MODELING OF DRIVER LANE CHOICE BEHAVIOR WITH ARTIFICIAL NEURAL NETWORKS (ANN) AND LINEAR REGRESSION (LR) ANALYSIS ON DEFORMED ROADS

Metin Mutlu Aydın[1], Mehmet Sinan Yıldırım[2] Orhan Karpuz[1] Kiarash Ghasemlou[3]

[1]Department of Civil Engineering, Gumushane University, Gumushane, Turkey
[2]Department of Civil Engineering, Gediz University, Izmir, Turkey
[1]Department of Civil Engineering, Gumushane University, Gumushane, Turkey
[3]Department of Urban and Regional Planning, Istanbul Technical University, Istanbul, Turkey

ABSTRACT

In parallel to the economic developments, the importance of road transportation was significantly increased in Turkey. As a result of this, long-distance freight transportation gains more importance and hence numbers of the heavy vehicles were significantly increased. Consequently, road surface deformations are observed on the roads as the increasing freight transportation and climatic conditions influence the road surface. Therefore, loss of functionality of the road surface is observed and drivers are much prone to accident due to their driving characteristics as they can have more tendencies to change their lanes not to pass through the deformation area. In this study, the lane changing behaviors of the drivers were investigated and both Artificial Neural Network (ANN) and Linear Regression (LR) models were proposed to simulate the driver behavior of lane changing who approach to a specific road deformation area. The potential of ANN model for simulating the driver behavior was evaluated with successive comparison of the model performances with LR model. While there was a slight performance increase for the ANN model with respect to LR model, it is quite evident that, ANN models can play an important role for predicting the driver behavior approaching a road surface deformation. It can be said that, approaching speed plays an important factor on the lane changing behavior of a driver. This can be criticized by the fact that, drivers with high approaching speeds more likely pass through the deformation to avoid the accidents while changing their lanes with a high speed.

KEYWORDS

Lane Changing Behavior, Artificial Neural Network (ANN) Model, Linear Regression (LR) Model, Road Surface Deformations.

1. INTRODUCTION

Prediction of motor vehicle behaviors approaching to a specific surface deformation on a highway is important for the safety of the transportation as the lane changing may results in traffic accidents. The road deformations should be considered as a significant hazard to the vehicles especially driving at high speeds. In this condition many vehicles try to avoid the deformation by rapid lane changing so this results in losing the control of the vehicle. Hence, head-to-head

collisions with the vehicles driving on the opposite lane are occurred. The ability to predict the behavior of a motor vehicle approaching to a deformation with dependent variables, such as speed, location, driver and vehicle characteristics allow to understand the impact of the deformations on the traffic accidents.

Artificial Neural Network (ANN) is a tool for modeling the statistical data. It is an effective way for dealing with the non-linear relations between system inputs and outputs. A neural network can handle the complex relationships inside the data pattern, generalize the data and come by a solution with a plausible error. A typical ANN application can be used effectively for recognition, classification, estimation and optimization in engineering problems with implementing a specific pattern recognition technique for a statistical dataset. Although, some traditional methods like regression analysis can be implemented to transportation problems [1]–[2], many studies indicate that ANN applications offer more accuracy in modeling [3]. Not only the ANN has several advantages such that there is no need of selecting a regression model to the data but also the ANN structure generally fits well with the data and shows a reliable performance. There are various ANN applications in the scope of the transportation engineering such as predicting the traffic safety, traffic forecasting, and transportation planning and traffic operations.

In the scope of this study, a multi-layer ANN and LR model was implemented to predict the driving behaviors approaching to road surface deformations. The training of the network was performed by a back-propagation algorithm with a data set consist of the vehicles type, approach speed, gender of drivers and the decision of the driver (changed his/her lane or continued through the deformation). Thus, the purpose of this study were to identify the most significant dependent factors that determine the possibility of lane changing approaching to a deformation and to evaluate the ANN and LR performance with the traditional regression methods.

Beside many studies, predicting the driver and vehicle characteristics is important for a better understanding the causes of accidents. Many studies are available in literature with dealing this topic. [4] implemented an ANN application for predicting the contribution of the environmental factors, driver characteristics and road conditions to the road accidents. [5] compared the capability of a neural network (NN) model with multiple discriminant analysis and logistic regression to predict vessel accidents. [6] proposed an ANN model for prediction of the accidents (number of accidents, injuries and deaths).The comparison of their ANN model with a nonlinear one shows the superiority of the ANN model over the nonlinear model. Their study, which compared the ANN and nonlinear models in terms of various error expressions, showed that the ANN model had better results against the nonlinear regression models. [7] implemented an ANN application which inputs the driver characteristics (age, gender, alcohol rate, safety belt usage), vehicle types and several other factors to predict the severity of the possible accident (light, medium or heavy with fatality).

