Modelling Trip Generation using Adaptive Neuro-Fuzzy Inference System in Comparison with Traditional Multiple Linear Regression Approach

Sameer Abu-Eisheh, Mohammad Irshaid

Department of Civil Engineering, An-Najah National University, Nablus, Palestine.

e-mail: sameeraa@najah.edu; mohammad.irshaid@outlook.com

***Abstract* - Development of trip generation models has been conducted mainly using the traditional Multiple Linear Regression approach, which sometimes might not necessarily result in appropriate models, especially with existence of many interrelated and complex relationships among several related socioeconomic variables. This study investigates the feasibility of using a relatively new approach, the Adaptive Neuro-Fuzzy Inference System, and compares the results with those using the traditional approach. This is conducted by developing a home-based general trip generation model for one of the Palestinian urban areas. The comparison between the two methods outcome and the associated validation results is done using the R-squared, RMSE, and MAE measures. The Adaptive Neuro-Fuzzy Inference System was found to be a useful tool and a promising technique for modelling household trip generation, which is shown to outperform the traditional approach, with more accurate results and closer predictions to actual values. Further exploration of the new approach in transportation studies is recommended.**

Keywords - trip generation, modelling, adaptive neuro-fuzzy inference system, multiple linear regression

I. INTRODUCTION

Trip-based approaches have been used for modelling travel behavior and in estimating the number of trips produced, attracted or exchanged between traffic zones, and distributed on the modes and routes in urban road networks. This have been conducted mainly through the four-step travel demand forecasting process, or what is known as the 'Urban Transportation Modeling System'. This process usually transfers urban activities into number of trips, and attempts to quantify the relationship between urban activities and travel, through modelling trip generation, trip distribution, mode choice, and traffic assignment.

The traditional approaches, mainly the Multiple Linear Regression (MLR), are still used almost universally. The MLR models usually capture the correlation patterns and study the relationship between variables that are considered as the determinants of trip maker behaviour (explanatory variables), such as the household characteristics, and variables that are considered to estimate the number of trips as indicators of travel behavior [1]. However, despite the fact that this approach is widely used, and could be easily constructed, estimated, and interpreted, sometimes it might not result in appropriate models, especially with the existence of interrelated and complex relationships among several related socioeconomic variables. In similar situations, the use of advanced modelling techniques, based on the artificial intelligence, have attracted much attention, which could enable effective elaboration for modelling such situation.

This study is devoted to investigating the feasibility of using a relatively new method for data analysis called the Adaptive Neuro-Fuzzy Inference System (ANFIS), as an alternative for the traditional MLR. The results using the two approaches on the same data set are compared. The application and comparison have been explored within the Palestinian context through the development of trip generation model for Salfit urban area. Through this study, a home-based general trip generation model (denoted by ALLTRIPS) was developed for Salfit, taken as the study area, with the intension to estimate the total number of trips generated by a household during a typical working day. This model was developed using the two approaches, MLR & ANFIS, followed by conducting a comparative analysis among their performance, in an attempt to seek the more accurate modelling approach.

II. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

The ANFIS is an artificial intelligence-based approach, first proposed by Jang in 1993, which is considered as an advanced modelling technique that has the capability of dealing with nonlinear and highly complex systems [2]. ANFIS integrates the best features of the Artificial Neural Network (ANN) and the Fuzzy Inference System (FIS) into a single framework.

The ANN is a machine learning-based system, developed for information processing, and usually used as a data analytical tool. It is structured by interconnected artificial neurons arranged in a systematic manner to form a layer pattern. Each layer is composed of several processing neurons that have specific function, which could be adaptive or fixed [3]. The ANN is not based on specific rules, but rather it is developed through trial and error procedure across successive calculations. Such system "learns" to perform tasks by considering examples (dataset), generally without being programmed with any task-specific rules. The ANN can adapt itself to self-organize its structure, when the sample input-output training is presented.

