations required and makes it suitable for rapid dynamic calculations.

In light of the above mentioned studies, it can be concluded that ANN have generalization capabilities. However, developing an appropriate structure and algorithm for ANN is an important issue for its beneficial employment for modeling.

Gazder and Hussain (2013) [33] used ANN for prediction of daily traffic on the King Fahd causeway. They tried to find the optimal specification of MLP neural networks by trying different combinations of hidden neurons, number of iterations and learning rates. Their work was further improved by Ahmed and Gazder (2013) [15] by comparing the performance of the best ANN with regression analysis. They concluded that ANN has superior performance than linear regression but it also has less explanatory powers.

2.1.6 Exogenous Factors for Traffic Forecasting

Factors which are not part of the system under study are termed as exogenous factors. These factors may include alternate routes, weather, economic parameters, etc. The applications of such factors for traffic forecasting has been investigated previously which is presented in this section.

The effect of weather parameters on traffic demand has been proven by researchers like Tsirigotis et al. (2011) [34] and Yang (1997) [35]. In these studies, the authors have used the weather as one of the modeling parameters and have reported it to be affecting the modeling results. Traffic pattern also changes with respect to time as well, this theory has been supported by De Jong et al. (2003) [36]. They proposed and proved that travel demand is affected by time of the day.

Goodwin (1996) [37] has proved in his research that changes in availability and service of one alternative also affects the travel demand on the other. Another special consideration for border transport is the political and economic conditions of both connected countries which have to be taken in to account. Some of the studies like Peterson (2004) [13] and Rietveld (1999) [14] have used Gross Domestic Product as the economic indicator in their models. But this parameter is only suitable for long-term predictions as its value does not change on a short-term basis.

2.2 Mode Choice Modeling

Determining choice of transportation mode is an important part of transportation planning. Two different approaches are commonly used for mode choice modeling: statistical and AI. Statistical methods are widely used to understand the factors influencing choices of the users in many fields including transport modes. On the other hand, AI techniques are used to give improved accuracies of regression and classification problems [9].

The number of travelers making a mode choice (N_{auto}) and the market share for each mode (S_{auto}) is predicted using the equations 3 and 4.

$$N_{auto} = \sum_n P_{auto}(n) \tag{3}$$

$$S_{auto} = \frac{1}{N} \sum_n P_{auto}(n) \tag{4}$$

Where P is the probability of choosing auto mode, n is the number of cases in the sample choosing auto mode, N_{auto} is the predicted number of auto travelers, N is the total number of travelers including all modes, and S_{auto} is the market share of the auto mode [38].

The mode choice models were initially developed using the logit and probit techniques in 1960s and 1970s. In that era, mostly aggregate data was used for mode choice modeling. But since disaggregate data gives a better understanding of the users' behavior, use of disaggregate data in mode choice modeling started during early 1970s and it was adopted by many public transport agencies during the 1980s. Since then, it has taken over the use of aggregate data [38], [39]. Disaggregate data is considered effective for two reasons. First, it gives more focus on differences in tastes and preferences between individuals of the same zone or region, explaining their taste heterogeneity which have reportedly increased accuracies [1], [40]. Secondly, the models which are based on disaggregate data are easy to transfer for temporal periods and different geographic regions. Innovations and advancements in data collection along with the analysis techniques and equipments made the utilization of disaggregate data for mode choice modeling much easier for planners and policy makers [38], [41], [42].

The logit and probit techniques were used extensively till the late 1980s. By that time, researchers realized the potential of AI in terms of handling massive amount of data and nonlinear modeling. Methods like ANNs and Fuzzy Logics became popular in all fields of engineering including transportation modeling. In the early 1990s, the use of traditional models declined greatly, especially the probit models. Nevertheless, logit models are still in use today and they are being applied with different variations like mixed multinomial and latent class. From 1990s onwards, substantial research has focused on exploring the integration of different methods, including AI techniques like ANNs, to increase their performance.

The proceeding subsections present the review of mode choice modeling techniques in a chronological order as referred above. It is done so that the current trend in this area is defined and the future research potential is identified.

2.2.1 Traditional Techniques for Mode Choice Modeling

The traditional mode choice models are based on "utility maximization" theory. These models assume that a person making a particular choice from a set of alternatives will look for the best alternative to maximize the benefits (utility) he receives. These models are used for modeling many transportation related problems such as modeling choice of travel mode, time, route and others [43]–[45]. These models are also known as binomial or multinomial logit and probit models based upon the number of modes/choices incorporated in them. Table 2-1 summarizes the advantages and disadvantages of each type of statistical model.

Table 2-1: Main characteristics of traditional mode choice techniques

Model	Advantages	Disadvantages	References
Probit Model (1960s–1990s)	The utility equations contain an error term which is mode specific and has a general covariance structure Utility functions can be simplified because of the use of mode-specific error terms. Significance and elasticity of each variable can be easily investigated through statistical tests.	Extra computations required due to mode-specific component Lacks taste heterogeneity Large co-variance matrix, difficult to develop with a large number of variables or modes	McFadden, 1974 [46]; Keane, 1992 [47]; Yai et al., 1997 [48]; Ortuzar and Willumsen, 2002 [2]
Logit Model (1970s–1990s)	Less computations are required as compared to probit models. Utility functions are of the same nature as a regression model. Significance and elasticity of each variable can be easily investigated through statistical tests.	Offers less mode-related flexibility Fixed properties that have to be satisfied Lacks taste heterogeneity Difficult to develop with a large number of variables.	Foerster, 1979 [49]; Horowitz et al., 1986 [38]; Bhat, 1998a [43]; Suzuki, 2000 [5]; Hensher and Greene, 2003 [50]
Nested Logit (1990s till present)	Independence of Irrelevant Alternative (IIA) property does not apply Inclusion of exogenous factors and policy issues. Multi-dimensional analysis is possible	Pre-classification of data and gathering rich quality of data Shared unobserved attributes are linked with only one of the choice dimensions.	Bhat, 1998a, 1998b [43], [44]; Lee and Waddell, 2010 [45]; Yu and Sun, 2012 [51]
Mixed Multinomial and Latent Class Logit Models (1990s till present)	Use of random coefficients. More suitable to capture individual taste heterogeneity. More flexibility than Multinomial and Nested Logit Model. Accounts for every individual through random components of choice through distribution of coefficients.	Selection of distribution function for the coefficients. More computations than other logit models. Requires better quality of data than logit models. Latent class models require the pre-classified date	Cohen and Kocis, 1980 [7]; Horowitz et al., 1986 [38]; Ben-akiva et al. [52], 2002; Cantarella and de Luca, 2005 [53]

2..2.1.1 Probit Models

Probit models have also been very popular among the traditional techniques for transportation planning, especially the mode and route choice models [48], [54]. The response of making a choice can be predicted by using equations 5 and 6.

$$U_i = V\ (X_i, S) + \varepsilon_i \tag{5}$$

$$P_i = P(U_i > U_j) \text{ - For all available alternatives } j \text{ other than } i \tag{6}$$

Where U_i is the utility of the i^{th} mode, V is the systematic component of the utility function which comes from the weighted sum of the input variables, X is the vector of explanatory variables, S is the vector of coefficients having the same size as X, and ϵ is the random or error component of the utility function, P_i is the probability of choosing the i^{th} mode, j is the total number of modes in the model. Equation 6 depicts the probability of choosing mode 'i' (P_i). The significance of the parameters and the accuracy of the model is determined via different statistical tests like t-tests for various coefficient values [2].

The main attraction for using the probit models is that it provides a general covariance structure for the alternative specific error terms. The covariance matrix for a probit model must be set or assumed during the modeling process and its size depends upon the number of explanatory variables and modes [48], [55]. In some situations, probit models have been used to achieve satisfactory results, but Horowitz (1991) [55] argued in his article that efforts to gain marginal improvements in the probit models will result in much more computational costs. Another issue for using probit models is the correct specifications of the variables which affect its performance [47].

Probit models require lots of computations because of their alternative specific utility structure. This requires the calculation of the error distribution for each alternative mode.

2.2.1.2 Logit Models

The intensive calculation requirement inspired the shift from probit model to less calculation dependent logit models. This is the reason that logit models became very popular and are still in use [46]. The probability of a traveler to choose a particular mode in logit models can be expressed equation 7.

$$P_k^i = \frac{e^{V_{ki}}}{\sum_j e^{V_{ki}}} \tag{7}$$

Where,

P_k^i = Probability of trip-maker i choosing mode k out of j alternatives

V_{ki} = Utility of alternative k for trip-maker i,

= $\beta_0 + \beta_i X$; linear form

Utility functions of logit models are developed to optimize the log-likelihood parameter of the model which is given in equation 8 [56].

$$L = \sum_{m=1}^{M} \log(P_m, m) \tag{8}$$

Where (P_m, m) is the probability calculated by the model for the observed mode 'm'. The overall fit of the model is determined by comparing the chi- square statistic at specific probability (mostly 5%) with the double difference of log-likelihood parameter of model with intercept and full model which is denoted by -2LL.

Logit models have been used extensively for modeling mode choices by the transportation researchers. Such models are useful in understanding the choice-making phenomenon and the relative significance of the variables [38], [57].

2.2.1.3 Nested Logit Models

Logit models have the inherent restriction of Independence of Irrelevant Alternatives, which implies that the probability ratio between any two choices is independent of changes in utility of any other choice. Nested Logit (NL) models present an easy and convenient approach for adding or removing or changing utility of alternatives without Independence of Irrelevant Alternative (IIA) constraints [58]. Nested Logit (NL) models are used to examine the interrelation between transportation and other exogenous factors such as, residential mobility, departure time or petroleum prices in the modeling process. NL models are developed by classifying the modes or choices in different sets (called nests) based upon a particular characteristic such as public or private or access location. Figure 2-2 shows a NL model for residential mobility.

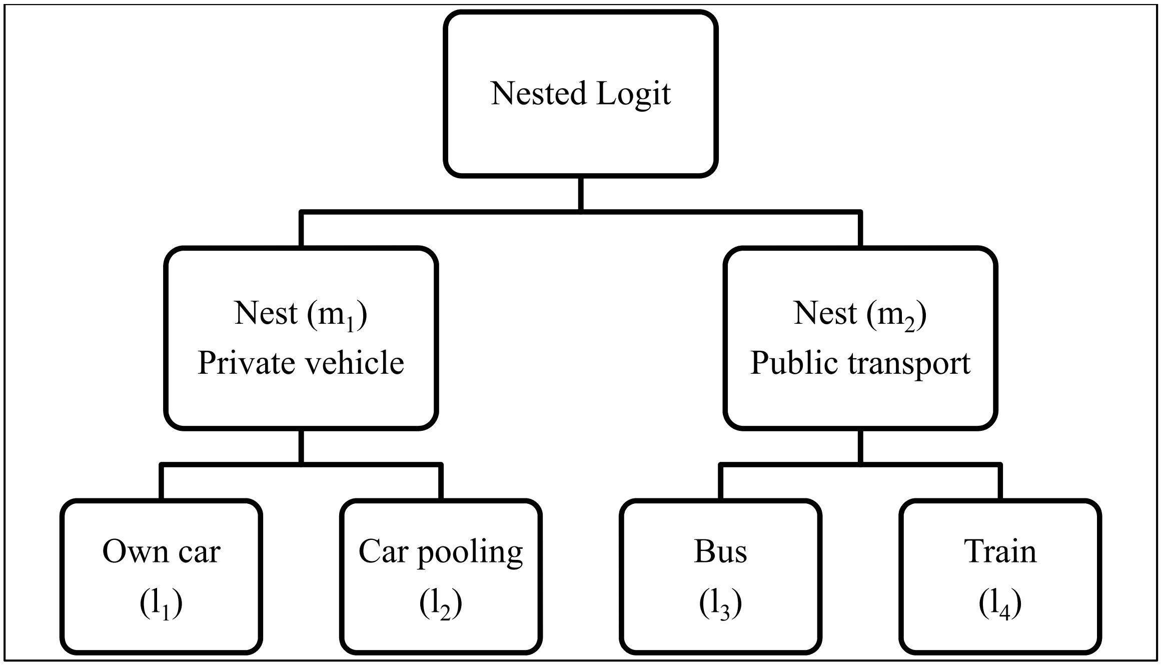

Figure 2-2: Nested Logit Model

The probabilities of choices are estimated using the equation 9.

$$P(l) = P(l \mid m) \cdot P(m) \tag{9}$$

Where P(l) represents the choice of the user to select mode 'l', given that the user has selected nest 'm'. Its conditional probability is calculated as per equation 10.

$$P(l \mid m) = \frac{e^{V_l \mu_l}}{\sum_{l' \varepsilon L_m} e^{V_l' \mu_l}} \tag{10}$$

Where V_l is the utility for mode 'l', μ_l is the scale parameter for 'l', and L_m is the total number of modes in nest 'm'.

The marginal probability of choosing nest 'm' is given by equation 11.

$$P(m) = \frac{e^{V'_{m'} \mu_m}}{\sum_{m' \varepsilon M} e^{V'_{m'} \mu_m}} \tag{11}$$

Where 'V' and 'μ' have the same meaning except that they are now for nest 'm'.

NL models account for the multi-dimensional nature of the transportation planning process and represent a step forward in the integration of transportation planning, land use and other social and economic factors. However, NL imposes an unnecessary restriction on the model in which the shared unobserved attributes can be associated with only one of the choice dimensions [45], [51].

2.2.1.4 Mixed Multinomial Logit Model

Mixed multinomial logit models can be defined as multinomial logit with random coefficients α drawn from a cumulative distribution function G (α, θ). The general form of these models is given in equation 12 and 13.

$$P_C(i \mid x, \theta) = \int L_C(i; x, \alpha) \cdot G(dx; \theta) \tag{12}$$

$$L_C(i;\ x,\ \alpha) = e^{x_i \alpha} \Big/ \sum_{j \varepsilon C} e^{x_j x} \tag{13}$$

Where C = 1…… J is the choice set; x = Variables included in the utility function; α = Random coefficient of variable; θ = Standard deviation of the distribution of α

The random parameter α is assumed to be arising from the taste heterogeneity of the decision makers and it is usually taken to be following a normal distribution function. This is the reason why mixed logit models are also known as the Random Coefficient Logit (RCL) models [50], [59].

The mixed logit models are considered to be the most accurate approach for discrete choice modeling currently available among statistical techniques. The researchers started utilizing them in the field of transportation in the late 1990s, as indicated by Bhat (1998a, 1998b) [43], [44] and McFadden and Train (2000) [59], and it has continued from then onwards. The main reason behind their introduction was that the mixed logit models align themselves much more with reality than most discrete choice models by adopting random coefficients. They also account for every individual having his own interrelated systematic and random components as they consider each alternative in its perceptual choice set(s). Ultimately, the forecasting power of the model is increased by considering the behavioral variability of the choice makers.

Researchers and practitioners are using the above mentioned models for revealed preference data, stated choice data and their combination. The continuing challenges faced with mixed logit models are derived mainly from the quality of the data. Mixed logit certainly demands better quality data than multinomial logit since it offers an extended framework to capture a greater amount of behavioral variability in choice making [51], [60].

Another variation to mixed logit models are the latent class models. These models assume random coefficients 'α' to be arising from a discrete distribution rather than a continuous distribution. The taste heterogeneity of the travelers is captured by the membership function of the discrete classes of utility function. This approach assumes that the population consists of 'Q' groups of individuals having fixed parameter set 'βq'. Greene and Hensher (2012) [60], in their study applied such model for freight transport routing. They observed that introducing random parameters within the groups of latent class models provides the best fit when compared to fixed parameter latent class, mixed logit model or multinomial logit model. Mixed and latent class models have been extensively applied because they do not have the restrictions of the standard logit and nested logit, thus offering more flexibility [51].

2.2.2 Artificial Intelligence Techniques for Mode Choice Modeling

Use of AI in transportation modeling has increased during the last two decades, due to its improved performance over traditional methods [61]. The major categories of AI techniques used for mode choice modeling and other transportation planning applications are the ANNs and Fuzzy logic systems.

2.2.2.1 Artificial Neural Networks

ANNs have been widely applied to various transportation problems, partly because of their abilities to be generic. Therefore they can easily be applied to any specific region or traffic conditions without major changes. They are used for accurately and conveniently simulating numerical model components. ANNs have been successfully used as an alternative to random utility theory for discrete choice modeling. In fact, they proved to be more successful than their traditional alternatives in some studies (logit and probit models) [62]. The other beneficial attribute of ANN for transportation planners include their ability to work with multi-dimensional data with large number of input and output variables. In this context, their primary usage has been for trip generation and distribution and modal split.

Artificial Neural networks are found to be suitable for transportation related problems due to their ability to work without prior information of the relationships between input and output parameters and to model nonlinear data. Because of such capabilities, ANNs are also considered suitable for dynamic modeling in conjunction with real-time information dissemination systems, for online application, which is the essence of modern Intelligent Transportation Systems (ITS) [4], [9], [63].

The most common types of applications that are used in transportation modeling are prediction, and classification [24], [53], [64]. Table 2-2 shows some application of the artificial neural networks in transportation choice modeling.

Table 2-2: Applications of artificial neural networks in transportation choice modeling

Application	Purpose [a]	References
Travel Mode Choice Modeling	FA, CLA	McFadden and Train, (2000) [56], Greene and Hensher (2012) [57], Ratrout and Rahman (2009([58]
Route Choice Modeling	CLA	Dia and Panwai (2007) [4], Bhat (1998b) [42], Yai et al., (1997) [46]
Trip Distribution Modeling/Forecasting	FA	Shir-Mohammadli et al., (2011) [60], Tortum et al., (2009) [65]
Predicting Household Automobile Choices	CLA	Shmueli et al., (1996) [69]
Analysis of Travel Behavior	CLA	Faghri and Hua (1995) [70]

[a]FA = Function Approximation, CLA = Classification

It should be noted that the accuracy of ANNs largely depends upon the richness of the training data. This data is fed to the ANNs for training themselves in understanding the phenomena or establishing the values of weights and biases [62], [64].

2.2.2.3 Integration of Artificial Neural Networks (ANNs) with Other Techniques

Artificial Neural networks can have the problem of over fitting to the training data, which reduces the accuracy of its prediction, due to being sensitive to initial weights and local minima. This problem is mainly due to the application of unclassified data to the system. One of the suggested treatments is to use a fuzzy interference with ANNs. Fuzzy logic systems are combination of fuzzifier with some fuzzy IF–THEN rules, fuzzy inference engine, and defuzzifier. The fuzzy IF–THEN rules are used to perform mapping from an input vector $x_T = [x_1,x_2, . . . ,x_n] \varepsilon R_n$ to an output $f \varepsilon R_n$. The basic flow of such systems is shown in Figure 2- [71].