2. METHODOLOGY

2.1. Logistic Regression Analysis

Logistic regression analysis (LRA) is a regression technique used for predicting a categorical target variable that has exactly two categories. Beside the multiple linear regression, it fits a s-shape curve to data as the predicted value is always between 0 an 1 (See Figure 1).
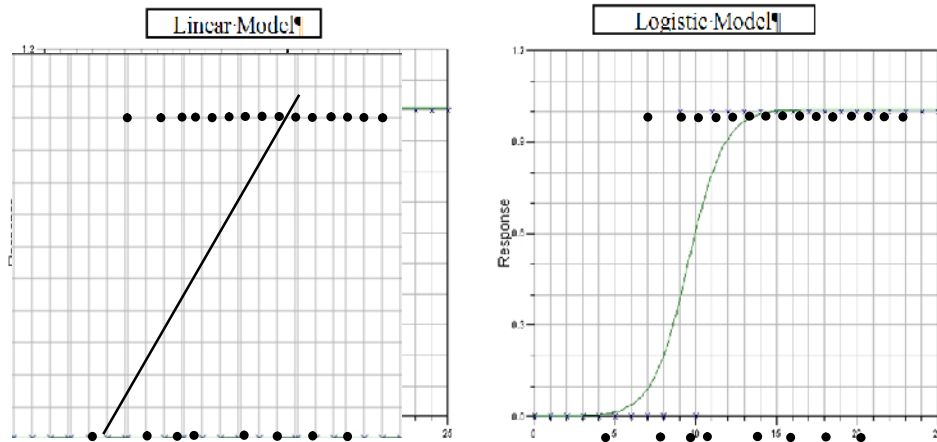
Figure 1. Data fitting of the linear model and the logistic regression model.

In the simplest case of the logistic regression, the decision of lane changing of a driver (changed his/her lane or continued through the deformation) can be described as one dichotomous outcome variable correlated with several continuous or discrete predictor variables (approach speed, gender of the driver, vehicle type etc.). As the plot of such data indicates two parallel lines of categorical outcome variables, fitting of a linear regression model is not applicable since these lines are difficult to be described with an ordinary least squares regression equation due to the dichotomy. However, logistic regression results in a s-shaped model. This model is curved at the ends and referred as "sigmoidal" or S-shaped model. Logistic regression method applies logit transformation to the dependent variable as the model predicts the logit of Y from X which is the natural logarithm of odds of Y. Specifically, logarithm of the odds of Y corresponds to the ratios of probabilities happening or not happening of Y. A general logistic model with multiple dependent variables can be described as following mathematical description.

$$\text{logit}(Y) = \ln\left(\frac{\tau}{1-\tau}\right) = a + b_1 X_1 + b_2 X_{2+\dots} \tag{1}$$

Where b_i is the regression coefficient of the independent variables either categorical or continuous and a is the axis intercept of the model. Typically a and bi can be estimated by the maximum likelihood method, which aims to maximize the likelihood of reproducing the data. Traffic data can be entered into the model as binary values (0 or 1) and dummy variables (e.g., 0 or 1) for categorical the predictors.

2.2. Artificial Neural Networks

The ANN is basically a mathematical representation of the human brain consist of numerous neurons connected to each other. Similar to the human brain, an ANN network is able to recognize the patterns in a data and suit well to the nonlinear characteristics of a system. Application of the ANN is well common for solution of many of engineering problems by modeling of nonlinear systems. In addition, the implementation of the ANN's in civil engineering problems of the field of transportation is quite common. An ANN consists of several data processing nodes called neurons. Basically the neurons or nodes are grouped in several layers called input, output and one or several hidden layers. A three-layer artificial neural network is shown in Figure 2.
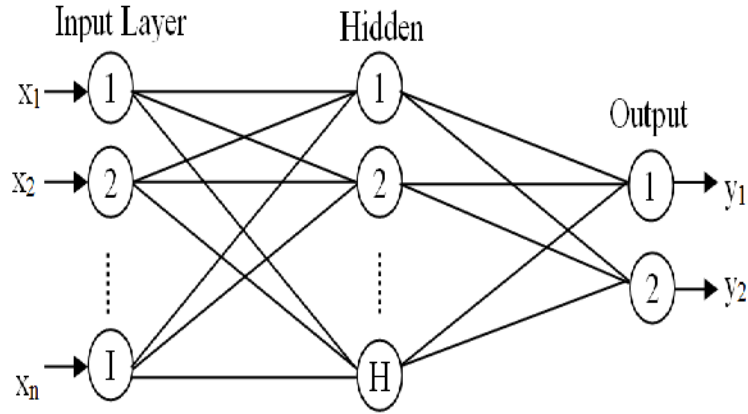
Figure 2. A typical ANN structure with three layers (Input, hidden and output).