The FIS, on the other hand, uses the fuzzy logic for modelling complex systems based on degrees of truth, represented by the degree of membership value in a fuzzy set, rather than the usual true or false (0 or 1) in the classic set. This approach has been proven to be an effective technique for dealing with variables that are linguistically specified, such as low, medium, or high, which may be defined by fuzzy sets [4]. For each variable, different Membership Functions (MFs) could be identified, which would reflect a specific fuzzy set. The MF is a curve that defines how each point in the input range is assigned to a membership value between 0 and 1. MFs can have several shapes, such as triangle, trapezoidal, and gaussian. However, the basic structure of a FIS consists of three conceptual components: a rule base, which contains a selection of fuzzy rules; a database which defines the MFs used in the fuzzy rules; and a reasoning mechanism, which performs the inference procedure upon the rules to derive an output [1]. Takagi-Sugeno FIS type was found to be widely used in the application of ANFIS method.

The ANFIS, hence, is an ANN that is functionally based on the model of Takagi-Sugeno FIS. ANFIS constructs a FIS from given input-output training dataset, and uses the learning capability of ANN for optimizing the fuzzy rules and MFs parameters, by training the system several times (epochs) using specific learning algorithm. This learning algorithm seeks to minimize the measure of error, usually the Root Mean Squared Error (RMSE), between the observed and predicted values. Two types of learning algorithm could be used, the backpropagation (gradient descent method for all parameters), and the hybrid (combines backpropagation for the input MFs, and least squares estimation for the output MFs). However, for more details Jang [2], Jang et al. [5], and Rutkowska [6] provide full insight into the working mechanism of these algorithms.

III. LITERATURE REVIEW

Several studies in different scientific disciplines have proved the effectiveness of ANFIS in modelling complex nonlinear systems over the conventional techniques. In transportation research, for example, Stojčić [7] reviewed the applications of the ANFIS from the year 1993 till 2018, considering many areas, such as vehicle routing, traffic control, safety, modelling, and traffic congestion; and concluded that it represents a promising modelling method, that able to show better performance compared with the traditional methods. In addition to the studies of Tortum et al. [8], Andrade et al. [9], and S. et al. [10], for modelling travel mode choice, who raised with the same conclusion of Stojčić [7].

In trip generation modelling, the application of ANFIS was not widely considered, and found to be limited to specific few studies, such as those of Přibyl & Goulias [1], Ahmadpour et al. [11], and Mahdavi & Mamdoohi [12], who all developed trip generation models for different purposes using ANFIS, and found that ANFIS is a potentially better data analytic method that performs more accurately (higher R-squared and less RMSE) than the MLR approach.

In Palestine, few specific studies that concerned with the development of trip generation models have been performed, as mentioned earlier. The most relevant here were the study of Dodeen [13], who utilized the MLR technique for developing trip generation models for Jericho City, based on the household socioeconomic characteristics; and the study of Amer [14], who examined the potential for spatial transferability of Dodeen models from Jericho City to Salfit City. The relevance of this study comes from the fact that it was pioneering and the first in its type that considers a comparison among the performance of trip generation models developed using ANFIS, and that developed using MLR.

IV. STUDY AREA AND DATA COLLECTION

Salfit City was selected as a case for this study, due to the availability of thoroughly and systematically collected data through the recent study conducted by Amer for the city [14]. These collected data included 309 households, which were divided into two sets, a set of 256-households considered for the estimation and the training process of the desired model (representing the minimum sample size of 10% of the total number of households), and the balanced set of 53-households considered for validation and testing purposes. Each collected household sample included data regarding the socioeconomic characteristics, represented by 12 explanatory variables, used as model inputs, which were expected to be relevant based on previous studies; and the associated total number of trips generated (ALLTRIPS as model output), as described in Table I.

V. METHODOLOGY

In order to achieve the desired objective of this research, and evaluate the more suitable modelling approach, the ALLTRIPS model was developed using the two competing approaches, MLR and ANFIS, by following three steps as described hereafter.

A. Developing ALLTRIPS Model Using MLR Approach

In this step, the backward elimination stepwise regression approach, based on the ordinary least-squares estimation technique, was considered for the estimation and calibration process (i.e., estimating the associated coefficients with the most relevant explanatory variables). Under this approach, the fitness of the regression analysis was evaluated based on different statistical tests that would quantify the significance of each explanatory variable that was assumed to be relevant.