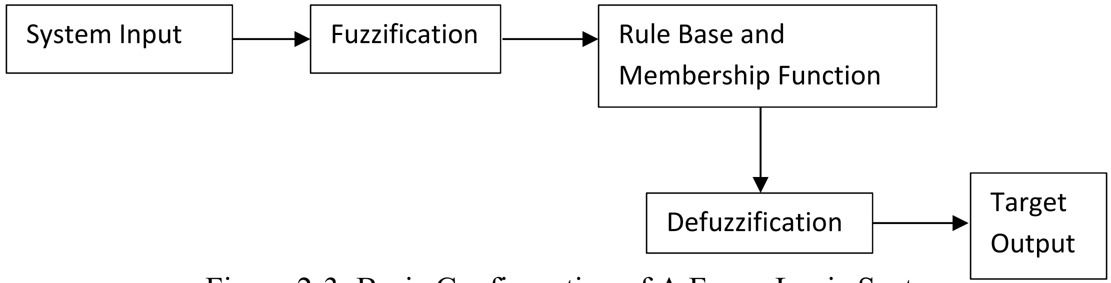

Figure 2-3: Basic Configuration of A Fuzzy Logic System

Using a Neuro-Fuzzy model also improves the flexibility and deals well with the uncertainty of human behavior. This model basically focuses on providing an appropriate architecture and suitable parameters for a target system with a set of input-output pairs [53], [67], [72].

Yin et al. (2002) adopted the above approach in order to reduce the processing time of network training and to improve the accuracy of prediction. They used fuzzy sets to make clusters of data sets that reduce the training time for ANNs. The model is divided into two modules: a Gate network (GN) used for fuzzy clustering and an Expert network (EN) which is composed of ANN to get the final forecasts [73].

This approach consists of two steps: representation of model by equivalent ANN and training of the ANN to minimize the error. The approach is helpful for transportation applications where travelers weigh their decisions on different parameters. In this case, ANNs are used to enforce fuzzy logic systems to enhance the understanding of travel behavior. Andrade et al. (2006) [72] used a similar approach to develop the utility functions for the Multinomial Logit Model. Their neuro-fuzzy multinomial logit (NFMNL) model provided better understanding and accuracy as compared with multinomial logit models.

RBFNN and Generalized Regression Neural Network (GRNN) are other types of ANN that are reported to give better accuracy than simple feedforward neural networks. GRNN works by drawing the function from the training data without any iterations and its accuracy is dependent upon the size of the training data.

Celikoglu (2006) [65] used ANNs to calibrate a model for individual traveler mode choice. The ANN was considered a sub-process for the nonlinear specification of utility functions, and used feedforward backpropagation neural network and multivariate linear regression (MVLR) analysis in addition to the above types for the calibration of a binary logit model for mode choice. In MVLR, a straight line best-fit equation is computed using the least-squares method in which squares of the errors are minimized for fitting the line to the data given.

Such studies have demonstrated that ANNs, irrespective of their type, give better results in discrete choice modelling than conventional techniques. Celikoglu (2006) [65] has shown that the selection of a particular technique cannot be done on single criteria. The networks with complex architecture and longer processing times such as radial basis neural networks are more accurate. On the other hand, a multivariate linear regression model is easy to develop and gives insight on the parameter elasticity of the model but it has a lower accuracy level. Hence, a trade-off has to be made between the complexity and modeling time and accuracy of the model.

Table 2-3: Examples of field applications of different mode choice modeling techniques

Technique	Study Area	Variables Used	Modes	Year	References
Artificial Neural Networks	Luxembourg	Location (place of work and residence); and related variables such as cost, availability, and abundance of public trans- portation. Travel	Private car, bus, train, cycling, or walking	2013	Omrani et al., (2013) [74]
Probit Models	Santiago (Chile)	Travel time, cost of travel, income of traveler	Car, sheared taxi, metro, bus, car-metro, sheared taxi-metro, bus-metro	1999	Bolduc (1999) [75]
Logit Model	Sydney, Canberra, and Melbourne (Australia)	Total cost (In-vehicle cost + time cost), travel time, terminal time, household income	Car, train, bus, plane	2012	Yu and Sun (2012) [51]
Nested Logit Model	Sydney, Canberra, and Melbourne (Australia)	Total cost (In-vehicle cost + time cost), travel time, terminal time, household income	Car, train, bus, plane	2012	Yu and Sun (2012) [51]
Mixed Multinomial Logit Model	Sydney, Canberra, and Melbourne (Australia)	Total cost (In-vehicle cost + time cost), travel time, terminal time, household income, choice specific constants	Car, train, bus, plane	2012	Yu and Sun (2012) [51]
Neuro-Fuzzy Model	Sapporo (Japan)	In-vehicle time, excess time, cost	Bus, subway, car	2006	Andrade et al. (2006) [72]
Multivariate Linear Regression	Istanbul (Turkey)	Number of trips for a set of origin and destination, mode specific time and costs	Cars, public transport	2006	Celikoglu (2006) [65]
Radial Basis Neural Networks	Istanbul (Turkey)	Number of trips for a set of origin and destination, mode specific time and costs	Cars, public transport	2006	[65]
General Regression Neural Networks	Istanbul (Turkey)	Number of trips for a set of origin and destination, mode specific time and costs	Cars, public transport	2006	[65]

The selection of a particular technique for a given scenario depends upon region, variables and modes considered in the model. Table 2-3 shows examples of field application of different mode choice modeling techniques from the recent literature. From the review of these examples logit models are still preferred in recent research because of their ability to depict the relationships between input variables and mode choices.

2.2.3 Border Mode Choice Modeling

Modeling behaviors for choosing travel modes across the border is a unique problem. It is mainly due to the involvement of people from different countries which have different backgrounds and behavioral characteristics [76]. This area of research is under explored and the literature is scarcely found on this topic.

A study worth mentioning in this regard is by Aljarad and Black (1995) [41]. They used the multinomial logit models for modeling mode choices of travelers between Saudi Arabia and Bahrain. These travelers are also the focus of this research. They developed models for two different corridors: Riyadh-Bahrain corridor and Eastern Province-Bahrain corridor. Riyadh is the capital of Saudi Arabia and it is approximately 400 km to the west of Eastern Province. They addressed the mode choices between car, bus and air travel in their model. They used disaggregate data for modeling and included socioeconomic and behavioral factors such as car ownership, decision time, etc. They proved the hypothesis that the factors influencing the mode choice of travelers from different origins are not the same in spite of having the same destination and purpose of travel.

Reggiani et al. (1997) [77] developed a binomial logit model and artificial neural network for modeling mode choices among the countries of European Union. They also presented a comparison between artificial neural networks and logit models based upon different statistics, most notable of which was Average Relative Variance (ARV), also commonly known as coefficient of determination (R^2). They reported that ANN model gives much accurate results than logit model. The equation 15 is used to calculate ARV which was applied on the aggregating the results of mode choice models i.e. flow of traffic for each mode.

$$ARV = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \tag{15}$$

Where,

y = Observed flow using mode 'i'

ŷ = Predicted flow using mode 'i', by the adopted model and

$\bar{y}$ = Average of the observed flow using mode 'i'

They pointed out the fact that ANNs are preferred for large samples of multidimensional data over logit models. But the use of ANNs to interpret causes of a certain behavior is limited as these networks only give the predicted values and not the relationships between variables. The selection of the appropriate architecture (including number of layers, hidden neurons, weights, etc.) and type of network is an important issue for getting valid results from ANN.

Some studies have also been found that have focused upon the aggregate mode choice demand. Rietveld (1999) [14] developed a multinomial logit model for calculating the total number of interregional trips between EU and non-EU countries. The model was based upon economic parameters of Gross Domestic Product (GDP) of origin and destination countries. GDP is the total market value of all the products and services produced within a country in a specific period. The focus of the study was on business trips between the countries. De Jong and Gunn (2001) [78] studied the effects of changes in fuel prices and travel time on car trips and public transport trips from historical data using traditional regression analysis. They used 15 years of data from 1985 onwards. The results of the study indicated that travel time has a bigger impact on the number of trips than travel cost for car travelers. It was also mentioned that the feedback regarding congestion would result in reducing the travel demand initially. This will make the travel conditions better in terms of travel time and may in turn induce more demand later.

Logit models have also been used in modeling air travel. Hensher et al. (2001) [79] modeled passengers' choice of airlines using multinomial logit model. They asked the travelers to state their choices for 32 fare regimes and then truncated the regimes to 4, 8, 16 and 24 and performed the survey again. They studied the change in accuracy of the model because of the design of questionnaire from more complex 32 regimes to simpler 4-regime data structure. The model results for different classification of fare regimes suggested that only a marginal improvement in the accuracy of the model is achieved by increasing the complexity of the data design.

Another interesting study was by Petersen (2004) [13], which dealt with the same type of study area as considered in this research i.e. the Oresund bridge between Sweden and Denmark. They studied the effects of different scenarios on the transportation demand going across the bridge. These scenarios were created using different assumptions of the values of GDP, infrastructure investment and demand forecasts of traffic. The model was created by nested logit model technique using the data collected before the bridge became operational. The aim of this study was to validate the models using the post-project data.

Table 2-4: Summary of Border Transport Modeling Techniques and Applications

Technique	Study Area	Variables	Modes	Year	References
Logit Model	Saudi Arabia - Bahrain	Monthly income, family trip (0 or 1), causeway aesthetic attractiveness, size of traveling group, Out of pocket cost, No. of departures by public carriers per day, decision time, automobile availability, out-of-vehicle time, in-vehicle time, weighted out of vehicle time (time multiplied by income)	Car, bus	1995	Aljarad and Black, 1995 [41]
Logit Model& Artificial Neural Networks	European Council Union member countries	Cost, travel time	Car, train	1997	Regianni et al., 1997 [77]
Logit Model	Europe Continent	GDP of origin and destination countries, dummy variables for ECU countries	Total number of trips	1999	Rietveld, 1999 [14]
Regression analysis	Europe	Fuel cost, travel time	Car, public transport	2001	de Jong and Gunn, 2001 [78]
Logit Model	Australia and New Zealand	Fare of alternative airline (centered and squared), dummy variables for membership in any promotion program, availability of limousine service, interaction of discounts and number of household adults, household income	Airlines operating between the two countries	2001	Hensher et al., 2001 [79]
Nested Logit Model	Denmark and Sweden (by Oresund Bridge Corridor)	GDP, origin and destination, generalized cost of travel between zones	Public transport, car, walk, bicycle, ferry/boat	2004	Petersen, 2004 [13]

Most of the studies found for border transport are related to politically or geographically connected countries like European Union, Australia-New Zealand, Scandanavian countries, and Saudi Arabia-Bahrain. From the review of all these studies, it can be said that logit model has been the predominant technique in this area. Another important observation is the inclusion of parameters like GDP, quality of service, decision time and border resistance factor in border transport models which are not included in inter-city mode choice models. Table 2-4 shows examples of border transport modeling techniques and applications.

Among the other variables, cost and time have been commonly considered in the studies; apart from that, a measure of economic condition has also been included in the form of GDP in most of the cases. Another important observation for border mode choice modeling is that logit models are more dominantly used in these studies as evident in table 2-4. One of the reasons behind the popularity of logit models is that the factors affecting the mode choice and their respective elasticities can be easily investigated with these models. Also ANNs are more preferred when the number of variables included in the model is large, which is not the case in most of the above studies.

2.3 Ensemble Learning

A new approach for using ANN is to use ensembles instead of a single network for modeling. This approach avoids the problems pertinent to ANNs such as dropping to local minima and over fitting to the training cases [12], [80]. Moreover, it is also useful in combining models of different types, architectures and input variables to get the desired output. Ensemble denotes a group of models with same output(s) which are developed in parallel or in series. The component models of an ensemble can differ in their training algorithms, input parameters or network architecture.

The approach has been successfully used in forecasting including internet traffic [81], video network traffic [82], economic parameters [80], air passenger traffic [83], traffic forecasting [84], [85]. Ensemble learning has also been used for time-series forecasting. Adhikari and Agarwal (2011) used homogeneous ensembles for time-series forecasting. They used weighted average method to calculate results for the ensembles and have reported to overcome the problem of over-fitting the training data [12].

Ensemble learning has also been used in short term traffic forecasting using per minute data by Chen and Zhang (2005) [11] and Chen and Chen (2007) [84]. Chen and Zhang used adaptive network based fuzzy inference system (ANFIS) to create ensembles. They concluded that the performance of the ensemble was better than the individual models. They acquired an accuracy range of from 10 to 20% [11]. Chen and Chen used RBFNN for ensemble development. They also investigated the effects of changing time resolution (scale of data), prediction horizon and number of input data points on the accuracy of the model. The accuracy range they acquired in their research was from 6 to 10% for predicting 16 minutes ahead using 80 minutes of traffic data. As a result of their research they concluded that increasing the number of input data points and the time resolution improves ensemble performance. While increasing the extent of prediction decreases the accuracy of the ensembles [84].

However, the use of ensembles for mode choice modeling is scarcely found in the literature. The most notable effort in this regard is by Nitsche et al., (2013) [86] who have combined AI technique hidden Markov model (HMM) with probabilistic technique for mode choice modeling. HMM is a model that takes an infinite number of possible sequences and each sequence is described by a finite probability distribution [87].

2.4 Summary of Literature

There are two types of techniques mainly available for transportation modelling; statistical and AI. Statistical techniques rely on the fixed mathematical foundations for modelling which requires sound knowledge of the problem before modelling. The popular statistical techniques for traffic forecasting include; linear and non-linear regression and ARMA. On the other hand, logit and probit models are commonly used statistical techniques for mode choice modeling. These techniques fail to handle the multi-dimensional and massive amount of data without increasing the calculations considerably.

AI techniques work without any prior information and have good generalization capabilities. ANN is the most commonly used AI technique used for transportation modeling. In spite of the benefits associated with the use of ANN, they are reported to over fit to training data and converge on local minima. In order to avoid these problems, hybrid models and ensemble learning are the two approaches used. Developing hybrid models requires in-depth knowledge, of all the techniques and methods used, for effective integration. Ensembles are more convenient to develop because only the outputs are combined without changing the framework of the technique(s). However, the application of ensemble learning for mode choice modeling is under explored.

Considering the advantages of ensemble learning and present trend of using ANN based ensembles, this research employs ensemble learning for traffic forecasting as well as mode choice modeling.

CHAPTER 3 TRAFFIC FORECASTING

In this chapter, development of traffic forecasting models for the incoming traffic from Bahrain to KSA through King Fahd causeway is presented. Moreover, a comparison has also been presented for different prediction horizons, input data points (look-back period) and variable sets.

The models for predicting traffic for different prediction horizons were developed using a range of 1 to 7 days as the input of time series data. This range was selected because daily traffic counts show a cyclic behavior from one day of the week to the same day of next week. For each prediction horizons different variable sets were used as the predictors. These variables included daily traffic, exogenous factors including daily number of flights and passengers arriving and departing from King Fahd international airport, weather (temperature and humidity), stock market indices for KSA and Bahrain and dummy variables for time-related factors such as weekday and summer vacations. A detailed description of these factors will be presented in the proceeding sub-section.

Traffic pattern changes with respect to time and may show peaks in certain periods [36]. Considering this fact, weekdays and yearly holidays have been used in this study as inputs to capture the change in daily travel demand on specific days.

As stated in section 1.1, King Fahd causeway provides a route for travelers from Dammam-Khobar metropolitan area (KSA) to Bahrain, air-travel is another alternate available for those travelers. Keeping this fact in mind, air-travel is also included in the predictive modeling. In order to consider the political and economic situation of the country, which affects its travel demand and pattern, stock market indices have been proposed to include in traffic forecasting models as the economic and political indicator.

The relationship between stock market indices on one hand, and political stability and economic prosperity on the other hand is established based upon the following premise. Political instability, especially when it is considered life-threatening, is expected to affect trip generation. Hence, it is acceptable to assume that border transport is affected by the political condition of the study area. Moreover, economic conditions have also been considered in the border transport modeling studies such as, Petersen (2004) [13], who used Gross Domestic Product (GDP) to predict traffic. Therefore, it can be claimed that border transport and stock markets are both affected by political stability as well as economic prosperity. Considering the fact that stock market indices are readily available, it was convenient and logical to use them as surrogate measure of political stability and economic prosperity which are difficult to quantify and measure in this study area.

In the following sub-sections, effects of adding exogenous factors, especially stock market indices, have been investigated in predicting traffic. Since King Fahd causeway is selected as the study area so the findings of this study are more relevant to cross-border transport.

3.1 Data Collection

The data for predictive modeling was collected from different sources which are given as follows. Traffic data for the causeway was planned to be collected for hourly basis from the causeway itself. But it could not be done because of difficulties in obtaining permission for data collection on the border area. Hence, the traffic data was provided by

Directorate general of Traffic in Bahrain which was handed over by Prof. Hashim Al-Madani from University of Bahrain.

Stock market indices were also considered important for the analysis. The daily stock indices for KSA stock market were collected from their official website (www.tadawul.com.sa). On the other hand, stock market indices for Bahrain stock market were provided by Prof. Al-Madani from Foud (2013) [88].

Weather parameters were collected for Dammam-Khobar metropolitan region (KSA) and Manama (Kingdom of Bahrain) through international weather monitoring website (www.tutiempo.net/en/Climate).

The common period for which all the above mentioned variables were available was from January 2003 to October 2013 which was utilized for traffic prediction model development. However, detailed data for longer periods of time will be beneficial for improving the accuracy of the predictive models.

3.2 Dataset

In this research; daily traffic counts on King Fahd causeway were used for the period ranging from 1st January 2003 to 21st October 2013. A complete description of variables of both datasets is given in Table 3-1.