To ensure the goodness of the developed model, several relevant statistical tests were conducted, which are 1) the Pearson's correlation and the Variance Inflation Factor (VIF) to check for multicollinearity, 2) the t-test to assess the significance of the individual regression coefficients, 3) the F-test for testing the overall significance of the model, and 4) the coefficient of determination (R-squared), the Root Mean Squared Error (RMSE), and the Mean Absolute Error (MAE) to assess the developed model goodness of fit.

B. Developing ALLTRIPS Model Using ANFIS Approach

The training dataset (i.e., 256-households sample), and the input explanatory variables were exactly the same as considered in the previous step. However, in this approach, two steps were considered for building the required model:

TABLE I. MODEL INPUTS AND OUTPUT VARIABLES DESCRIPTION

Variable	Description	Average	Standard Deviation	Range
The Selected Explanatory Variables (Model Inputs)				
SIZE	Number of persons (household size)	3.99	1.749	8
EMP	Number of persons in the household who are employed	1.38	0.899	4
EDU	Number of persons in the household who are receiving education	1.63	1.529	7
AGE_A	Number of persons in the household who are under 16 years	1.24	1.373	6
AGE_B	Number of persons in the household who are between 17 and 30 years	1.14	0.977	4
AGE_C	Number of persons in the household who are between 31 and 50 years	0.92	0.805	2
AGE_D	Number of persons in the household who are between 51 and 64 years	0.51	0.728	3
AGE_E	Number of persons in the household who are above 65 years	0.19	0.485	2
DRIVE	Number of persons in the household who are licensed drivers	1.18	0.964	6
CAR	Number of cars owned by the household	0.54	0.605	3
INC	Monthly household Income (Thousand New Israeli Shekel)	4.740	2.993	29.500
HHTYP	House Type: 1 if Independent Residence, 0 if Apartment	83.8% Independent Residence		
The Selected Dependent Variable (Model Output)				
ALLTRIPS	Total number of daily general trips generated by the household	6.42	3.050	17 (1-18)

1) Generating the Initial FIS: The Neuro-Fuzzy Designer in MATLAB was utilized for this purpose. An initial Takagi-Sugeno FIS could be generated using the Grid Partitioning function that is embedded in the above tool, which allows the user to determine the desired number and type of the input MFs that were associated with each input variable.

This function splits the range of each input variable into equal intervals based on the selected number of the input MFs, and creates particular decision rules. One fuzzy rule along with one linear output MF will be created for each input MFs combination. For example, if there are three input variables with three MFs for each variable, 27 fuzzy rules (3MFs x 3 MFs x 3MFs = 27) along with 27 output linear MFs will be created automatically.

2) Optimizing the FIS Using Learning Algorithms: The learning algorithm would use the training dataset to optimize the fuzzy rules and the MFs parameters, by training the FIS several times (epochs), until reaching a prespecified measure of error (RMSE), or a selected number of training epochs, which could be satisfied first.

In order to obtain the optimum configuration of the ANFIS, with the minimum possible RMSE and MAE, several options were considered, as per the following:

- The number of inputs MFs: three functions for each input variable, as one MF may create estimation errors, and more than three may increase the computational cost of ANFIS model.
- The type of inputs MFs: gaussian, trapezoidal, and triangle, which are the popular types.
- The type of learning algorithm: backpropagation and hybrid algorithms, which are provided by the MATLAB.
- The number of training epochs: 1, 5, 10, 50, 100, 200, 500, and 1000 epochs, which were selected based on personal discretion (trial and error procedure), as there is no simple way to determine what should be the optimal one.

C. Conducting Comparative Analysis

Three statistical performance measures were considered for the comparison between the developed ALLTRIPS model per each approach, namely the R-squared, RMSE, and MAE. These measures were also used to assess the accuracy of each model. Moreover, the validation assessment was achieved by comparing the predicted number of trips with the actual values, for the additional 53-households sample, which were not used in the estimation or training process. The better performing and more suitable approach was then determined based on several evaluation criteria, such as the higher value of R-squared, the lower RMSE & MAE, and the much closer outputs to the actual values.