Table 3-1: Summary of variables in dataset

Variable	Description
Variable A*	Dummy variable indicating period from 24th to 1st of each Islamic month (0 or 1)
Ramadan*	Dummy variable indicating period from 24th of 9th Islamic month to to 7th of 10th Islamic month (0 or 1)
Hajj*	Dummy variable indicating period from 1st to 14th of 12th Islamic month (0 or 1)
Name of Gregorian month	Name of each Gregorian month labeled by integers (1 to 12)
Summer vacations	Dummy variables indicating months of June and July (0 or 1)
Day of Gregorian month	Day of the month from 1 to 31
Name of weekday	Name of the weekday labeled by integers (1 to 7)
Daily stock index KSA	Daily index of Saudi stock market
Daily stock index Bahrain	Daily index of Bahraini stock market
Daily temperature KSA	Daily temperature recorded in Dhahran (KSA) in ^{0}C
Daily humidity KSA	Daily humidity recorded in Dhahran (KSA) in %
Daily temperature Bahrain	Daily temperature recorded at Bahrain international airport in ^{0}C
Daily humidity Bahrain	Daily humidity recorded at Bahrain international airport in %
Total traffic incoming	Total number of vehicles per day entering KSA through King Fahd causeway
Total traffic outgoing	Total number of vehicles per day exiting KSA through King Fahd causeway
Daily flights arriving	Total number of flights per day arriving at King Fahd international airport from Bahrain international airport
Daily passengers arriving	Total number of passengers per day arriving at King Fahd international airport from Bahrain international airport
Daily flights departing	Total number of flights per day departing from King Fahd international airport for Bahrain international airport
Daily passengers departing	Total number of passengers per day departing from King Fahd international airport for Bahrain international airport

*Shifting in Gregorian calendar each year

For the analysis, 19 variables were available including daily traffic counts of vehicles entering and exiting KSA, day and name of the month, name of weekday, daily temperature and humidity values in KSA and Kingdom of Bahrain, dummy variables for Hajj, Ramadan and summer vacations and daily stock market index for KSA and Bahrain.

Ramadan and Hajj are important religious occasions for Muslims occurring in 9th and 12th lunar month respectively, during these periods a 2 week holiday is announced in both countries. Similarly in summer vacations, educational institutes specially schools are closed. These variables were included based upon the assumption that some unique pattern of travel demand is expected to be captured in the holiday periods as people travel more for recreational purposes (KSA to Bahrain) and social visits (Bahrain to KSA).

Moreover, dummy variable 'A' represents the period in which salary is distributed to government employees in KSA at the end of each Islamic month. The Islamic month does not always coincide with the Gregorian month. The purpose of including this variable was to test whether people travel more for shopping and recreation when they receive their salary.

Daily stock indices are also included as indicators for political and economic stability of connected countries. The reason behind this inclusion is that overall stability of the country plays a role in its trips attraction and trip generation as mentioned in the studies discussed in previous section. Saudi Stock Exchange, also known as Tadawul, is the only stock exchange in Saudi Arabia. It is supervised by a government organization called Capital Market Authority. It lists 156 publicly traded companies (as of September 2, 2012). Bahrain stock exchange was established in 1987, it was converted in to a share-holding company in 2010 and named as Bahrain Bourse (BHB). There are 50 companies listed on BHB currently. It operates as an autonomous institution under the supervision of an independent Board of Directors, which is chaired by the Governor of the Central Bank of Bahrain.

The descriptive statistics for the numerical variables in the dataset is given in table 3-2. It can be observed from this table that the weather in Dammam-Khobar (KSA) and Bahrain are very close to each other. It is expected since both areas are only 25 Km apart. However, they were also considered for modeling and similar to other variables their relevance to travel demand was examined by performing correlation analysis. The correlation analysis is presented in the next section.

3.3 Correlation Analysis

The first step was to determine the variables from the available set that have statistically significant correlation with the traffic coming in to KSA. Since traffic on opposite directions are highly correlated to one another so this correlation analysis can also be considered valid for traffic going out from KSA. It is expected that these variables would be useful in understanding traffic pattern and predicting travel demand. For this purpose, a correlation analysis was performed using the Pearson correlation coefficient (r) given by the equation 13.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\left(\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)\left(\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\right)} \tag{15}$$

Table 3-2: Descriptive statistics of variables in dataset (January 2003 – October 2013)

	Mean	Minimum	Maximum	Standard Deviation
Daily Total traffic (vpd)	18522	1897	37033	5324.12
Daily temperature Bahrain (^{0}C)	27.4	10.3	39.6	6.82
Daily humidity Bahrain (%)	56.1	14	95	12.86
Daily temperature KSA (^{0}C)	27.51	7.8	40.7	7.87
Daily humidity KSA (%)	46.85	10	96	17.52
Daily stock index KSA	7656.79	2500	19600	3158.72
Daily stock index Bahrain	1704.38	1035.3	2902.68	526.16
Daily flights arriving	2	0	14	1.43
Daily passengers arriving	190	0	1402	141.2
Daily flights departing	2	0	16	1.52
Daily passengers departing	181	0	1968	147.77

Only those variables were selected for further analysis that have correlation coefficient statistically different from zero at 5% probability. The results of the correlation analysis are given in table 3-3. On the basis of correlation analysis, 13 variables were used in the predictive models. It should be noted that most travel-related, like opposite side traffic and alternate traffic mode, have the higher correlation coefficient values. Moreover, weekdays also have correlation coefficient statistically different from zero and thus considered to be affecting travel demand. Humidity of Bahrain has the highest correlation coefficient among the weather parameters. This may be due the fact that temperature is normally same on both sides of the border. Moreover, weather of Bahrain becomes more important as majority of travelers go to Bahrain for recreational purpose.

Furthermore, the hypothesis that stock market indices can be used as the economic and political indicator is also justified as their correlation index is statistically different from zero at 5% probability. The dummy variables for salary disbursement period, Hajj vacation period and summer vacations also have correlation coefficients greater than zero at specified probability. This means that travel demand changes during these periods of vacations and religious holidays. Ramadan vacations did not have correlation coefficients statistically different than '0' which is expected due to the fact that Muslims fast during part of this period and may not like to travel. The negative sign of the coefficients show that travel demand increases when the value of variable decreases as can be seen from humidity correlation coefficients.

Table 3-3: Correlation analysis for dataset with Traffic Incoming to KSA (2003 – 2013)

Variable	Correlation Coefficient	P-value	Significance at 5% probability
Variable A	0.05	0.004	Yes
Ramadan	-0.03	0.074	No
Hajj	0.08	0.000	Yes
Name of Gregorian month	0.00	0.834	No
Summer vacations	0.06	0.001	Yes
Day of Gregorian month	0.00	0.621	No
Name of weekday	-0.33	0.000	Yes
Daily stock index KSA	0.12	0.000	Yes
Daily stock index Bahrain	0.13	0.000	Yes
Daily temperature KSA	0.03	0.128	No
Daily humidity KSA	-0.01	0.664	No
Daily temperature Bahrain	0.02	0.348	No
Daily humidity Bahrain	-0.12	0.000	Yes
Total traffic incoming	1.00	0.000	Yes
Total traffic outgoing	0.63	0.000	Yes
Daily flights arriving	0.33	0.000	Yes
Daily passengers arriving	0.32	0.000	Yes
Daily flights departing	0.33	0.000	Yes
Daily passengers departing	0.30	0.000	Yes

3.4 Methodology

The modeling was done in three steps, also described in figure 1-1, to investigate the effects of adding stock market indices and other variables in the traffic forecasting models as inputs. In the first step, only traffic was used as the predictor, in the next step traffic and stock market indices were used as inputs to the predictive models. The final step was to use all the variables that were selected based upon correlation analysis for modeling as mentioned in section 3.2.

In each step, predictive models were developed for daily, weekly, monthly, quarterly and yearly prediction horizons using the time series data of input variables. These prediction horizons were selected because input variables, including traffic counts, salary disbursement period, weather parameters and/or stock market indices, show cyclic behavior during these periods. Therefore, this study also explored the extent to which these parameters are effective in forecasting traffic. For each horizon, time series data from 1 to 7 days were used i.e. seven models were developed for each prediction horizon. This range was set based upon the fact that traffic peaks are repeated weekly. The outputs of these models were combined in an ensemble for the final predictions. The ensemble was a separate ANN model which takes the output of each component model as the input. The process of ensemble development is depicted in figure 3-1.

All the models and ensembles were developed using 10-fold cross-validation process and MLPs. Section 2.1.5 gives a brief description of MLPs. This type of ANN has been successfully employed for predictive modeling including time-series forecasting as well [20], [21]. 10-fold cross-validation means that the sample was divided in to 10 parts and each part was used as the test set for different models of the same architecture. The performance of the model with that specific architecture is given as the average performance on the 10 test sets. This process takes care of any sampling bias for model development that may occur due to very large or small values being selected as the training set. It has been reported to give realistic measure of model performance [89]. Relative Mean Square Error (RMSE) was used as the metric of performance. It is calculated as shown in equation 16 [90].

$$RMSE = \frac{\sum_{i=1}^{n}\left(\frac{(Actual_i - Prediction_i)^2}{Actual_i^2}\right)}{n} \tag{16}$$

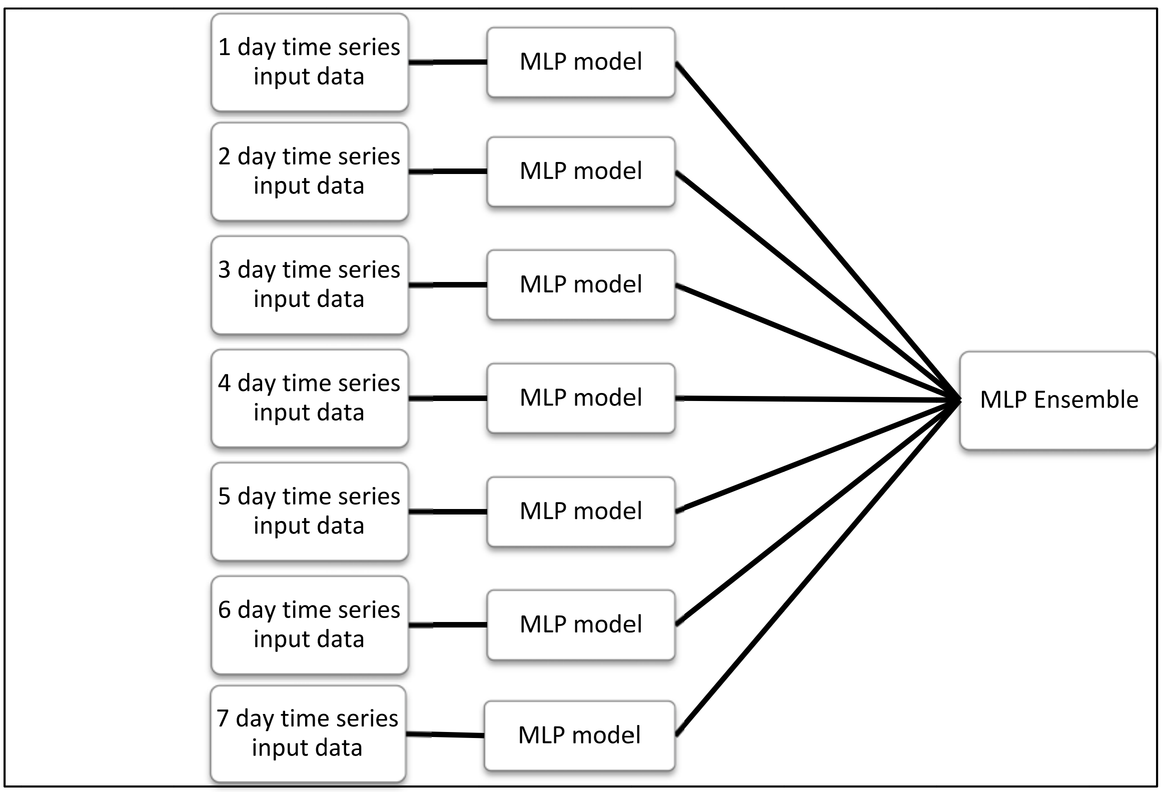

Figure 3-1: ANN Ensemble Development

Before beginning the steps of predictive modeling, optimum architecture of ANN was found using various iterations. RMSE values were compared all iterations using different architectures of ANN and the optimum architecture was selected based upon the lowest RMSE value. The best models extracted from these iterations are shown in table 3-4.

Table 3-4: Configuration of best models

Model	Steps Used as Input	Variables Used
5 hidden neurons, hyperbolic function	1	Incoming traffic only
4 hidden neurons, logistic function	2	
4 hidden neurons, hyperbolic function	3	
4 hidden neurons, hyperbolic function	4	
6 hidden neurons, logistic function	5	
8 hidden neurons, hyperbolic function	6	
7 hidden neurons, hyperbolic function	7	
9 hidden neurons, logistic function	Ensemble	
3 hidden neurons, hyperbolic function	1	Incoming traffic and stock market indices
7 hidden neurons, logistic function	2	
9 hidden neurons, hyperbolic function	3	
11 hidden neurons, hyperbolic function	4	
8 hidden neurons, hyperbolic function	5	
9 hidden neurons, logistic function	6	
8 hidden neurons, hyperbolic function	7	
9 hidden neurons, logistic output function	Ensemble	
7 hidden neurons, hyperbolic function	1	All factors having correlation coefficients statistically higher than '0' at 5% probability
7 hidden neurons, hyperbolic function	2	
5 hidden neurons, hyperbolic function	3	
5 hidden neurons, hyperbolic function	4	
6 hidden neurons, hyperbolic function	5	
8 hidden neurons, hyperbolic function	6	
16 hidden neurons, logistic function	7	
9 hidden neurons, logistic function	Ensemble	

It can be observed from the above table that a fixed number of hidden neurons or activation function cannot be used for different number of input variables. Also increasing the number of neurons with the increase in variables is not valid. The maximum number of hidden neurons needed for the forecasting models was 16, which was used for yearly prediction horizon with maximum number of input parameters. Less than 10 hidden neurons gave the best RMSE values in most of the cases.

3.5 Predictive Modeling Results and Discussion

The results from the predictive models in terms of RMSE are given in table 3-5. From review of this table few trends can be observed. The accuracies of the models decrease with the increase in prediction horizon which is shown by comparing row 3 with 12 or 21. On the other hand, accuracies generally improve with the increase in time series data input as observed from row 3 to 9 or 12 to 19. The use of ensemble learning proved to be beneficial in improving the accuracy of predictions.

Another important observation is that the addition of stock market indices improved the accuracy of the models from 1 to 4%, so their use in border traffic demand modeling is applicable. Accuracies of the models improved marginally upon the addition of other significant variables. Figure 3-2 shows the comparative improvement in accuracies of the ensembles for each prediction horizon when different sets of variables are used.

Table 3-5: Root Mean Square Error (RMSE) for predictive models

Steps Used as Input	Using Only Daily Incoming Traffic	Using Daily Incoming Traffic and Stock Market Indices	Using All Significant Variables
Daily Prediction			
1	0.050	0.048	0.018
2	0.048	0.043	0.017
3	0.033	0.033	0.017
4	0.036	0.033	0.017
5	0.031	0.031	0.017
6	0.032	0.030	0.017
7	0.019	0.018	0.017
Ensemble	0.018	0.019	0.012
Weekly Prediction			
1	0.100	0.096	0.095
2	0.097	0.096	0.090
3	0.097	0.088	0.085
4	0.095	0.088	0.082
5	0.097	0.088	0.080
6	0.092	0.087	0.075
7	0.092	0.075	0.075
Ensemble	0.090	0.066	0.060
Monthly Prediction			
1	0.174	0.176	0.144
2	0.174	0.168	0.144
3	0.172	0.167	0.140
4	0.173	0.162	0.138
5	0.174	0.161	0.135
6	0.175	0.158	0.132
7	0.171	0.151	0.134
Ensemble	0.134	0.130	0.129
Quarterly Prediction (3 months)			
1	0.193	0.189	0.169
2	0.174	0.174	0.167
3	0.176	0.170	0.160
4	0.175	0.166	0.158
5	0.175	0.163	0.155
6	0.177	0.160	0.154
7	0.172	0.157	0.150
Ensemble	0.165	0.151	0.150
Yearly Prediction			
1	0.209	0.210	0.171
2	0.208	0.180	0.169
3	0.204	0.180	0.168
4	0.204	0.168	0.165
5	0.199	0.181	0.162
6	0.190	0.180	0.162
7	0.189	0.164	0.161
Ensemble	0.188	0.148	0.145

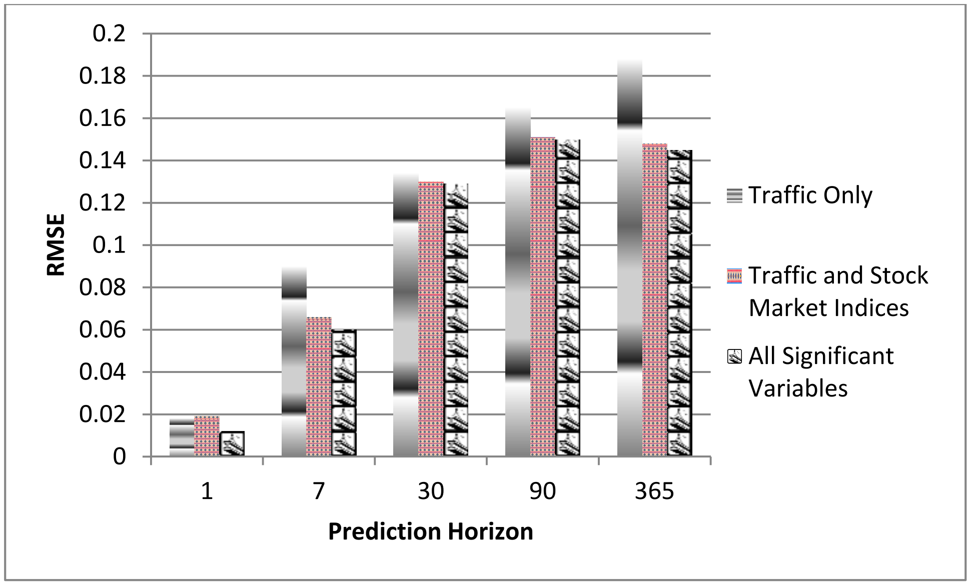

Figure 3-2: Comparison of Ensembles

Addition of only stock market indices affects the accuracies of the models more significantly when the prediction horizon is longer than 1 day. However, if other significant variables are added than much significant improvement is observed for daily predictions. So it can be concluded that for daily predictions, it is better to use all significant variables while for longer prediction horizons use of traffic with stock market indices can give comparable accuracies without the use of any other predictor. It is expected because cyclic nature of traffic is easily captured in long-term predictions while daily predictions are more random in nature and require more explanatory power in terms of input variables.

Another important trend is that ensemble accuracies, with only traffic data as input, degrade more rapidly with increase in prediction horizon from 1 to 7. While other ensembles' performance degrade more rapidly when prediction horizon changes from 7 to 30. This is the reason that all ensembles give same accuracy for 1 month prediction.