VI. Model Development and Comparison

The estimated number of trips include all possible trips that are generated by a household, regardless of their types or purposes. This number should lie between one trip per day as a minimum possible value, and up to 18 trips per day as a maximum value. the following illustrate the development process for this model considering both the MLR and ANFIS.

A. ALLTRIPS Model Using MLR

Equation (1) suggests the best explanatory variables, among others, that would explain the number of total daily general trips generated by a household. The regression results for this model are summarized in Table II.

ALLTRIPS = 2.388 + 1.046 EMP + 1.237 EDU + 0.424 DRIVE … (1)

The ALLTRIPS tends to increase with the increase in the household persons who are employed, receiving education, or having a driving license. This directly proportional and positive relationship is logical and as expected. The t-values for each coefficient were larger than two, which indicates that the selected explanatory variables are statistically significant at 99.99% level of significance. Moreover, the large F-value indicates that the null hypothesis that the explanatory variables EMP, EDU, and DRIVE have no impact on the ALLTRIPS is rejected statistically at the 99.99% level of significance (i.e., these variables jointly affect ALLTRIPS).

B. ALLTRIPS Model Using ANFIS

Referring to Table III, the optimum structure of the ANFIS was achieved first at three gaussian MFs for each input variable (i.e., EMP, EDU, and DRIVE), hybrid learning algorithm, and 1000 training epochs. An initial FIS was first developed, using the same input variables that were used in the regression approach. Three gaussian MFs for each input variable were then identified. Consequently, 27 fuzzy rules along with 27output linear MFs were created.

TABLE II. REGRESSION RESULTS FOR THE ALLTRIPS MODEL

Variables	**Coefficient**	**Standard Error**	**t-value**	**Sig.**	**VIF**
Intercept	2.388	0.223	10.721	0.000	
EMP	1.046	0.132	7.945	0.000	1.287
EDU	1.237	0.073	16.926	0.000	1.029
DRIVE	0.424	0.123	3.452	0.001	1.262
R-squared: 0.659		**Adjusted R-squared: 0.654**		**RMSE: 1.7112**	
MAE: 1.3882		**F-value: 161.982**		**Sample Size: 256**	

TABLE III. RESULTED RMSE AND MAE FOR DIFFERENT ANFIS MODEL CONFIGURATIONS – ALLTRIPS

Learning Algorithm	Index	Training Epochs							
		1	5	10	50	100	200	500	1000
Gaussian MFs (3 MFs for each input)									
Backpropagation	RMSE	6.9612	6.8865	6.7931	6.0037	4.7763	2.4853	1.6729	1.6653
	MAE	6.3188	6.2504	6.1646	5.4286	4.2249	1.9868	1.3269	1.3230
Hybrid	RMSE	1.5114	1.5108	1.5101	1.5068	1.5028	1.4970	1.4895	1.4880*
	MAE	1.1504	1.1486	1.1467	1.1393	1.1340	1.1288	1.1220	1.1203*
Trapezoidal MFs (3 MFs for each input)									
Backpropagation	RMSE	6.9590	6.8757	6.7718	5.9453	4.8184	2.9659	1.6110	1.5769
	MAE	6.3171	6.2419	6.1476	5.3760	4.2410	2.2799	1.2937	1.2599
Hybrid	RMSE	1.5453	1.5451	1.5448	1.5434	1.5400	1.5288	1.5288	1.5288
	MAE	1.2063	1.2061	1.2059	1.2017	1.1943	1.1825	1.1825	1.1825
Triangle MFs (3 MFs for each input)									
Backpropagation	RMSE	6.9604	6.8825	6.7854	6.0190	5.0845	3.3590	1.6817	1.5533
	MAE	6.3181	6.2467	6.1573	5.4380	4.5152	2.7353	1.3532	1.2389
Hybrid	RMSE	1.5289	1.5196	1.5178	1.5152	1.5130	1.5130	1.5130	1.5130
	MAE	1.1896	1.1751	1.1689	1.1649	1.1616	1.1616	1.1616	1.1616

*The lowest possible error obtained by the optimum configuration

The associated parameters with the inputs and outputs MFs were optimized at the end using the hybrid learning algorithm for 1000 training epochs.

Table IV summarizes the results associated with the optimized FIS for estimating the ALLTRIPS. Fig. 1 reflects the equivalent ANFIS architecture for the developed FIS. Fig. 2 illustrates the initial and the optimized gaussian MFs (before and after training) for each input variable.

TABLE IV. ANFIS OPTIMUM CONFIGURATION SUMMARY

R-Square: 74.18%	**RMSE: 1.4880**
MAE: 1.1203	**Sample Size: 256**
ANFIS Sum of Squares:	1628.324
Residual Sum of Squares:	566.786
Total Sum of Squares:	2195.109

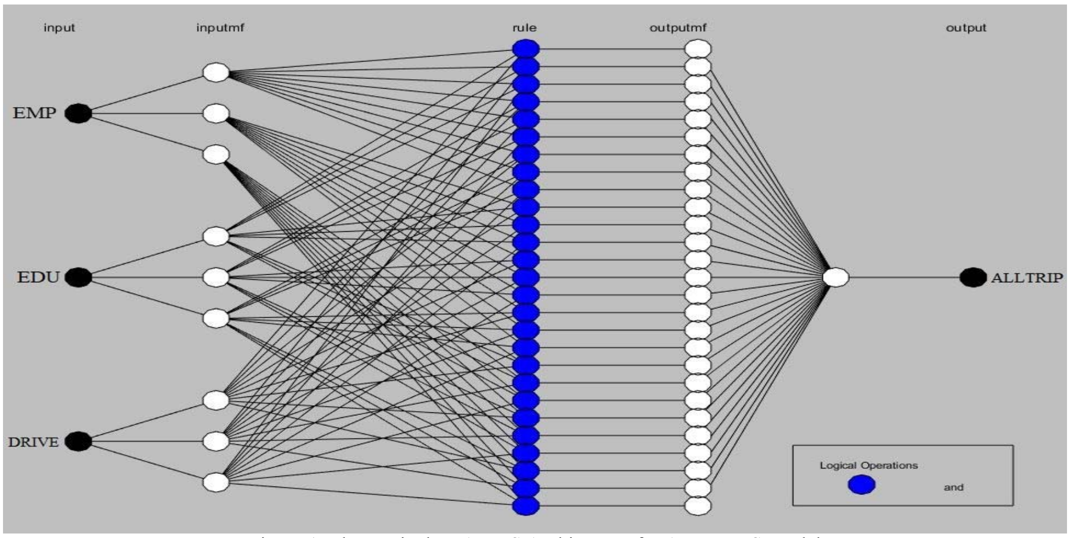

Figure 1. The Equivalent ANFIS Architecture for ALLTRIPS Model

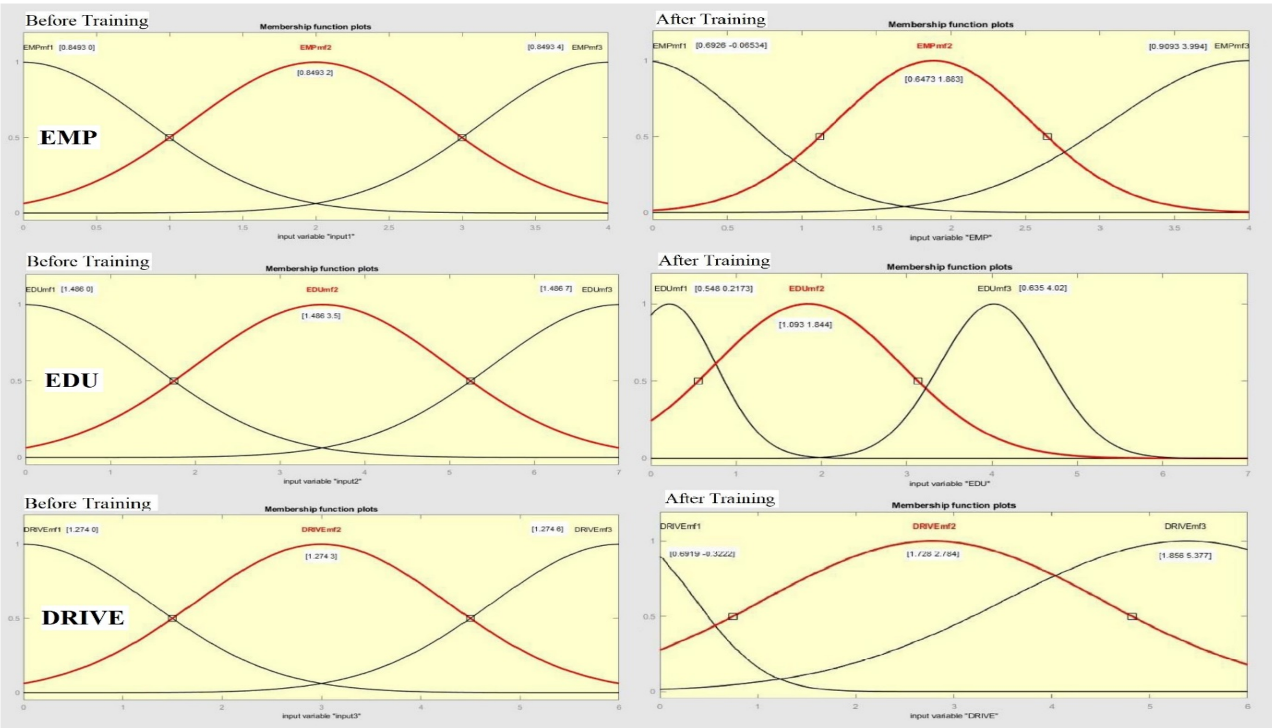

Figure 2. The Initial and The Optimized Gaussian MFs For Each Input Variable

C. Models Comparison and Validation

Table V illustrates a comparison between the performance of the ALLTRIPS model developed using the MLR approach, and that developed using the ANFIS approach. It could be noticed that the ANFIS was able to develop more accurate model with better performance compared with the MLR approach. The R-squared value was raised by 8.33%, from 65.85% in MLR to 74.18% in ANFIS. Moreover, a considerable reduction in the RMSE, the MAE, and the Residual Sum of Squares (RSS) could be noticed, where the percentage difference for each of these measures of error were -13.04%, -19.30%, and -24.39%, respectively.

TABLE V. ALLTRIPS PERFORMANCE COMPARISON

	MLR		ANFIS	Difference	% Difference
RMSE	1.7112	>	**1.488**	-0.2232	-13.04%
MAE	1.3882	>	**1.1203**	-0.2679	-19.30%
R-squared	65.85%	<	**74.18%**	+8.33%	--
RSS	749.604	>	**566.786**	-182.812	-24.39%