It can also be concluded that MLP with 5 – 9 hidden neurons using hyperbolic function give best RMSE values for most of the cases. Ensembles were also found to give better accuracies than all the single models in almost all the cases.

CHAPTER 4 MODE CHOICE MODELING

Mode choice models for travel between Dammam-Khobar metropolitan area (KSA) and Bahrain are presented in this chapter. A complete description of data collection process has also been given alongwith the guidelines used for data collection. Different types of models have been developed, namely; logit, ANN and ensemble. The accuracy of these models has been compared for different number of input variables and mode choices. The nature of border transport differs from an intra-city or intra-region transport as discussed before in section 2.2.3. Consequently, mode choice models for inter-city or intra-city travel cannot be easily applied to border transport. Therefore, mode choice models for border transport are scarcely found in the literature. Majority of the models found thus far in the literature are using the traditional technique of logit models.

4.1 Identification of Research Problem

It was decided to collect disaggregate data through a questionnaire interview survey, because it is more effective in capturing travelers' behavior [38]. The questionnaire encompassed most of the variables considered important by other researchers in this area. From the review of studies mentioned in table 2-4, it has been observed that use of ANNs and ensembles in this area and their comparison to logit technique is under-explored, so this issue is further investigated in this research. Previous studies in this area neglected the effects of causeway congestion on the travelers' choice and induced demand of any new mode (like ferry or rail). This research takes both these factors in to account. Moreover, Aljarad and Black (1995) [41], who also covered the route between Khobar-

Dammam (KSA) and Bahrain for mode choice modeling, did not collect travelers' data from Bahrain which was considered in this study.

4.2 Data Collection

A questionnaire was formed to seek information about the traveler and their trips between Dammam-Khobar (KSA) and Bahrain. Traveler information included nationality, occupation, car-ownership, and previous traveled to/from Bahrain (Yes or No), whereas trip information covered, purpose, frequency, mode, time and cost of travel. Moreover, questions to capture the behavior of the user for selecting one of the existing modes (car, airplane and bus) or switching to another potential mode; ferry or train, were also part of the interview. Ferry is a small ship in which people can travel with their cars and have mobility within Bahrain. This mode can be perceived as entertaining by some travelers while seen as unsafe by other.

An important consideration for the data collection process was the number of travelers to be interviewed by the survey. Sudman (1976) [91] suggested that the sample should be large enough to encompass minimum of 100 elements in each major category of the population. So this rule was used as a guideline for sample collection. The major categories considered in this study and the number of responses for each of them is given in table 4-1. The travelers for the bus mode was not considered for mode choice models because it was found from the survey results that the bus riders are captive blue-collar travelers.

First of all, an online pilot survey was conducted in a focused community comprising mainly of university students, faculty and staff. It served as a useful pilot survey. 100

respondents submitted their responses most of which were car travelers. This pilot survey was used for improving the questionnaire to be used ahead. In this survey, the respondents were also asked to give their preferred values for travel time and cost for potential modes of train and ferry service. The average value of travel time and cost of train were 40 minutes and 40 SAR respectively. While the average travel time and cost for ferry were 60 minutes and 50 SAR respectively. These values were used in further survey as the travel cost and time of these services.

Table 4-1: Statistics of Major Categories

Category	Options	Total
Nationality	Saudi	245
	Non-Saudi	409
Car Ownership	Yes	448
	No	206
Previously Traveled to Bahrain	Yes	516
	No	137
Mode of Travel	Car	409
	Airplane	107

It was decided to conduct the survey through random sampling in several shopping malls of Khobar-Dammam metropolitan area of KSA and Bahrain on different days and times. This approach was helpful in collecting the responses from people who have not traveled on this route but may travel team when train or ferry service is started in future. The people who have not traveled previously to Bahrain were included in the analysis to calculate the induce demand for the potential modes (train and ferry). This was done based upon the assumption that any new mode or service may induce new demand because of better or different characteristics than existing modes. Hence it was also observed in the survey that people who have not traveled on this route before were interested to use it, if new modes are offered.

A team of 5 people was employed in KSA while a team of 3 people from Bahrain side was collecting data. The malls nearest to the causeway in both countries were selected for this purpose. The samples were collected randomly in the same time frame in KSA and Bahrain. Different teams were deployed in both countries consisting of people who can communicate in multiple languages, including Arabic, English, Urdu, Hindi and Bengali. Arabic and English are the languages mainly spoken and taught in both countries, while others are useful to communicate with the large community of Asian expatriates. The questionnaire was developed in English as attached in appendix A.

The sample data was monitored and analyzed on daily basis to ensure that at least 100 responses for each major category are collected [91]. The survey was continued until this objective was reached which resulted in collection of more than 654 responses. Among which 485 were from Dammam-Khobar area (KSA) and 170 were from Bahrain. The ratio of responses from both countries was consistent to the ratio of population of the two adjoining areas of the causeway which is 0.29 (Bahrain to Eastern province KSA).

Currently, there are three modes operating currently between Khobar-Dammam (KSA) and Kingdom of Bahrain; car, bus and airplane. From the pilot survey and the data of actual survey, 23 responses were collected for the bus service. It was also found that car and airplane are the modes commonly considered for traveling on this route. Therefore 100 samples for each of these modes were also ensured. Bus is the least preferred mode and mostly taken by blue-collar labor class people which is arranged by their employers to take transit flight from Bahrain. Therefore, bus travelers are mainly captive riders and its inclusion in the mode choice modeling was not considered beneficial. The modeling for the present situation was performed for a binomial choice problem with car and airplane as the available choices.

The collected data presented a good mix in terms of nationality, occupation and car ownership. The data revealed that recreation, shopping and catching flight are the main reasons for traveling to Bahrain opted by 31%, 27% and 17% respectively.

Travel time, cost and privacy were indicated as the main reasons for selecting their mode by approximately 65% of the travelers. Other reasons for mode choice were mobility, convenience and safety. Car travelers represented approximately 80% of the responses and the rest were mainly airplane travelers. Majority of the respondents said that they would be willing to shift to train or ferry, if one of these modes is started in the future, as shown in table 4-2.

Table 4-2: Statistics of Collected Data for New Modes

Category	Options	Responses from Bahrain	Responses from KSA	Total (%)
Decision to Shift to Train	Yes	116	404	510 (78%)
	No	52	92	144 (22%)
Decision to Shift to Ferry	Yes	61	317	378 (58%)
	No	109	167	276 (42%)

4.3 Modeling Methodology

The modeling process was accomplished in three phases. The first phase was to develop models for existing situation i.e. modeling the mode choice of the travelers who have traveled to Bahrain previously using one of the existing modes (car and airplane). The bus mode was not included in modeling at this stage because of captive nature of its riders and difficulties in data collection. The second phase was to model the mode choice of these travelers after the inclusion of new modes i.e. train and ferry. For this purpose, first train and then ferry service was included in the model as the third mode. Lastly, models were developed to include existing as well new travelers (people who do not have any travel experience to Bahrain but may travel in future attracted by new modes). Table 4-3 presents complete description of each phase.

Mode choice models were developed using both ANNs and logit model. The general form of logit model is given in equation 7 in sub-section 2.2.1.2. A comparison was made between the accuracy of both techniques. Lastly, ensembles were created by developing ANN models from the probabilities calculated by logit model for mode choices. The flow of this logit-ANN ensemble is depicted in figure 4-1.

Table 4-3: Description of Mode Choice Modeling Phases

<table>
<tr><th>Phase</th><th>Travel modes included in the analysis</th><th>Travelers included in the analysis</th></tr>
<tr><td>Phase – I</td><td>Car and airplane</td><td>Existing travelers</td></tr>
<tr><td rowspan="2">Phase – II</td><td>Car, airplane, and train</td><td>Existing travelers</td></tr>
<tr><td>Car, airplane and ferry</td><td>Existing travelers</td></tr>
<tr><td rowspan="2">Phase – III</td><td>Car, airplane and train</td><td>Existing and new travelers</td></tr>
<tr><td>Car, airplane and ferry</td><td>Existing and new travelers</td></tr>
</table>

The dataset was divided in to two parts; for training and testing of the models respectively. The train set consists of 75% of the samples. Accuracy was used as the performance measure of the models which is calculated as given in equation 17.

$$Accuracy = \frac{CP}{TP} \tag{17}$$

Where CP is the number of correct predictions and TP is number of total predictions. For each model; accuracies for total predictions and predictions for each mode was calculated [86]. The accuracy of each model was calculated by taking the average of 10 runs of model development. In each run, different samples were randomly selected for training and testing of the models (logit, ANN and ensemble). This was done to avoid sample biasness which may result in inconsistent results.

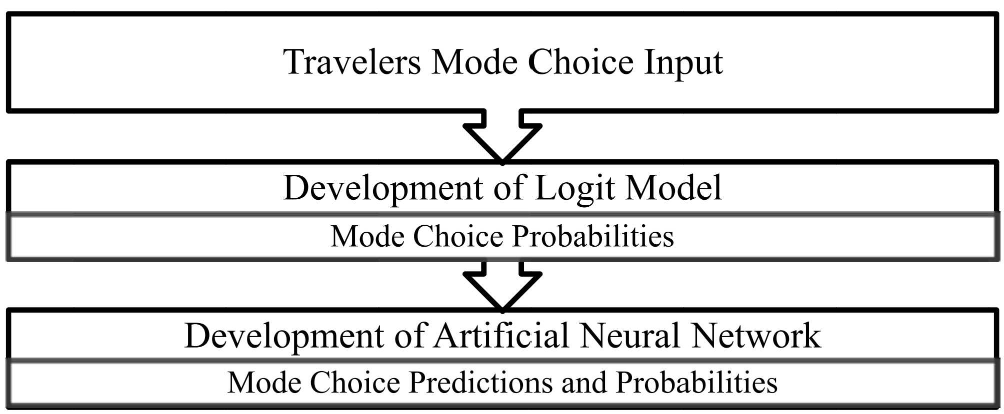

Figure 4-1: Logit-ANN Ensemble

4.4 Modeling Existing Situation – Phase I

A list of variables available for modeling is presented in table 4-4. The first column of the table shows the category of the variable. 516 survey responses were used to model this scenario.

The architecture for ANN was selected after several iterations of changing number of neurons and network type. Three types of ANN were tested; MLP, RBFNN and PNN. All these types are explained in section 2.1.5.

Table 4-4: Variables for Modeling - Phase I

Category	Variable	Scale Type
Traveler information	Nationality	Discrete (1 to 5)
	Age	Continuous
	Occupation	Discrete (1 to 12)
	Car ownership	Discrete (0 or 1)
Travel information	Travel time	Continuous
	Travel cost	Continuous
	Trips per year	Continuous
	Frequency of traveling alone	Discrete (0 for never, 1 for sometimes, 2 for always)
Purposes of travel	Recreation	Discrete (0 or 1)
	Shopping	
	Catching flight	
	Social visit	
	Business	
	Education	
	Resident	
	Health	
Reasons for choosing the selected mode	Time	Discrete (0 or 1)
	Convenience	
	Privacy	
	Mobility	
	Cost	
	Avoiding causeway congestion	
	Safety	
	Other	

Table 4-5 presents the results for the accuracy of all models and ensemble on the test set. The logit model was developed using airplane as the base mode and it is represented in equation 18. Its log-likelihood parameter was 189.96 which is greater than chi-square distribution value at 5% probability for 9 degrees of freedom. The variables for logit models were filtered after comparing their coefficients t-statistics with standard t-distribution at 95% probability i.e. 1.96 and dropping the variables with illogical signs. The t-statistic for each coefficient is given as subscripts in equations 18 and 19 with coefficients. The change in base mode did not change the accuracy of the logit model.

$$U_{car} = 1.40 + 0.15(TY)_{3.75} - 1.18(FLT)_{-2.81} + 0.84\,(REC)_{2.85} - 0.81(FTA)_{3.42} + 1.14(P)_{1.97} - 4.12(CONG)_{-3.24} - 0.003(TC)_{-1.85} - 0.001(TT)_{-2.14} \quad (18)$$

Where,

- U_{car} = Utility of car mode
- TY = Trips/year
- FLT = Catching flight as the trip purpose
- REC = Recreation as the trip purpose
- FTA = Frequency of traveling alone
- P = Privacy as the purpose of travel
- CONG = Avoiding congestion as the reason for mode choice
- TC = Travel cost
- TT = Travel time

It can be observed from the utility function that frequency of travel, flight and recreation as the purpose of travel, causeway congestion, privacy, travel cost and time are important factors for attraction or rejection of mode choices. Travel cost and time had lowest coefficient values in the utility function as compared to other factors.

In order to expedite the improvement in accuracies of logit models, an ensemble was created using the probabilities calculated by the logit model. Model presented in equation 18 was used for calculating modal probabilities since equation 18 and 19 give the same accuracies. These probabilities were fed as input to an ANN network in the second step.

From table 4-5 it was found that ANN gives the least accuracies for overall predictions as well as predicting cars. Accuracies of ensembles and single logit model were close in terms of overall predictions. Single logit models were used to calculate the market share for each mode using equations 2 and 3 mentioned in section 2.2. Table 4-6 shows the number of travelers and market share for each mode for Phase I.

Table 4-5: Accuracy for Models - Phase I

Model	$Accuracy_{Total}$	$Accuracy_{Car}$	$Accuracy_{Airplane}$
ANN	0.80	0.80	0.63
Logit Model	0.86	0.89	0.74
Logit-ANN ensembles	0.85	0.94	0.62

Table 4-6: Number of Travelers and Market Share for Each Mode

Mode	Car	Airplane
Number of Travelers	416	100
Market Share	0.81	0.19

It can be observed from table 4-7 that car is the most preferred mode of travel. This observation is consistent with the results of the survey.

4.5 Modeling with New Modes – Phase II

The second phase in planning was to model the change in existing travelers' behavior if a new mode is introduced. Two modes were proposed in this regard; train and ferry. Each mode was tested in turn to test which one has a greater effect on the travelers' mode choice. The travelers were asked whether they will choose to go by ferry or train given a specified fare and travel time. These parameters were calculated from the pilot survey done online in which the respondents were asked to suggest cost and time for these services. In this phase, more variables were used in modeling as the respondents were also asked to give reasons for rejection of new modes if they do so.

4.5.1. Effects of Train Service

Table 4-7 gives a complete list of variables which were used for modeling to analyze the effects of adding train service to the existing scenario. After changing the mode choices for the travelers according to their responses for the new mode, it was found that almost all the travelers who travelled by airplane previously now opted for the new train service. There were only 9 respondents who still opted for airplane which are not enough observations for modeling. So it can be assumed that airplane will become obsolete after the introduction of the new train service. Therefore, the models in this phase are built for two modes only; car and train using 507 observations.

Table 4-7: Variables for Modeling - Phase II with Train Service

Category	Variable	Scale Type
Traveler information	Nationality	Discrete (1 to 5)
	Age	Continuous
	Occupation	Discrete (1 to 12)
	Car ownership	Discrete (0 or 1)
Travel information	Travel time	Continuous
	Travel cost	Continuous
	Trips per year	Continuous
	Frequency of traveling alone	Discrete (0 for never, 1 for sometimes, 2 for always)
Purpose of travel	Recreation	Discrete (0 or 1)
	Shopping	
	Catching flight	
	Social visit	
	Business	
	Education	
	Resident	
	Health	
Reasons for choosing the selected mode	Time	Discrete (0 or 1)
	Convenience	
	Privacy	
	Mobility	
	Cost	
	Avoiding causeway congestion	
	Safety	
	Other	
Reasons for not choosing new mode (If applicable)	Lack of mobility	Discrete (0 or 1)
	Travel time	
	Lack of privacy	
	Rigid schedule	
	High fare	
	Unsafety	
	Other	

The utility function calculated for logit model was calculated using car as the base mode (equation 19). The log-likelihood parameters (-2LL) for this model was 166.73, which satisfies the chi-square test. Hence the logit model was statistically fit for mode choice modeling in this phase.

$$U_{train} = 11.006 - 3.84(TSCH)_{-4.80} - 0.20(TC)_{-6.67} \quad (19)$$

Where;

- U_{train} = Utility of train mode
- TSCH = Flexible schedule for rejecting train service
- TC = Travel cost

Moreover, travel cost and flexibility in schedule will affect the attraction for train service negatively. However, utility of train is more sensitive to rigidity of schedule as compared to travel cost since the former has higher coefficient value.

Table 4-8 gives the test accuracies of the models developed for this scenario. As observed in section 4.3, both logit model and ensemble give higher accuracies than single ANN models. Moreover, the accuracies of ensemble and logit model are practically the same.

Table 4-8: Accuracy of Models - Phase II with Train Service

Model	$Accuracy_{Total}$	$Accuracy_{Car}$	$Accuracy_{Train}$
ANN	0.92	0.87	0.93
Logit Model	0.95	0.96	0.95
Logit-ANN Ensemble	0.97	0.97	0.97

At this stage, ensemble is used for calculating the market share. This is because it shows consistent accuracies, for overall predictions as well as predicting car and airplane, as compared to logit model. Table 4-9 gives the number of travelers and market share for car and train service in the new scenario.

Table 4-9: Number of Travelers and Market Share - Phase II with Train Service

Mode	Car	Train
Number of Travelers	115	392
Market Share	0.23	0.77

It can be observed that introduction of train service will attract a major share of the car mode. In addition to that, it will also change the mode choice of almost all the present airplane travelers as indicated by the survey responses.

4.5.2. Effects of Ferry Service

Another potential mode that can be employed as an alternate to present modes is a ferry service. Table 4-10 gives the list of variables used for analysis of this scenario. Mobility was not included as one of the reasons for not choosing the ferry service, because people can travel with their cars in a ferry. The analysis in this situation was carried out for 3 modes; ferry, car and airplane and 516 responses were used.