For validation purposes, testing the 53-household sample, the predicted results and the actual values for both modelling approaches were within acceptable conformity, with better results associated with ANFIS. Table VI summarizes a comparison of several statistical measures between the actual and predicted values of this sample. The means of the predicted values using MLR and ANFIS were compared with the mean of the actual values using the t-test. The results indicated the two means are not significantly different from the actual mean at 90% level of significance. The predicted mean and standard deviation using ANFIS were closer to the actual values than with the MLR. Furthermore, the difference between the summation of the actual ALLTRIPS and of the related predicted values using ANFIS (9.34 trip) were smaller as compared with the MLR (19.65 trips).

TABLE VI. ACTUAL AND PREDICTED ALLTRIPS DESCRIPTIVE STATISTICS

	Actual		MLR		ANFIS
Mean	**6.81**		**6.44**		**6.64**
St. Deviation	3.563		2.639		3.026
Sum	361		341.35		351.66
Difference	0		**19.65**		**9.34**
Count	53		53		53
t-test	t-critical =1.66	>	t-statistic = 0.608	>	t-statistic = 0.265
	α = 0.1	<	P-value = 0.5448	<	P-value = 0.7917

VII. CONCLUSION

Through this study, the robust comparison and validation process reveals that the ANFIS represents a promising modelling technique, that can be a good competitor for MLR approach, especially, especially when dealing with interrelated and complex relationships among several, including behavioural socioeconomic, variables. It was found that the ANFIS can be used successfully for modelling ALLTRIPS, and able to outperform the traditional MLR approach, with more accurate and closer predictions to actual values. The ANFIS is potentially better data analytic method for complex systems, which needs to be explored more in-depth and compared to more sophisticated regression techniques that are already in use in transportation researches.