Table 4-10: Variables for Modeling - Phase II with Ferry Service

Category	Variable	Scale Type
Traveler information	Nationality	Discrete (1 to 5)
	Age	Continuous
	Occupation	Discrete (1 to 12)
	Car ownership	Discrete (0 or 1)
Travel information	Travel time	Continuous
	Travel cost	Continuous
	Trips per year	Continuous
	Frequency of traveling alone	Discrete (0 for never, 1 for sometimes, 2 for always)
Purpose of travel	Recreation	Discrete (0 or 1)
	Shopping	
	Catching flight	
	Social Visit	
	Business	
	Education	
	Resident	
	Health	
Reasons for choosing the selected mode	Time	Discrete (0 or 1)
	Convenience	
	Privacy	
	Mobility	
	Cost	
	Avoiding causeway congestion	
	Safety	
	Other	
Reasons for not choosing new mode (If applicable)	Travel time	Discrete (0 or 1)
	Lack of privacy	
	Rigid schedule	
	High fare	
	Unsafe	
	Other	

The utility functions calculated for logit model using airplane as the base mode are given in equations 20 and 21. In this case, a multinomial problem had been encountered with 3 modes, so there were two utility functions calculated for two modes. The probability of the third mode was calculated as the remainder from 100%. The log-likelihood parameter for logit model was 449.512 which is higher than standard chi-square value at 5% probability.

$$U_{car} = 3.48 + 0.43(TY)_{4.22} - 1.56(FTA)_{-4.32} - 0.01(TC)_{-5.07} - 0.01(TT)_{-3.00} \quad (20)$$

$$U_{ferry} = 9.29 + 0.41(TY)_{4.28} - 0.97(FTA)_{-2.62} - 0.08(TC)_{-8.52} - 0.01(TT)_{-3.64} \quad (21)$$

Where;

- U_{car} = Utility of car mode
- U_{ferry} = Utility of ferry mode
- TY = Trips per year
- FTA = Frequency of traveling alone
- TT = Travel time
- TC = Travel cost

It can be observed from equations 20 and 21 that utility of car is more sensitive to traveling alone while utility of ferry is more sensitive to travel cost. So it can be said people traveling alone are more likely to take ferry as compared to car. Travel time and trips per year have the approximately same effect on car and ferry utilities.

Table 4-11 presents the accuracies of the models developed in this stage. ANN model show the lowest accuracies and ensemble have higher accuracies than logit model and ANN for overall as well individual mode predictions.

Table 4-11: Accuracy of Models - Phase II with Ferry Service

Model	$Accuracy_{Total}$	$Accuracy_{Car}$	$Accuracy_{Ferry}$	$Accuracy_{Airplane}$
ANN	0.83	0.81	0.87	0.53
Logit Model	0.87	0.89	0.87	0.80
Logit-ANN Ensemble	0.89	0.92	0.89	0.82

The probabilities generated by the ensemble are used to calculate market share for car, airplane and ferry service as shown in table 4-12. It can be observed that after the introduction of ferry service to the existing situation, it will become the most preferred mode for the travelers. However, many travelers would still prefer to use their own car for travel. The market share of airplane will also reduce with the introduction of this mode.

Table 4-12: Number of Travelers and Market Share - Phase II with Ferry Service

Mode	Car	Ferry	Airplane
Number of Travelers	191	279	41
Market Share	0.37	0.54	0.08

4.5.3. Effectiveness of Train and Ferry Service

A comparison of market shares between the existing situation and after the introduction of new modes train and ferry service is presented in figure 4-2. The market share of airplane after the introduction of train is indicated as '0'. This is because almost all the existing travelers preferred the new mode in their response of the survey.

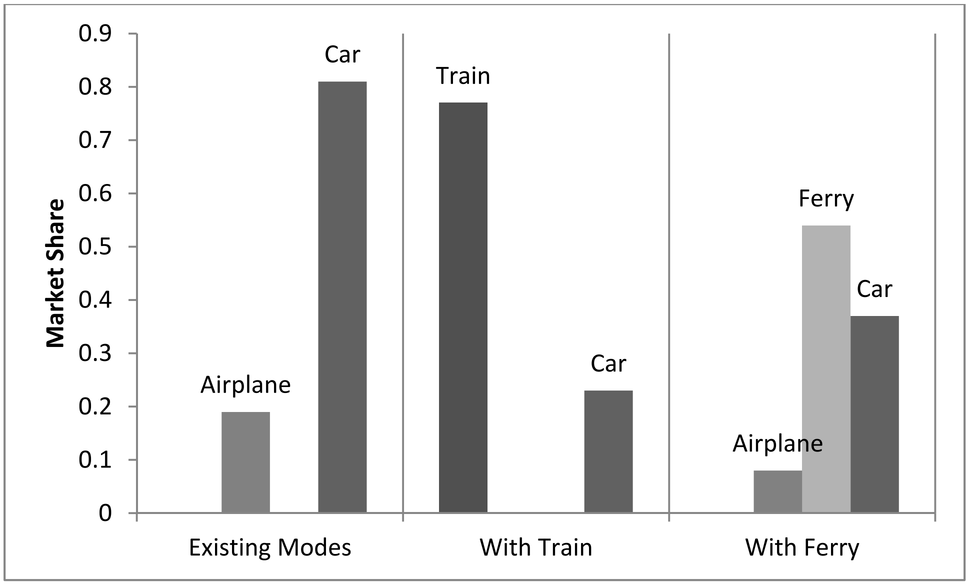

Figure 4-2: Comparison of Market Shares

Present airplane travelers are more flexible in terms of shifting their mode than present car travelers. However, train service will attract more travelers of both car and airplane as compared to ferry service.

Almost all the surveyed present airplane travelers (except 9) have preferred train service over their existing mode. Moreover, train service attracts a larger portion of present car travelers as compared to ferry service. Car mode keeps the second largest share with the introduction of any new mode.

4.6 Modeling Existing and New Travelers with New Modes – Phase III

It is often observed as result of a road development project, that due to higher level of service, it also attracts the travelers who were not using that route before [37]. Similarly, introduction of a new mode can induce new travelers to consider traveling between Bahrain and KSA who presently do not like the existing modes. So it is essential to take in to account this induced demand for determining realistic market shares of any new mode. Keeping this factor in mind, the survey also covered the people who do not have any prior experience of traveling to Bahrain/KSA. They were asked if they would like to consider traveling between KSA and Bahrain in the future if a train or ferry service is started.

Mode choice models were recalibrated with the inclusion of these potential travelers, who do not have any prior experience of using the existing modes. These travelers will be referred to as new travelers henceforth. The number of variables used in this situation also reduced since some of them were not applicable to new travelers like number of trips, frequency of traveling alone, reasons for choosing existing mode and others. Therefore, only traveler information and travel cost and time will be used in the modeling. The list of variables used in this phase is given in Table 4-13.

Table 4-13: Variables for Modeling - Phase III

Category	Variable	Scale Type
Traveler information	Nationality	Discrete (1 to 5)
	Age	Continuous
	Occupation	Discrete (1 to 12)
	Car ownership	Discrete (0 or 1)
	Previous travel to Bahrain	Discrete (0 or 1)
Travel information	Travel time	Continuous
	Travel cost	Continuous

4.6.1. Effects of Train Service after Inclusion of New Travelers

The mode choice models were calibrated for two modes only; car and train. As it was observed in section 4.4.1 almost all the existing airplane travelers preferred train over their existing mode of travel. The new travelers do not add any share for the existing modes so airplane mode was still not considered for modeling. There were 625 responses used for modeling this scenario.

The utility function estimated for train is given in equation 22 using car as the base mode. The log-likelihood parameter for logit model was calculated to be 204.39 which is higher than chi-square at 5% probability for 1 degree of freedom.

An observation from logit model analysis was that cost is the most important trip-related utility parameter which affects the utility of train negatively. This factor is important for all travelers; whether new or existing.

$U_{train} = 12.43 - 0.23(TC)_{-7.14}$ (22)

Where;

- U_{train} = Utility of train mode
- TC = Travel cost

Table 4-14 presents the accuracies of the models developed for inclusion of train service with new travelers. Single logit model and ensemble give equal overall accuracies for this binomial choice problem which his higher than ANN model.

Table 4-14: Accuracy of Models - Phase III with Train Service

Model	$Accuracy_{Total}$	$Accuracy_{Car}$	$Accuracy_{Train}$
ANN	0.92	0.32	0.94
Logit Model	0.97	0.97	0.99
Logit-ANN Ensemble	0.97	0.96	0.99

Ensemble was used to calculate the market shares which are shown in table 4-15. The market share of the train service has increased from 0.77 to 0.78 by the inclusion of new travelers so it can be said that 1% new demand will be attracted to this route if train service is started.

Table 4-15: Number of Travelers and Market Share - Phase III with Train Service

Mode	Car	Train
Number of Travelers	114	511
Market Share	0.22	0.78

4.6.2. Effects of Ferry Service after Inclusion of New Travelers

Similar to phase II, section 4.4.2., the models were developed for predicting modal choices among car, airplane and ferry using 613 responses. The logit utility functions were estimated for ferry and car and given in equation 23 and 24 using car as the base mode. The log-likelihood parameter was estimated to be 564.60 which is greater than chi-square distribution at 5% probability for 3 degrees of freedom.

$$U_{ferry} = 8.78 + 0.51(CAR)_{1.04} - 0.08(TC)_{-8.34} - 0.001(TT)_{-3.57} \quad (23)$$

$$U_{car} = 1.66 + 1.59(CAR)_{4.01} - 0.004(TC)_{-4.19} - .001(TT)_{-2.42} \quad (24)$$

Where;

- U_{ferry} = Utility for ferry mode using airplane as the base mode
- U_{car} = Utility for car mode
- CAR = Car ownership
- TT = Travel time
- TC = Travel cost

Car ownership, travel time and cost are the most important factors for modal utility in this situation. Travel time and cost negatively affect the utilities of car and ferry while car ownership has positive effect on their utilities. Utility of ferry is more sensitive to car ownership and travel cost as compared to car. Hence, car owners are more likely to travel by car.

Table 4-16 presents the accuracies of the models developed with new travelers after the introduction of ferry service. Accuracies of predicting airplane and car are lower as compared to ferry because the former have lesser number of responses. Moreover, logit model is under-fitting to the predictions of airplane which has the least number of observations and over-fitting to ferry which has majority of observations. Hence logit model has worst accuracy for predicting airplane and best accuracy for predicting ferry which is practically feasible. Ensemble shows consistently high accuracies for overall and individual mode predictions, so it is more reliable than logit model. ANN has the least overall accuracy in this case which is similar to the observation of table 4-15.

Table 4-16: Accuracy of Models - Phase III with Ferry Service

Model	$Accuracy_{Total}$	$Accuracy_{Car}$	$Accuracy_{Ferry}$	$Accuracy_{Airplane}$
ANN	0.69	0.54	0.77	0.64
Logit Model	0.84	0.65	0.99	0.32
Logit-ANN Ensemble	0.86	0.83	0.88	0.76

Market shares were calculated using probabilities given by ensemble and are presented in table 4-17. It can be observed that airplane becomes the least dominant mode in this situation and ferry has the majority of marker share.

Table 4-17: Number of Travelers and Market Share _ Phase III with Ferry Service

Mode	Car	Ferry	Airplane
Number of Travelers	191	382	40
Market Share	0.31	0.62	0.06

4.6.3. Increase in Demand for Train and Ferry Service

Comparison of market shares of train and ferry service before and after the inclusion of new travelers is presented in figure 4-3. From this figure, it can be observed that the increase in market share of the new mode by the inclusion of new travelers is 1% and 8% for train and ferry service respectively. Since new travelers add more to the market share of ferry service than train service so it can be said that ferry is more attractive to the new travelers as compared to train.

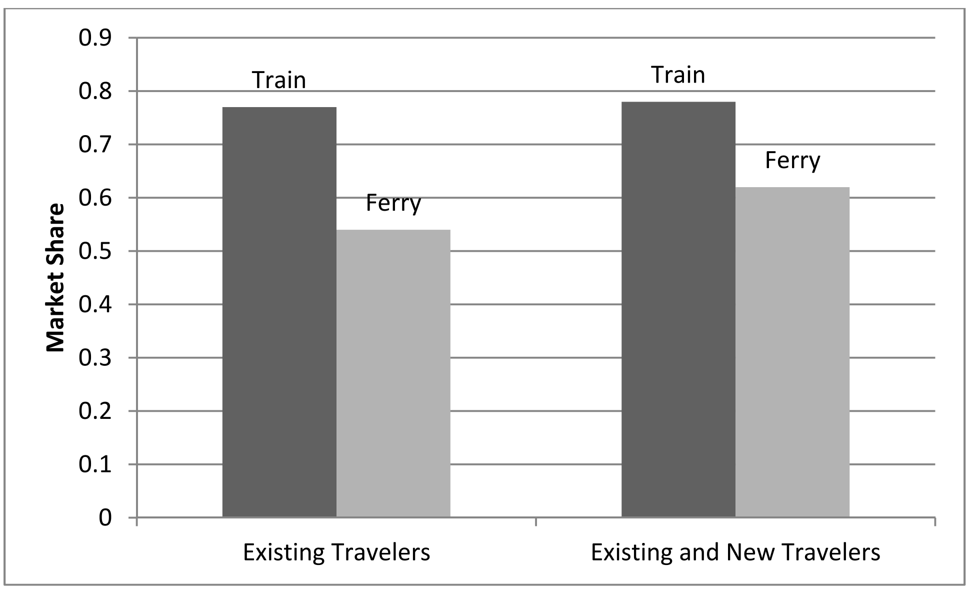

Figure 4-3: Change in Market Shares by New Travelers

However, even after this increase, overall market share of ferry service is less than that of train. It is because train service is more effective in diverting the existing travelers to itself irrespective of the mode chosen presently by them. Hence, effectiveness of train service for this route, between Dammam-Khobar (KSA) and Bahrain, is further reinforced by these observations.

4.7 Effects of Train and Ferry Fare on Probabilities

Using the logit models developed for phase III (for both existing and new travelers), probabilities of mode choices were calculated for different fares of train and ferry. The proceeding sub-sections present the result and discussion for this analysis.

4.7.1. Travel Cost & Probability for Train Mode

Utility equation 22 was used for calculating probability of train, while the probability of car was estimated by subtracting probability of train from 100%. Figure 4-4 shows the effects of change in train fare on probabilities of train and car.

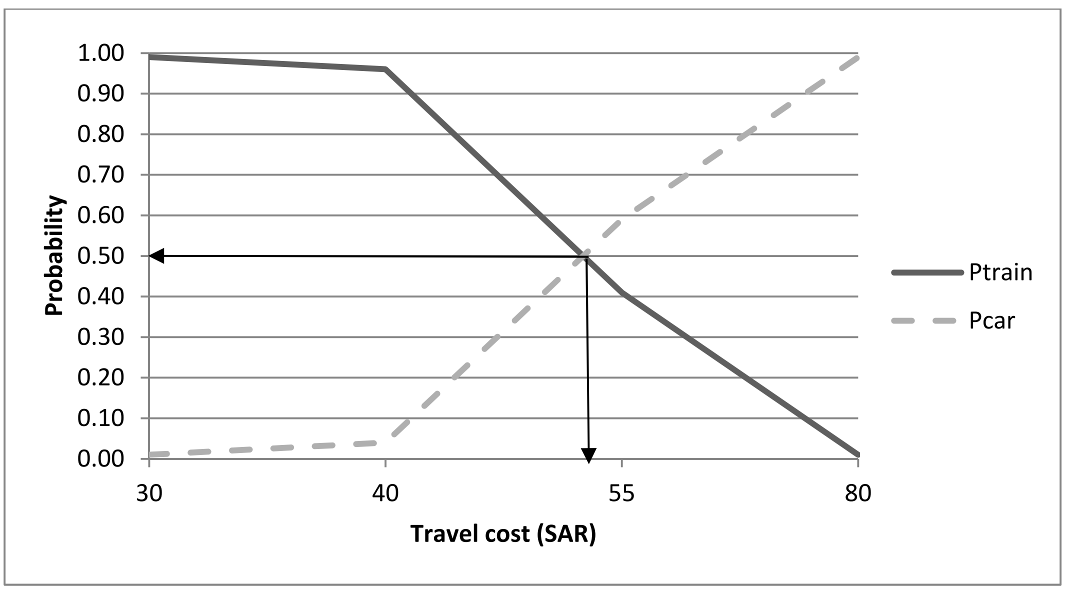

Figure 4-4: Change in Probability of Train and Car

Figure 4-4 shows that train has higher probability when its fare is equal or below SAR 55. Therefore, the optimum fare for train can be selected as SAR 50 for its effective operation.

4.7.2. Travel Cost & Probability for Ferry Mode

Utility functions shown in equations 23 and 24 were used to calculate probability of selecting ferry and car. Since, these utility functions contain travel cost, time and car ownership, so the probabilities were calculated for car owners and as well as for travelers without cars using average value of travel time for the sample (i.e. 125 minutes). Figure 4-5 shows the change in mode choice probabilities for car owners while figure 4-6 shows the same for travelers who do not own car(s).

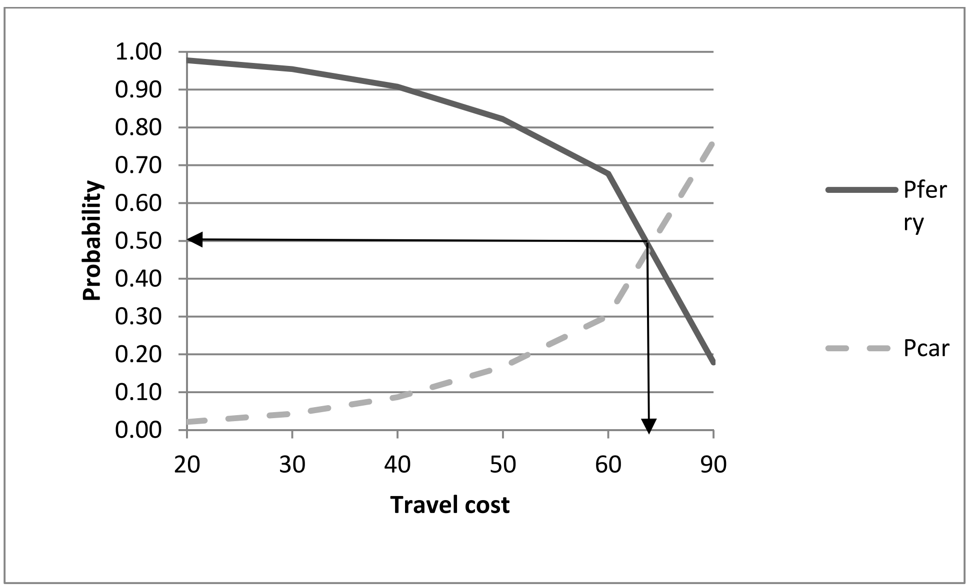

Figure 4-5: Change in Probability of Ferry and Car for Car Owners

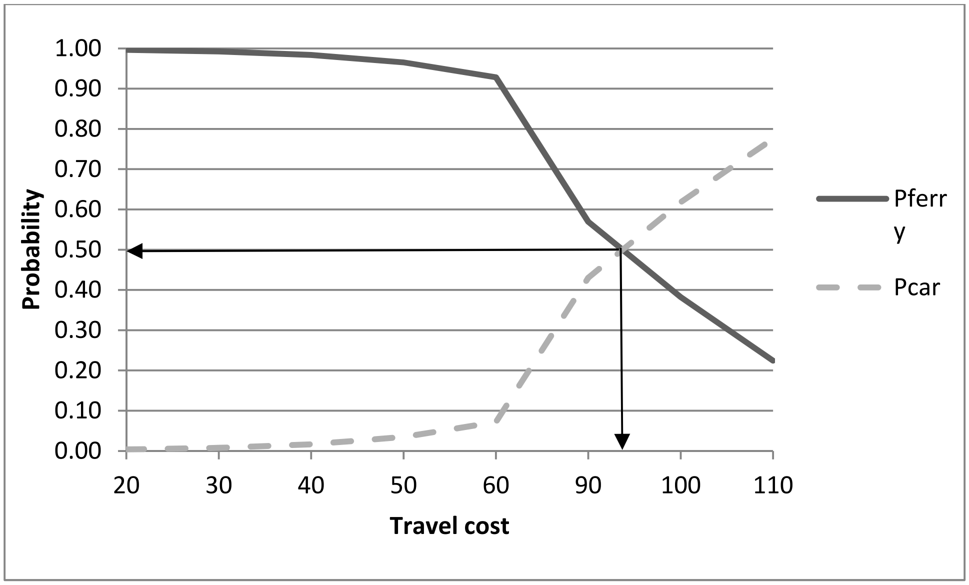

Figure 4-6: Change in Probability of Ferry and Car for Travelers without Cars

It can be observed from figures 4-5 that, ferry has higher probability than car for car owners when its fare is equal or less than SAR 70. On the other hand, travelers who do not own a car have higher probability of ferry when fare is equal or less than SAR 95, as shown in figure 4-6. Since, travelers who do not own car will not have the option to travel by their car so the acceptable fare for them is higher than car owners.

4.8 Discussion

The outcomes drawn from mode choice modeling exercise can be divided in to two categories; modeling outcomes and planning outcomes. Modeling outcomes focus upon the effectiveness of the techniques used for modeling. Planning outcomes are with regards to the behavior of travelers for the existing as well as new modes. They are presented in the following sub-sections.

4.8.1. Modeling Outcomes

The architecture of ANN models in each phase was selected after comparing accuracies from several networks architectures and types. The networks presented in table 4-18 gave the best results in respective situations for single ANN and ensemble, which were different from one another. MLPs and RBFNNs were the best models in most of the situations.

Single logit models perform better than ANN models in almost all situations. The probabilities calculated by logit models were used as an input to the second step of ensemble learning. It was found that the logit-ANN ensemble models give better results than ANN models in all cases. They also give consistently high accuracies for multinomial choice cases. However, the difference in accuracies between logit models and ensembles for binomial choice problem was not practically significant. This means that, single logit model can give accurate results for binomial mode choice problems.

Table 4-18: Best ANN Models for Mode Choice Modeling

Situation		Number of modes	Number of input variables	Model/Ensemble	Type	Number of hidden neurons
Existing modes	Phase I	2	23	Single ANN Model	RBFNN	13
				ANN Model used in Ensemble	MLP	7
Introduction of Train	Phase II	2	30	Single ANN Model	MLP	17 (2-layers)
				ANN Model used in Ensemble	RBFNN	9
Introduction of Ferry	Phase II	3	29	Single ANN Model	MLP	17
				ANN Model used in Ensemble	PNN	388*
Introduction of Train	Phase III	2	07	Single ANN Model	MLP	11
				ANN Model used in Ensemble	PNN	312*
Introduction of Ferry	Phase III	3	07	Single ANN Model	MLP	50
				ANN Model used in Ensemble	PNN	307*

*Equal to number of observations in training set

4.8.2. Planning outcomes

In the current scenario with three modes; on this route between Dammam-Khobar (KSA) and Bahrain, car is the most dominant mode. Bus is the least preferred mode and mainly taken by blue-collar labor class travelers for catching flights from Bahrain.

From the survey it was also observed that present car travelers are most resistant to switch to new modes if they are introduced. A considerable portion of these travelers will still prefer to use car instead of a new mode like train or ferry. When market shares were re-calculated after introducing train and ferry service in turn, it was found that train service will capture a larger share as compared to a ferry service. This is because train service attracts all almost the present airplane travelers and a major share of existing car travelers. The situation remains the same with the inclusion of new travelers as well. But inclusion of new travelers increases more share to ferry service than train which is different from existing travelers.

Tables 4-19 and 4-20 show the comparison of significant factors for different phases. It can be observed that travel cost has significant effect on mode utilities in all phases. Travel time is also important factor for present situation as well as for ferry service.

Table 4-19: Comparison of Significant Factors for Modeling Choice in Phase I and II

Phase I	Phase II	
	Train Service	Ferry Service
• Trips/year • Catching flight as the trip purpose • Recreation as the trip purpose • Frequency of traveling alone • Privacy as the reason for mode choice • Avoiding congestion as the reason for mode choice • Travel cost • Travel time	• Rigid schedule for rejecting the mode • Travel cost	• Trips per year • Frequency of traveling alone • Travel cost • Travel time

Car ownership becomes an important factor for mode choice decisions when new as well existing travelers are considered for ferry service. Among the available factors for new travelers (see table 4-14), same factors which were significant for existing travelers have been found significant in the utility functions for the case of new and existing travelers.

Table 4-20: Comparison of Significant Factors for Modeling Choice in Phase II and III

Train Service		Ferry Service	
Existing Travelers (Phase II)	New and Existing Travelers (Phase III)	Existing Travelers (Phase II)	New and Existing Travelers (Phase III)
• Rigid schedule for rejecting the mode • Travel cost	• Travel cost	• Trips per year • Frequency of traveling alone • Travel time • Travel cost	• Car ownership • Travel time • Travel cost

Chapter 5 CONCLUSIONS AND RECOMMENDATIONS

This study was conducted for comprehensive modeling of border transport in the Arabian Gulf countries. The travel between Dammam-Khobar region in Kingdom of Saudi Arabia (KSA) and Kingdom of Bahrain is selected for the analysis. This link is vital for travelers between KSA and Bahrain and also for travelers from other countries of this region. The travel demand on this route is increasing with time and results in congestion problems on routine problems. In order to improve the quality of travel on this route, short as well as long term measures are required. Accurate traffic forecasts play a vital role for devising short term operational measures to improve flow of traffic. One the other hand, long term travel demand management can be effectively done by providing new mode options to the travelers. Mode choice models are helpful in determining the change in behavior of travelers due to introduction of new modes, which will affect the effectiveness and feasibility of new modes.

Keeping short and long term issues of this route; this study was conducted in two parts; traffic forecasting and mode choice modeling. Firstly, suitable methods and techniques were reviewed from the literature. As a result of which it was found that Artificial Neural Networks (ANNs) based ensembles were not used for border transport modeling before. Therefore this study utilizes this contemporary approach for both types of modeling.

For traffic forecasting analysis, data was collected from different sources which are as follows. Traffic data for this study was provided by General Directorate of traffic in Bahrain, stock market index for KSA and weather parameters for both countries were collected from online resources. Air-traffic data was provided by King Fahd International

Airport Authority and the stock market index for Bahrain was provided by Director of Bahrain stock market. Disaggregate travelers' data for mode choice modeling was collected through interview-based questionnaire survey conducted in Bahrain and KSA in the shopping malls near King Fahd causeway.

For traffic forecasting, ensembles were created using 7 component models of multi-layer perceptrons (MLPs). Each component model had different range of input time series (look-back) data. The ensemble was developed by feeding the outputs (traffic forecasts) of the component models to another MLP as inputs. All MLP models were created using 10-fold cross-validation technique to avoid sampling errors.

Mode choice ensembles created in two steps. The first was to develop logit models using all the available input parameters. In the second step, mode choice probabilities calculated from the logit model were used as input to an ANN model. The architecture of all ANN models and utility function for logit models was selected after trying different types of ANN s including; Multi-layer perceptrons (MLPs), Radial Basis Function Neural Networks (RBFNNs) and Probabilistic Neural Networks (PNNs) with varying number of hidden neurons.

This study has investigated the use of stock market indices for border traffic forecasting as a surrogate measure of political and economic conditions of the countries. The use of stock indices in this regard is not found in present literature for border traffic forecasting. An ensemble consisting of logit and ANN models has also tried for mode choice modeling. This is the first that this type of ensemble has been used for mode choice modeling. More detail about the use of stock market indices and proposed ensemble is given in the proceeding sections.

5.1 Traffic Forecasting

A number of factors were used for traffic forecasting models including time-related parameters (name of the day of week, month, holidays), weather parameters (temperature and humidity) of both countries, stock market indices of both countries and number of passengers and flights for air-travel between Dammam and Bahrain. As an initial filter, correlation analysis was carried out using the Pearson correlation coefficient. All the variables which have coefficient value statistically different from '0' at 5% probability were considered for modeling. Based upon the correlation analysis, the parameters included in traffic forecasting models include traffic on the other side of the road, number of passengers and flights for air-travel, weekdays, salary disbursement period, summer vacation, religious holidays (Hajj) and humidity of Bahrain. This is due to the fact that many people travel to Bahrain for recreational or social visits therefore these factors affect their travel decisions.

Considering the fact that predicted traffic data was across border, stock market indices of both countries were used as indicators of political stability and economic prosperity. This hypothesis proved valid by the results of the correlation analysis as well as the results of predictive models. The models showed an improvement in performance when stock market indices were used as the input in addition to traffic counts.

Predictive models were developed using input look-back period from 1 to 7 days, for different prediction horizons (1, 7, 30, 90 and 365 days). An ensemble was created for each prediction horizon. Three trends were observed from these models and ensembles. Firstly, the accuracies of predictive models increase as the range of input look-back period was increased. Secondly, the accuracies tend to decrease with the increase in prediction horizons. Also, ensembles perform better than the component models.

It was further observed that the use of only traffic and stock market indices can give sufficient accuracy without the use of any other exogenous or traffic related factor for prediction horizons of 30 days and more. However, for daily and weekly predictions, the use of other parameters also adds to the accuracy. This trend can be attributed to the fact that short-term predictions require more explanatory power in terms of input variables. Another explanation for this trend can be that the economic and political conditions, captured by the stock market indices, affect traffic with a lagging period of more than one week. The appropriate variable set for each prediction horizon is shown in table 5-1.

Table 5-1: Variable Sets Recommended for Different Prediction Horizons

Prediction Horizon	Variable Set
1-day (daily)	Traffic counts (incoming and outgoing), humidity of Bahrain, Hajj vacations, salary disbursement period, summer vacations, stock market indices of KSA and Bahrain, daily number of flights and passengers by airplane between Dammam and Bahrain (incoming and outgoing)
7-days (weekly)	Traffic counts (incoming and outgoing), humidity of Bahrain, Hajj vacations, salary disbursement period, summer vacations, stock market indices of KSA and Bahrain, daily number of flights and passengers by airplane between Dammam and Bahrain (incoming and outgoing)
30-days (monthly)	Traffic counts (incoming and outgoing), stock market indices of KSA and Bahrain
90-days (quarterly)	Traffic counts (incoming and outgoing), stock market indices of KSA and Bahrain
365-days (yearly)	Traffic counts (incoming and outgoing), stock market indices of KSA and Bahrain

5.2 Mode Choice Modeling

Disaggregate data, for current and potential new travelers, were collected through an interview survey. The survey was done in shopping malls near King Fahd causeway in KSA as well as Bahrain. There were 654 valid responses collected and the expected error of 4% was calculated for this sample size.

From analysis of the survey responses it was found that in the current scenario car is the most dominant mode while bus is the least preferred mode. Bus mode was found to be least preferred and its travelers were mainly blue-collar captive riders. Due to these issues, bus travelers were not included in the mode choice models.

The survey also included questions on hypothetical modes of train and ferry service, if they are introduced in future. It was observed that car travelers are most resistant in adopting new modes. When market shares were re-calculated after introducing train and ferry service in turn, it was found that train service will capture a larger share as compared to a ferry service. The market shares for these new modes were also calculated with the inclusion of new potential travelers. These are travelers who may travel on this route in future because of the better service or new features offered by new modes. Train service had a greater market share than ferry service after the inclusion of new travelers as well. However, it was observed that new travelers are more attracted to ferry service. This trend was different from the existing travelers so their inclusion in modeling for better market share prediction is justified.

From the logit model, it was found that travel cost is a significant parameter for mode utilities in the present situation as well as with train or ferry service. The optimum fare, below which hypothetical mode (train or ferry) has more probability than car, for train was estimated to be approximately SAR 50 and that for ferry ranges from SAR 70 to 95.

Three types of mode choice models were developed for each situation; single ANN, single logit and logit-ANN ensemble. The types of ANNs used for mode choice modeling were multi-layer perceptrons (MLPs), probabilistic neural networks (PNNs) and radial basis function neural networks (RBFNNs).

Single Logit models give better prediction results than single ANN models in almost all cases. It was also found that the logit-ANN ensemble give better accuracy than single ANN models for all cases. Ensembles also give consistently high accuracies for multinomial problems (more than 2 choices). For binomial choice problems, logit model was found to be sufficiently accurate and the use of ensemble was not practically feasible.

5.3 Recommendations

The following recommendations are made for transportation planners and researchers after taking the observations of this study in to account.

- All relevant factors should be included in the predictive modeling for accurate daily and weekly forecasts
- It is recommended to collect detailed travelers' data for accurate mode choice modeling when trip data is not available as for new modes or new travelers
- Other types of ensembles combining different types of ANNs should be explored

for future research

- Machine learning techniques can be used for finding optimum architecture of ANN (number of hidden neurons and layers) to give best accuracies
- The logit-ANN ensemble is recommended to be used for mode choice modeling, especially for complex problems involving more than 2 choices
- A comparison between ferry and train service can be made by presenting both options simultaneously. This will also help to investigate the factors affecting the demand of each of these modes

REFERENCES

[1] H. Al-Ahmadi, "Development of Intercity Mode Choice Models for Saudi Arabia," *J. King Abdulaziz Univ. Sci.*, vol. 17, no. 1, pp. 3–20, 2006.

[2] J. D. D. Ortuzar and L. G. Willumsen, *Modelling Transport*, Third. London: John Wiley & Sons Inc., 2002, pp. 199–205.

[3] C. R. Bhat, "Incorporating Observed and Unobserved Heterogenity in Urban Work Travel Mode Choice Modeling," *Transp. Sci.*, vol. 34, no. 2, pp. 228–238, 2000.

[4] H. Dia and S. Panwai, "Modelling drivers' compliance and route choice behaviour in response to travel information," *Nonlinear Dyn.*, vol. 49, no. 4, pp. 493–509, Jan. 2007.

[5] Y. Suzuki, "The relationship between on-time performance and airline market share: a new approach," *Transp. Res. Part E Logist. Transp. Rev.*, vol. 36, no. 2, pp. 139–154, Jun. 2000.

[6] J. Y. T. Wang, "Appraisal of Factors Influencing Public Transport Patronage," in *Proceedings of the 32nd Australasian transport research forum*, 2009, pp. 1–11.

[7] G. Cohen and M. Kocis, "Components of Change in Urban Travel," *Transp. Res. Rec.*, vol. 775, pp. 42–47, 1980.

[8] U. G. O. Lachapelle, "Public transit use as a catalyst for an active lifestyle: mechanisms, dispositions and hindrances," University of British Columbia, 2010.

[9] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. Part C Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, Jun. 2011.

[10] P. K. Edara, "Mode Choice Modeling Using Artificial Neural Networks," Virginia Polytechnic Institute and State University, 2003.

[11] D. Chen and J. Zhang, "Time series prediction based on ensemble anfis," in *Proceedings of the Fourth International on Machine Learning and Cybernetics*, 2005, no. August, pp. 18–21.

[12] R. Adhikari and R. K. Agarwal, "A Homogeneous Ensemble of Artificial Neural Networks for Time Series Forecasting," *Int. J. Comput. Appl.*, vol. 32, no. 7, pp. 1–8, 2011.

[13] T. Petersen, "Modelling Cross-border tTransport in Öresund : Scenario Assumptions , Model results and a Before-and-after Database," *Rev. Reg. Res.*, pp. 173–210, 2004.

[14] P. Rietveld, "Obstacles to Openness of Border Regions in Europe," in *Gaining advantage from open borders. An active space approach to regional development*, 2001, pp. 79–96.

[15] A. Ahmed and U. Gazder, "Traffic Forecasting for King Fahd Causeway : Comparison of Parametric Technique with Artificial Neural Networks," in *Proceeding of the International Conference on Advances in Structural, Civil and Environmental Engineering -- SCEE 2013*, 2013, vol. 1, no. 1, pp. 75 – 79.

[16] Federal Deprartment of Transportation, "Traffic Forecasting," Resource Guidebook by Intermodal Systems Development Unit, Florida, 2006.

[17] R. Reel, S. Friedman, V. Mysore, F. Tabatabaee, and V. Majano, *Project Traffic Forecasting*. Florida Department of Transportation, 2012.

[18] A. M. Martin and A. M. Nancy, "NCHRP Report 365: Travel Estimation Techniques for Urban Planning," 1998.

[19] A. Stathopoulos, L. Dimitriou, and T. Tsekeris, "Fuzzy modeling approach for combined forecasting of urban traffic flow," *Comput. Civ. Infrastruct. Eng.*, vol. 23, pp. 521–535, 2008.

[20] P. Cortez, M. Rio, M. Rocha, and P. Sousa, "Internet Traffic Forecasting using Neural Networks," in *2006 International Joint Conference on Neural Networks*, 2006, pp. 2635–2642.

[21] P. Cortez, M. Rio, M. Rocha, and P. Sousa, "Multi-scale Internet traffic forecasting using neural networks and time series methods," *Expert Syst.*, vol. 29, no. 2, pp. 143–155, 2012.

[22] Y. Zhang and Y. Liu, "Comparison of Parametric and Nonparametric Techniques for Non-peak Traffic Forecasting," *World Acad. Sci. Eng. Technol.*, pp. 8–14, 2009.

[23] A. Nagare and S. Bhatia, "Traffic Flow Control using Neural Network," *Int. J. Appl. Inf. Syst.*, vol. 1, no. 2, pp. 50–52, 2012.

[24] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial Neural Networks: A Tutorial," *IEEE*, pp. 31–44, 1996.