It is to be stated that, and through this study, a particular emphasis was given to the effect of different design options in building the desired ANFIS model. The development process involves 48 possible configurations, and the one with the lowest error was selected to represent the optimum ANFIS structure. It was noticed that the use of gaussian MFs along with hybrid learning algorithm would usually minimize the measure of error faster, with less training epochs, than other options.

REFERENCES

[1] O. Přibyl & K. Goulias, “Application of Adaptive Neuro-Fuzzy Inference System to Analysis of Travel Behavior,” Transportation Research Record: Journal of the Transportation Research Board, No.1854,. Washington, D.C., PP: 180-188, 2003.

[2] R. Jang, “ANFIS: Adaptive Network-Based Fuzzy Inference Systems”, IEEE Transactions on Systems, Man, and Cybernetics, Vol 23, No. 3, PP: 665–685, 1993.

[3] V. Profillidis & G. Botzoris, “Modeling of Transport Demand: Analyzing, Calculating, and Forecasting Transport Demand,” Elsevier Inc., Science Publishers, 2018.

[4] R. Jang & N. Gulley, “Fuzzy Logic Toolbox™ User's Guide, MATLAB R2018b,” The MathWorks, Inc., 2018.

[5] R. Jang, C. Sun, & E. Mizutain, “Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence,” Prentice Hall, Inc., 1997.

[6] D. Rutkowska, “Neuro-Fuzzy Architectures and Hybrid Learning,” Springer-Verlag Berlin Heidelberg GmbH, 2002.

[7] M. Stojčić, “Application of The ANFIS Model in Road Traffic and Transportation: A Literature Review From 1993 To 2018,” Operational Research in Engineering Sciences: Theory and Applications, Volume-1, Issue-1, PP: 40-61, 2018.

[8] A. Tortum, N. Yayla, & M. Gökdag, “The Modeling of Mode Choices of Intercity Freight Transportation with The Artificial Neural Networks and Adaptive Neuro-Fuzzy Inference System,” Expert Systems with Applications. Vol.36, Elsevier Ltd, PP: 6199-6217, 2009.

[9] K. Andrade, K. Uchida, & S. Kagaya, “Development of Transport Mode Choice Model by Using Adaptive Neuro-Fuzzy Inference System,” Transportation Research Record: Journal of the Transportation Research Board, No. 1977, Washington, D.C., PP: 8-16, 2006.

[10] M. S., C. R. Sekhar, & E. Madhu, “Development of Neuro-Fuzzy-Based Multimodal Mode Choice Model for Commuter in Delhi,” IET Intelligent Transport Systems, Volume 13, Issue 2, PP: 243-251, 2019.

[11] M. Ahmadpour, W. Yue, & M. Mohammadzaheri, “Neuro-fuzzy Modelling of Workers Trip Production,” 2009, Retrieved from (Accessed: 2019, Mar 20): https://www.semanticscholar.org/paper/Neuro-fuzzy-modelling-ofworkers-trip-production-Ahmadpour Yue/67a562de25044a1bf3157a 166b2488ff478f268f.

[12] R. Mahdavi & R.Mamdoohi, “Comparison of Linear Regression Model and ANFIS Model for Prediction of Arbitrary Trip Attraction: Case Study Qazvin City,” Presented at the 17th International Conference of Traffic and Transportation Engineering, Tehran, Iran, 2018.

[13] A. Dodeen, “Developing Trip Generation Models Utilizing Linear Regression Analysis: Jericho City as a Case Study,” Master thesis, An-Najah National University, Nablus, Palestine, 2014.

[14] A. Amer, “Transferability of Trip Generation Models for Palestinian Cities: Jericho and Salfit,” Master thesis, An-Najah National University, Nablus, Palestine, 2017.