[25] D. F. Specht, "Probabilistic Neural Networks," *Neural Networks*, vol. 3, pp. 109–118, 1990.

[26] H. Chen, S. Grant-muller, L. Mussone, and F. Montgomery, "A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting," *Neural Comput. Appl.*, no. 10, pp. 277–286, 2001.

[27] G. E. P. Box and D. A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *J. Am. Stat. Assoc.*, vol. 65, no. 332, pp. 1509–1526, 1970.

[28] F. Qiao, H. Yang, and W. H. K. Lam, "Intelligent simulation and prediction of traffic flow dispersion," *Transp. Res. Part B*, vol. 35, pp. 843–863, 2001.

[29] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res. Part C Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, Jun. 2005.

[30] T. Ma and B. Abdulhai, "Genetic Algorithm-Based Optimization Approach and Generic Tool for Calibrating Traffic Simulation Parameters," *Transp. Res. Rec.*, vol. 1800, no. 1800/2002, pp. 6–15, 2002.

[31] W.-C. Hong, Y. Dong, F. Zheng, and S. Yung, "Hybrid evolutionary algorithms in a SVR traffic flow forecasting model," *Appl. Math. Comput.*, vol. 217, no. 15, pp. 6733–6747, 2011.

[32] Y.-H. Tang and B. Xi, "Dynamic forecasting of traffic volume based on quantificational dynamics : A nearness perspective," *Sci. Res. Essays*, vol. 5, no. 4, pp. 389–394, 2010.

[33] U. Gazder and S. A. Hussain, "Traffic Forecasting for King Fahd Causeway Using Artificial Neural Networks," in *UKSIM - AMSS 15th International Conference on Modeling and Simulation*, 2013, pp. 1–5.

[34] L. Tsirigotis, E. I. Vlahogianni, and M. G. Karlaftis, "Does Information on Weather Affect the Performance of Short-Term Traffic Forecasting Models?," *Int. J. Intell. Transp. Syst. Res.*, vol. 10, no. 1, pp. 1–10, Aug. 2011.

[35] Q. Yang, "A Simulation Laboratory for Evaluation of Dynamic Traffic Management Systems," Massachusetts Institute of Technology, 1997.

[36] G. de Jong, A. Daly, M. Pieters, C. Vellay, M. Bradley, and F. Hofman, "A Model for Time of Day and Mode Choice Using Error Components Logit," *Transp. Res. Part E*, vol. 39, no. 3, pp. 245–268, 2003.

[37] P. B. Goodwin, "Empirical evidence on induced traffic A review and synthesis," *Transportation (Amst).*, no. 23, pp. 35–54, 1996.

[38] J. L. Horowitz, F. S. Koppelman, and S. R. Lerman, "A Self-instructing course in disaggregate mode choice modeling," 1986.

[39] M. Abou-Zeid and D. M. Scott, "Current Issues in Mode Choice Modeling," *Transportation (Amst).*, vol. 38, no. 2011, pp. 581–585, 2011.

[40] K. Train and D. McFadden, "The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models," *Transp. Res.*, vol. 12, pp. 349–353, 1978.

[41] S. N. Aljarad and W. R. Black, "Modeling Saudi Arabia-Bahrain corridor mode choice," *J. Transp. Geogr.*, vol. 3, no. 4, pp. 257–268, 1995.

[42] N. Sanko and T. Morikawa, "Temporal Transferability of Updated Alternative-Specific Constants in Disaggregate Mode Choice Models," *Transportation (Amst).*, vol. 37, no. 2010, pp. 203–219, 2010.

[43] C. R. Bhat, "Analysis of Travel Mode and Departure Time Choice for Urban Shopping Trips," *Transp. Res. Part B Methodol.*, vol. 32, no. 6, pp. 361–371, 1998.

[44] C. R. Bhat, "Accommodating Flexible Substitution in Multi-Dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice," *Transp. Res. Part B Methodol.*, vol. 32, no. 7, pp. 455–466, 1998.

[45] B. H. Y. Lee and P. Waddell, "Residential mobility and location choice: a nested logit model with sampling of alternatives," *Transportation (Amst).*, vol. 37, no. 4, pp. 587–601, Apr. 2010.

[46] D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in *In Frontiers in Econometrics*, 1974, pp. 105–142.

[47] M. P. Keane, "A Note on Identification in the Multinomial Probit Model," *J. Bus. Econ. Stat.*, vol. 10, no. 2, pp. 193–200, 1992.

[48] T. Yai, S. Iwakura, and S. Morichi, "Multinomial Probit With Structured Covariance for Route Choice Behavior," *Transp. Res. B*, vol. 31, no. 3, pp. 195–207, 1997.

[49] J. F. Foerster, "Mode Choice Decision Process Models: A Comparison of Compensatory and Non-Compensatory Structures," *Transp. Res.*, vol. 13A, pp. 17–28, 1979.

[50] D. A. Hensher and W. H. Greene, "The Mixed Logit model : The state of practice," *Transp. 30*, pp. 133–176, 2003.

[51] L. Yu and B. Sun, "Four Types of Typical Discrete Choice Models: Which are You Using?," *IEEE*, no. 2, pp. 298–301, 2012.

[52] M. Ben-akiva, D. Mcfadden, K. Train, and A. Borsch-Supan, "Hybrid Choice Models : Progress and Challenges," no. January. pp. 163–175, 2002.

[53] G. E. Cantarella and S. de Luca, "Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models," *Transp. Res. Part C Emerg. Technol.*, vol. 13, no. 2, pp. 121–155, Apr. 2005.

[54] D. S. Bunch, "Estimability In The Multinomial Probit Model," *Transp. Res. B*, vol. 258, no. I, pp. 1–12, 1991.

[55] J. L. Horowitz, "Reconsidering the Multinomial Probit Model," *Transp. Res. B*, vol. 25, no. 6, pp. 433–438, 1991.

[56] O. A. Khan, "Modeling Passenger Mode Choice Behavior Using Computer Aided Stated Preference Data," Queensland University of Technology, Queensland, 2004.

[57] C. R. Bhat, "A heteroscedastic extreme value model of intercity travel mode choice," *Transp. Res. Part B Methodol.*, vol. 29, no. 6, pp. 471–483, Dec. 1995.

[58] F. S. Koppelman and C. Bhat, "A Self Instructing Course in Mode Choice Modeling : Multinomial and Nested Logit Models," 2006.

[59] D. McFadden and K. Train, "Mixed MNL Models for Discrete Responce," *J. Appl. Econom.*, vol. 15, no. November, pp. 447–470, 2000.

[60] W. H. Greene and D. A. Hensher, "Revealing Additional Dimensions of Preference Heterogeneity in a Latent Class Mixed Multinomial Logit Model," *Appl. Econ.*, vol. 45, no. 14, pp. 1897–1902, 2012.

[61] N. T. Ratrout and S. M. Rahman, "A Comparative Analysis of Currently Used Micriscopic and Macroscopic Traffic Simulation Software," *Arab. J. Sci. Eng.*, vol. 34, no. 1, pp. 121–133, 2009.

[62] D. A. Hensher and T. T. Ton, "A Comparison of the Predictive Potential of Artificial Neural Networks and Nested Logit Models for Commuter Mode Choice," *Transp. Res. Part E*, vol. 36, no. 2000, pp. 155–172, 2000.

[63] M. Shir-mohammadli, S. N. Shetab-Bushehri, H. Poorzahedy, and S. R. Hejazi, "A comparative study of a hybrid Logit – Fratar and neural network models for trip

distribution : case of the city of Isfahan," *J. Adv. Transp.*, vol. 45, no. September 2010, pp. 80–93, 2011.

[64] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transp. Res. Part C Emerg. Technol.*, vol. 21, no. 1, pp. 148–162, Apr. 2012.

[65] H. B. Celikoglu, "Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling," *Math. Comput. Model.*, vol. 44, no. 7–8, pp. 640–658, Oct. 2006.

[66] F. Tillema, K. M. van Zuilekom, and M. F. A. M. van Maarseveen, "Comparison of neural networks and gravity models in trip distribution," *Comput. Civ. Infrastruct. Eng.*, vol. 21, pp. 104–119, 2006.

[67] Y. Zhang and Y. Xie, "Travel Mode Choice Modeling with Support Vector Machines," *Transp. Res. Rec. J. Transp. Res. Board*, no. 15, pp. 141–150, 2008.

[68] A. Tortum, N. Yayla, and M. Gukdag˘, "The modeling of mode choices of intercity freight transportation with the artificial neural networks and adaptive

neuro-fuzzy inference system," *Expert Syst. with Appl. An Int. J.*, vol. 36, no. 3, pp. 6199–6217, 2009.

[69] D. Shmueli, I. Salomon, and D. Shefer, "Neural network analysis of travel behavior: evaluating tools for prediction," *Transp. Res. Part C Emerg. Technol.*, vol. 4, no. 3, pp. 151–166, 1996.

[70] A. Faghri and J. Hua, "Roadway seasonal classification using neural networks," *J. Comput. Civ. Eng.*, vol. 9, no. 3, pp. 209–215, 1995.

[71] D. Teodorovic, "Fuzzy logic systems for transportation engineering : the state of the art," *Transp. Res. Part A*, vol. 33, pp. 337–364, 1999.

[72] K. Andrade, K. Uchida, and S. Kagaya, "Development of Transport Mode Choice Model by Using Adaptive Neuro-Fuzzy Inference System," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1977, no. 1, pp. 8–16, Jan. 2006.

[73] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. Part C*, vol. 10, pp. 85–98, 2002.

[74] H. Omrani, O. Charif, P. Gerber, A. Awasthi, and P. Trigano, "Prediction of Individual Travel Mode with Evidential Neural Network Model," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2399, no. 1, pp. 1–8, Dec. 2013.

[75] D. Bolduc, "A practical technique to estimate multinomial probit models in transportation," *Transp. Res. Part B*, vol. 33, pp. 63–79, 1999.

[76] R. Buehler, "Determinants of transport mode choice: a comparison of Germany and the USA," *J. Transp. Geogr.*, vol. 19, no. 4, pp. 644–657, Jul. 2011.

[77] A. Reggiani, P. Nijkamp, and W. F. Tsang, "European Freight Transport Analysis Using Neural Networks and Logit Models," 1997.

[78] G. de Jong and H. F. Gunn, "Recent Evidence on Car Cost and Time Elasticities of Travel Demand in Europe Gerard de Jong and Hugh Gunn," *J. Transp. Econ. Policy*, vol. 35, no. 2, pp. 137–160, 2001.

[79] D. A. Hensher, P. R. Stopher, and J. J. Louviere, "An exploratory analysis of the effect of numbers of choice sets in designed choice experiments : an airline choice application," *J. Air Transp. Manag.*, vol. 7, no. 1, pp. 373–379, 2001.

[80] J. Lin and B. Zhu, "A Novel Neural Network Ensemble System for Economic Forecasting," *Commun. Comput. Inf. Sci.*, vol. 2, pp. 1227–1233, 2007.

[81] M. L. F. Miguel, M. C. Penna, J. C. Nievola, and M. E. Pellenz, "New models for long-term Internet traffic forecasting using artificial neural networks and flow based information," *2012 IEEE Netw. Oper. Manag. Symp.*, pp. 1082–1088, Apr. 2012.

[82] A. E. D. Mahmoud Hussein, "Enhanced Ensemble Learning Technique With A Case Study Of Video Network Traffic Forecasting," Cairo University, 2005.

[83] Y. Bao, T. Xiong, and Z. Hu, "Forecasting Air Passenger Traffic by Support Vector Machines with Ensemble Empirical Mode Decomposition and Slope-Based Method," *Discret. Dyn. Nat. Soc.*, vol. 2012, pp. 1–12, 2012.

[84] L. Chen and C. L. P. Chen, "Ensemble Learning Approach for Freeway Short-Term Traffic Flow Prediction," in *2007 IEEE International Conference on System of Systems Engineering*, 2007, pp. 1–6.

[85] Y. J. Yu and M.-G. Cho, "A Short-Term Prediction Model for Forecasting Traffic Information Using Bayesian Network," in *2008 Third International Conference on Convergence and Hybrid Information Technology*, 2008, pp. 242–247.

[86] P. Nitsche, P. Widhalm, S. Breuss, N. Brändle, and P. Maurer, "Supporting large-scale travel surveys with smartphones – A practical approach," *Transp. Res. Part C Emerg. Technol.*, Dec. 2013.

[87] S. R. Eddy, "Hidden Markov Models," *Curr. Opin. Struct. Biol.*, vol. 6, no. 3, pp. 361–365, 1996.

[88] F. Rashid, "Stock Market Indices for Bahrain Stock Market," 2013.

[89] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *International Joint Conference on Arti cial Intelligence (IJCAI)*, 1995, pp. 1137–1145.

[90] R. J. Hyndman and A. B. Koehler, "Another Look at Measures of Forecast Accuracy," *Int. J. Forecast.*, vol. 22, no. May, pp. 679–688, 2005.

[91] S. Sudman, *Applied Sampling*. New York: Academic Press, 1976, pp. 1–50.

APPENDIX A: SURVEY QUESTIONNAIRES

#

Questionnaire for Pilot Survey

1. Nationality:

□ Saudi National □ Bahraini National

□ Other Gulf countries □ Other Arab Countries

□ South Asian (India, Pakistan, Bangladesh, Sri Lanka, Nepal, Bhutan)

□ Other Asian Countries

□ Western Countries (Europe, USA, Canada and Australia) □ African

2. Age: ___________________

3. Gender:

□ Male □ Female

4. Occupation:

□ Doctor

□ Engineer

□ Lawyer

□ Accountant/Financial Affairs

□ Marketing

□ Farmer

□ Teacher/Researcher

□ Student

□ Labor

□ Technician

□ Driver

□ Other (________________________)

5. What is your Purpose of traveling to Bahrain (Select all that apply)?

□ Residence

□ Business/Work

□ Connecting flight from Bahrain/Saudi Arabia

□ Education

□ Other (Specify) _________

□ Recreation/Sight Seeing

□ Social Visit

□ Health

□ Shopping

6. How often do you travel through the King Fahd causeway?

(Example: for 3 trips per week; write 3 in the blank and select 'Week' from the options)

____________ trips per (select one from the given)

□ Day

□ Week

□ Month

□ Year

7. Monthly Household Income/Salary (In SAR): _____________________

8. How many cars you own or available to you for travel?

□ 0

□ 2

□ 1

□ More than 2

9. How often do you travel with a group or family, when traveling to/from Bahrain?

□ Never

□ Often

□ Seldom

□ Every time

10. Number of travelers in the group, traveling to/from Bahrain:

□ I do not travel with a group/family □ 2

□ 3 □ 4

□ 5 □ More than 5

11. What is your preferred mode of travel for travel to/from Bahrain?

□ Car □ Bus □ Airplane

12. Reason(s) for Choosing the above selected Mode (Select all that apply):

□ Because I can travel on convenient times □ Because it costs less

□ Because it safer □ Because I can have privacy

□ Because total travel time (door-to-door) is less

□ Because I am not affected by congestion on the causeway

□ Because it provides mobility within Bahrain □ Other (Specify) _________

13. If you had any experience of traveling on the causeway, please answer the following:

a. Expected Time Spent in Crossing the Causeway (from the time of paying tolls till the end of causeway): _________________________(hrs./min.)

b. Expected time from your house/bus station to the causeway toll gate: _________________________(hrs./min.)

c. Expected time from end of the causeway to your destination (house, work, airport, etc.): _________________________(hrs./min.)

(Adding all parts of Q 13 give total travel time from origin to destination)

14. If you know in advance (through internet, radio, traffic signs, SMS) that there is congestion on King Fahd causeway at the time of your travel, what will you do?

□ Do nothing □ Choose another Mode

□ Choose another Time □ Choose another Day

□ Cancel the trip

15. If a train service is started for this route, what would you prefer as the one-way fare per passenger for train service? SAR______________________________

16. What would you expect as the total travel time (including immigration and other formalities) for train service from Khobar to Bahrain station?

_____________ (hrs/min)

17. If the train service operates at your expected fare and travel time, would you like to take it instead of your current mode of travel?

□ Never □ Seldom

□ Often □ Every time

18. Reason(s) for Choosing this train service (Select all that apply):

□ Because it costs less □ Because I do not have a car

□ Because it is safer

□ Because total travel time (door-to-door) is less

□ Because I will not be affected by congestion on the causeway

□ Other (Specify) _________

19. If the train service is extended to Qatar, will you use it to travel to or from Qatar?

□ Never □ Seldom

□ Often □ Every time

20. Reason(s) for Choosing train service to travel to or from Qatar (Select all that apply):

□ Because it costs less □ Because I do not have a car

□ Because it is safer

□ Because total travel time (door-to-door) is less

□ Because I will not be affected by congestion on the causeway

□ Other (Specify) _________

21. If a ferry (small ship in which people can board with their vehicles) service is started for this route, what would you prefer as the fare per passenger for ferry service?

SAR_____________________________

22. What would you expect as the total travel time (including immigration and other formalities) for ferry service from Khobar to Bahrain station?

__________________ (hrs/min)

23. If the ferry service operates at your expected fare and travel time, would you like to take it instead of your current mode of travel?

□ Never □ Seldom

□ Often □ Every time

24. Reason(s) for Choosing ferry service (Select all that apply):

□ Because it costs less □ Because I do not have a car

□ Because it is safer

□ Because total travel time (door-to-door) is less

□ Because I will not be affected by congestion on the causeway □ Other (Specify) _________

Questionnaire for Final Survey

1. Nationality:

□ Saudi National □ Bahraini National

□ Other Gulf countries □ Other Arab Countries

□ Asian □ African

□ Western Countries (Europe, USA, Canada and Australia)

2. Age: ___________________

3. Occupation:

□ Doctor □ Engineer □ Lawyer □ Accountant/Financial Affairs

□ Marketing □ Farmer □ Teacher/Researcher □ Student

□ Labor□ Technician □ Driver □ Own Business

□ Other (_________________)

4. Monthly Household Income/Salary (In BD): _____________________

5. Do you own a car?

□ Yes □ No

6. Do you travel to Bahrain/Saudi Arabia?

□ Yes □ No

(If ‘No’ go to question 16)

7. How many trips to Bahrain/Saudi Arabia per year? _______________

8. Purpose of Travel:

□ Residence □ Recreation/Sight Seeing

□ Business/Work □ Social Visit

□ Connecting flight from Bahrain/Saudi Arabia □ Health

□ Education □ Shopping

□ Other (Specify) _________

9. Do you travel alone?

□ Yes □ No □ Sometimes

10. What is your preferred mode of travel?

□ Car □ Bus □ Airplane

11. Reason(s) for preferring this mode (Select all that apply):

□ Convenient times □ Costs less □ Safer □ Privacy

□ Less total travel time □ Mobility within Bahrain

□ Avoiding causeway congestion □ Other (Specify) _________

12. Average time from your house to the causeway/airport:

_________(min.)

13. Average time spent in crossing the Causeway (only for car and bus travelers):
_________________________(min.)

14. Average time from end of the causeway/airport to your destination:
_________________________(min.)

15. Cost of travel ____________________ (BD)

16. If a train service is started for this route with fare of 40 SAR per passenger (one way) taking 40 minutes to travel, would you use it to travel to Bahrain/Saudi Arabia?

□ Yes □ No

17. If not why (Select all that apply)?

□ Schedule will not be flexible □ High fare □ Not Safe

□ No Privacy □ More total travel time

□ No Mobility within Bahrain □ Not interested to visit Bahrain

□ Other (Specify) _________

18. If a ferry service is started for this route with fare of 50 SAR per passenger (one way) taking 60 minutes to travel, would you use it to travel to Bahrain/Saudi Arabia?

□ Yes □ No

19. If not why (Select all that apply)?

□ Schedule will not be flexible □ High fare □ Not Safe

□ No Privacy □ More total travel time

□ No Mobility within Bahrain □ Not interested to visit Bahrain

□ Other (Specify) _________

APPENDIX B: RESULTS OF ITERATIONS FOR TRAFFIC FORECASTING

Predictions Using Traffic Only

1-Day Prediction Using 1-Step

Model	Relative MSE	MAPE
1 hidden neuron, hyperbolic function	0.055	0.140
2 hidden neuron, hyperbolic function	0.053	0.138
3 hidden neuron, hyperbolic function	0.053	0.137
4 hidden neuron, hyperbolic function	0.051	0.136
5 hidden neuron, hyperbolic function	0.050	0.134
6 hidden neuron, hyperbolic function	0.051	0.135
5 hidden neuron, logistic function	0.050	0.135
2 layers (5+1) hidden neuron, hyperbolic function	0.099	0.166

1-Day Prediction Using 2-Step

Model	Relative MSE	MAPE
2 hidden neuron, hyperbolic function	0.051	0.136
3 hidden neuron, hyperbolic function	0.048	0.133
4 hidden neuron, hyperbolic function	0.048	0.133
4 hidden neuron, logistic function	0.048	0.133
2 layers (4+2) hidden neuron, logistic function	0.075	0.165

1-Day Prediction Using 3-Step

Model	Relative MSE	MAPE
3 hidden neuron, hyperbolic function	0.035	0.114
4 hidden neuron, hyperbolic function	0.033	0.111
5 hidden neuron, hyperbolic function	0.035	0.114
4 hidden neuron, logistic function	0.035	0.114
2 layers (4+2) hidden neuron, hyperbolic function	0.085	0.167

1-Day Prediction Using 4-Step

Model	Relative MSE	MAPE
4 hidden neuron, hyperbolic function	0.036	0.109
5 hidden neuron, hyperbolic function	0.034	0.109
4 hidden neuron, logistic function	0.035	0.110
2 layers (4+2) hidden neuron, hyperbolic function	0.049	0.128

1-Day Prediction Using 5-Step

Model	Relative MSE	MAPE
5 hidden neuron, hyperbolic function	0.033	0.105
6 hidden neuron, hyperbolic function	0.031	0.103
7 hidden neuron, hyperbolic function	0.032	0.105
6 hidden neuron, logistic function	0.031	0.102
2 layers (6+3) hidden neuron, logistic function	0.035	0.104

1-Day Prediction Using 6-Step

Model	Relative MSE	MAPE
6 hidden neuron, hyperbolic function	0.033	0.105
7 hidden neuron, hyperbolic function	0.032	0.100
8 hidden neuron, hyperbolic function	0.032	0.100
8 hidden neuron, logistic function	0.033	0.100
2 layers (8+4) hidden neuron, hyperbolic function	0.035	0.104

1-Day Prediction Using 7-Step

Model	Relative MSE	MAPE
7 hidden neuron, hyperbolic function	0.019	0.076
8 hidden neuron, hyperbolic function	0.020	0.077
7 hidden neuron, logistic function	0.019	0.078
2 layers (7+3) hidden neuron, hyperbolic function	0.019	0.078

All Predictions Using Best Configurations

Model	Steps Used	Prediction Horizon	Relative MSE	MAPE
5 hidden neuron, hyperbolic function	1	1	0.050	0.134
		7	0.100	0.123
		30	0.174	0.222
		90	0.193	0.220
		365	0.209	0.222
4 hidden neuron, logistic function	2	1	0.048	0.133
		7	0.097	0.120
		30	0.174	0.213
		90	0.174	0.214
		365	0.208	0.221
4 hidden neuron, hyperbolic function	3	1	0.033	0.111
		7	0.097	0.119
		30	0.172	0.163
		90	0.176	0.186
		365	0.204	0.189
4 hidden neuron, hyperbolic function	4	1	0.036	0.109
		7	0.095	0.118
		30	0.173	0.154
		90	0.175	0.186
		365	0.204	0.186
6 hidden neuron, logistic function	5	1	0.031	0.102
		7	0.097	0.122
		30	0.174	0.212
		90	0.175	0.212
		365	0.199	0.215
8 hidden neuron, hyperbolic function	6	1	0.032	0.100
		7	0.092	0.121
		30	0.175	0.182
		90	0.177	0.184
		365	0.190	0.196
7 hidden neuron, hyperbolic function	7	1	0.019	0.076
		7	0.092	0.118
		30	0.171	0.178
		90	0.172	0.181
		365	0.189	0.164

Performance of Ensembles

Model	Relative MSE	MAPE
1-Day Prediction		
9 hidden neurons, logistic output function	0.0177	0.0740
7-Day Prediction		
9 hidden neurons, logistic output function	0.090	0.117
30-Day Prediction		
9 hidden neurons, logistic output function	0.134	0.171
90-Day Prediction		
9 hidden neurons, logistic output function	0.165	0.180
365-Day Prediction		
9 hidden neurons, logistic output function	0.188	0.158

Prediction with Traffic and Stock

1-Day Prediction Using 1-Step

Model	Relative MSE	MAPE
3 hidden neuron, hyperbolic function	0.048	0.136
4 hidden neuron, hyperbolic function	0.052	0.138
3 hidden neuron, logistic function	0.054	0.137
2 layers (3+2) hidden neurons, hyperbolic function	0.055	0.136

1-Day Prediction Using 2-Step

Model	Relative MSE	MAPE
6 hidden neuron, hyperbolic function	0.046	0.132
7 hidden neuron, hyperbolic function	0.046	0.132
7 hidden neuron, logistic function	0.043	0.129
2 layers (7+3) hidden neurons, logistic function	0.046	0.129

1-Day Prediction Using 3-Step

Model	Relative MSE	MAPE
9 hidden neuron, hyperbolic function	0.033	0.116
10 hidden neuron, hyperbolic function	0.034	0.118
9 hidden neuron, logistic function	0.034	0.116
2 layers (9+4) hidden neurons, hyperbolic function	0.033	0.111

1-Day Prediction Using 4-Step

Model	Relative MSE	MAPE
10 hidden neuron, hyperbolic function	0.035	0.112
11 hidden neuron, hyperbolic function	0.033	0.113
12 hidden neuron, hyperbolic function	0.035	0.115
11 hidden neuron, logistic function	0.036	0.119
2 layers (11+5) hidden neurons, hyperbolic function	0.036	0.106

1-Day Prediction Using 5-Step

Model	Relative MSE	MAPE
8 hidden neuron, hyperbolic function	0.031	0.108
9 hidden neuron, hyperbolic function	0.033	0.107
10 hidden neuron, hyperbolic function	0.034	0.106
8 hidden neuron, logistic function	0.034	0.110
2 layers (8+4) hidden neurons, hyperbolic function	0.034	0.106

1-Day Prediction Using 6-Step

Model	Relative MSE	MAPE
8 hidden neuron, hyperbolic function	0.041	0.109
9 hidden neuron, hyperbolic function	0.033	0.104
10 hidden neuron, hyperbolic function	0.033	0.104
9 hidden neuron, logistic function	0.030	0.105
2 layers (9+4) hidden neurons, logistic function	0.039	0.106

1-Day Prediction Using 7-Step

Model	Relative MSE	MAPE
9 hidden neuron, hyperbolic function	0.021	0.081
8 hidden neuron, hyperbolic function	0.018	0.079
7 hidden neuron, hyperbolic function	0.020	0.080
8 hidden neuron, logistic function	0.020	0.078
2 layers (8+4) hidden neurons, hyperbolic function	0.020	0.078

All Predictions Using Best Configurations

Model	Steps Used	Prediction Horizon	Relative MSE	MAPE
3 hidden neuron, hyperbolic function	1	1	0.048	0.136
		7	0.096	0.123
		30	0.176	0.221
		90	0.189	0.212
		365	0.210	0.200
7 hidden neuron, logistic function	2	1	0.043	0.129
		7	0.096	0.120
		30	0.168	0.207
		90	0.174	0.177
		365	0.180	0.185
9 hidden neuron, hyperbolic function	3	1	0.033	0.116
		7	0.088	0.119
		30	0.167	0.199
		90	0.170	0.175
		365	0.180	0.184
11 hidden neuron, hyperbolic function	4	1	0.033	0.113
		7	0.088	0.120
		30	0.162	0.200
		90	0.166	0.173
		365	0.168	0.176
8 hidden neuron, hyperbolic function	5	1	0.031	0.108
		7	0.088	0.122
		30	0.161	0.196
		90	0.163	0.175
		365	0.181	0.177
9 hidden neuron, logistic function	6	1	0.030	0.105
		7	0.087	0.120
		30	0.158	0.163
		90	0.160	0.176
		365	0.180	0.169
8 hidden neuron, hyperbolic function	7	1	0.018	0.079
		7	0.075	0.119
		30	0.151	0.165
		90	0.157	0.173
		365	0.164	0.159

Performance of Ensembles

Model	Relative MSE	MAPE
1-Day Prediction		
9 hidden neurons, logistic output function	0.019	0.075
7-Day Prediction		
9 hidden neurons, logistic output function	0.066	0.114
30-Day Prediction		
9 hidden neurons, logistic output function	0.130	0.1511
90-Day Prediction		
9 hidden neurons, logistic output function	0.151	0.162
365-Day Prediction		
9 hidden neurons, logistic output function	0.148	0.147

Prediction with All Significant Variables

1-Day Prediction Using 1-Step

Model	Relative MSE	MAPE
7 hidden neuron, hyperbolic function	0.018	0.074
8 hidden neuron, hyperbolic function	0.018	0.075
7 hidden neuron, logistic function	0.019	0.076
2 layers (7+3) hidden neurons, hyperbolic function	0.023	0.081

1-Day Prediction Using 2-Step

Model	Relative MSE	MAPE
7 hidden neuron, hyperbolic function	0.018	0.075
8 hidden neuron, hyperbolic function	0.019	0.075
7 hidden neuron, logistic function	0.021	0.076
2 layers (7+3) hidden neurons, hyperbolic function	0.022	0.079

1-Day Prediction Using 3-Step

Model	Relative MSE	MAPE
8 hidden neuron, hyperbolic function	0.019	0.075
7 hidden neuron, hyperbolic function	0.018	0.075
6 hidden neuron, hyperbolic function	0.017	0.073
5 hidden neuron, hyperbolic function	0.017	0.072
5 hidden neuron, logistic function	0.018	0.073
2 layers (5+2) hidden neurons, hyperbolic function	0.022	0.078

1-Day Prediction Using 4-Step

Model	Relative MSE	MAPE
5 hidden neuron, hyperbolic function	0.017	0.075
6 hidden neuron, hyperbolic function	0.019	0.074
4 hidden neuron, hyperbolic function	0.020	0.075
5 hidden neuron, logistic function	0.020	0.073
2 layers (5+2) hidden neurons, hyperbolic function	0.021	0.077

1-Day Prediction Using 5-Step

Model	Relative MSE	MAPE
5 hidden neuron, hyperbolic function	0.028	0.080
6 hidden neuron, hyperbolic function	0.017	0.074
7 hidden neuron, hyperbolic function	0.022	0.078
6 hidden neuron, logistic function	0.020	0.073
2 layers (6+3) hidden neurons, hyperbolic function	0.022	0.076

1-Day Prediction Using 6-Step

Model	Relative MSE	MAPE
8 hidden neuron, hyperbolic function	0.017	0.073
7 hidden neuron, hyperbolic function	0.022	0.079
9 hidden neuron, hyperbolic function	0.031	0.087
8 hidden neuron, logistic function	0.020	0.074
2 layers (5+4) hidden neurons, hyperbolic function	0.026	0.077

1-Day Prediction Using 7-Step

Model	Relative MSE	MAPE
8 hidden neuron, hyperbolic function	0.024	0.078
7 hidden neuron, hyperbolic function	0.024	0.079
9 hidden neuron, hyperbolic function	0.031	0.083
6 hidden neuron, hyperbolic function	0.022	0.074
5 hidden neuron, hyperbolic function	0.030	0.078
6 hidden neuron, logistic function	0.025	0.073
12 hidden neuron, hyperbolic function	0.023	0.076
15 hidden neuron, hyperbolic function	0.020	0.079
2 layers (10+5) hidden neurons, hyperbolic function	0.022	0.076
18 hidden neuron, hyperbolic function	0.029	0.083
2 layers (10+6) hidden neurons, hyperbolic function	0.024	0.074
4 hidden neuron, hyperbolic function	0.023	0.076
3 hidden neuron, hyperbolic function	0.021	0.076
16 hidden neuron, hyperbolic function	0.021	0.081
14 hidden neuron, hyperbolic function	0.024	0.079
15 hidden neuron, logistic function	0.025	0.076
16 hidden neuron, logistic function	0.017	0.073

All Predictions Using Best Configurations

Model	Steps Used	Prediction Horizon	Relative MSE	MAPE
7 hidden neuron, hyperbolic function	1	1	0.018	0.074
		7	0.095	0.119
		30	0.144	0.150
		90	0.169	0.160
		365	0.171	0.149
7 hidden neuron, hyperbolic function	2	1	0.017	0.074
		7	0.090	0.111
		30	0.144	0.152
		90	0.167	0.151
		365	0.169	0.151
5 hidden neuron, hyperbolic function	3	1	0.017	0.072
		7	0.085	0.122
		30	0.140	0.153
		90	0.160	0.161
		365	0.168	0.150
5 hidden neuron, hyperbolic function	4	1	0.017	0.075
		7	0.082	0.117
		30	0.138	0.158
		90	0.158	0.159
		365	0.165	0.151
6 hidden neuron, hyperbolic function	5	1	0.017	0.074
		7	0.080	0.116
		30	0.135	0.153
		90	0.155	0.155
		365	0.162	0.151
8 hidden neuron, hyperbolic function	6	1	0.017	0.073
		7	0.075	0.112
		30	0.132	0.159
		90	0.154	0.160
		365	0.162	0.155
16 hidden neuron, logistic function	7	1	0.017	0.073
		7	0.075	0.119
		30	0.134	0.146
		90	0.150	0.160
		365	0.161	0.150

Performance of Ensembles

Model	Relative MSE	MAPE
1-Day Prediction		
9 hidden neurons, logistic output function	0.012	0.061
7-Day Prediction		
9 hidden neurons, logistic output function	0.060	0.103
30-Day Prediction		
9 hidden neurons, logistic output function	0.129	0.130
90-Day Prediction		
9 hidden neurons, logistic output function	0.150	0.150
365-Day Prediction		
9 hidden neurons, logistic output function	0.145	0.132

VITAE

PERSONAL INFORMATION

Name	:	**UNEB GAZDER**
Nationality	:	PAKISTANI
Languages	:	English, Urdu.
E: mail address	:	*unebgazdar@gmail.com*
Permanent address	:	Flat # 9, 6th Floor, Rc4/409, Ranchoreline, Karachi, Pakistan.

QUALIFICATIONS

Education	:	**M.E.M. (Construction)**
		NED University of Engineering and Technology Karachi, Pakistan, December 2008
		BE (Civil Engineering)
		NED University of Engineering and Technology Karachi, Pakistan, December 2005
Software	:	AutoCAD, MS Project, @Risk

EXPERIENCE

Company	**:**	**King Fahd University of Petroleum and Minerals**
Designation	:	Lecturer B
Duration	:	From Feb. 2011 till date
Place	:	Ad-Dhahran, KSA
Company	**:**	**NED University of Engineering & Technology**
Designation	:	Assistant Professor (On Study Leave)
Duration	:	From Sep. 2008 till date
Place	:	Karachi, Pakistan
Company	**:**	**Consteel Construction**
Designation	:	Design Engineer
Duration	:	From Dec. 2005 to Jun. 2006
Place	:	Karachi, Pakistan

SEMINARS, CONFERENCES AND WORKSHOPS ATTENDED

- Seminar on "Safety of Road Users – Challenges & Solutions" organized by Ministry of Communications Pakistan and National Highway and Motorway Police held on 16th November, 2009 at Marquee, Pearl Continental Hotel Karachi, Pakistan.
- 5th seminar on Urban & Regional Planning "Urban Planning in Market Economy Situation" organized by Department of Architecture and Town Planning, NED University of Engineering and Technology held on May 08, 2010 at NED auditorium Karachi, Pakistan.

- First International Conference on Infrastructure Engineering In Developing Countries, Organized by US-AID – NED University held on 1st to 3rd July 2010 at Ramada Plaza Hotel, Karachi Pakistan.
- First National Workshop on Construction Materials, held at 27th November 2010 in NED University of Engineering & Technology, Karachi Pakistan.
- 15th international conference on computer modelling and simulation, UKSim2013, held on April 10-12, 2013, Cambridge UK

RELATED RESEARCH

- Research paper on "Traffic Forecasting for King Fahd Causeway: Comparison of Parametric Technique with Artificial Neural Networks", published in International Journal of Advancements Civil Structural and Environmental Engineering – IJACSE, Volume 1, Issue 1, on 15th May 2013, pp. 75 – 79.
- Research paper on "Traffic Forecasting for King Fahd Causeway Using Artificial Neural Networks", published in proceedings of 15th international conference on computer modelling and simulation, UKSim2013, held on April 10-12, 2013, Cambridge UK.