In a typical feed-forward ANN structure, signal is introduced to the input layer neurons and transmitted to the next layers through the connections between the neurons until the signal reaches the output layer. Considering a typical 3-layer feed forward network each neuron in a layer is connected to the next layer neurons and there exist no interconnectivity among the neurons of same layer. The intensity of the passing signal is modified with the multiplication of the connection weights while transferring through the network and a network output is obtained from the neurons of output layer. Basically, neurons can be considered as the basic processing units of a neural network. In Figure 3, a typical neuron of an ANN is shown.
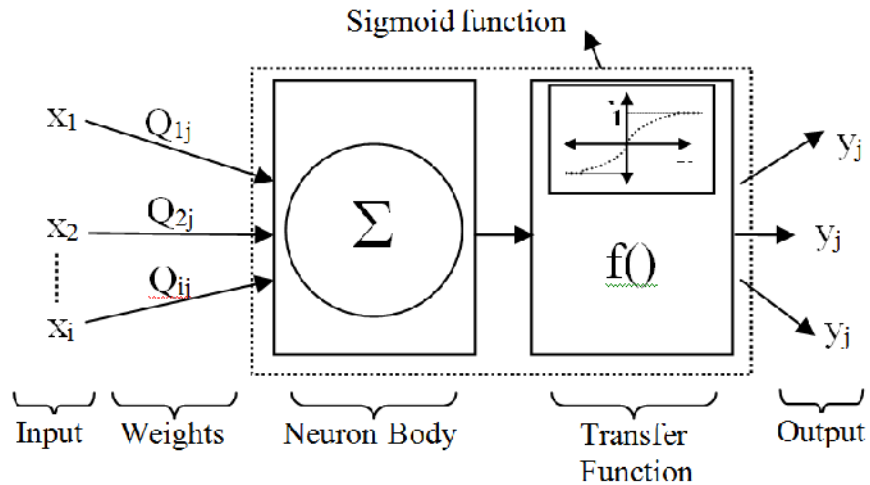
Figure 3. Detailed view of an ANN neuron.

Each neuron sums the weighted inputs from its connections and passes the sum through a specific transfer function to yield its signal output. Beside many transfer functions, the S-shaped sigmoid function is generally used. Hence, output signal is bounded between values of 0 and 1. Considering the transfer function, the output of a neuron in any layer is described as:

$$y_j = \left(\sum Q_{ij} X_i \right) = \frac{1}{1 + e^{-\left(\sum Q_{ij} X_i \right)}} \tag{2}$$

Beside many learning algorithms, training of an ANN is often described by altering the weights of the nodal connections with a training set of input and output data. Particularly, connection weights are initialized randomly or based on a previous experience. The weights are altered while introduction of data to minimize the difference between the network output and desired output. When the network error minimized beyond a specific limit, weights are stabilized and the network is considered to be trained for further predictions. With the back-propagation algorithm based on the generalized delta rule [8], the calculated network error is propagated back through the network and weights are adjusted. After the training process, network performance can be evaluated with a validation data mostly in the manner of mean square error (MSE) determined with the equation given below:

$$MSE = \frac{(O_e \ O_m)^2}{n-1}$$ (3)

Where O_e is the expected value, O_m is the predicted value from the model and n is the size of the data-set. To achieve optimal learning and avoid overtraining, excessive care should be taken while modeling. Determination of the numbers of the hidden layer neurons is also considered to be an important factor influencing the ANN performance as with high neuron numbers, training will take considerable time and over fitting of the data is quite possible [9].

3. OBSERVATIONS AND DATA ANALYZING

In the scope of the study, the traffic data were obtained for a two-lane road between Erzurum and Bingöl provinces of Turkey. The geometric characteristics of the deformation on the road surface are given in Table 1. The reason of selecting this location was the high intensity of the traffic flow including various types of vehicles. All observations were performed between the 7:00 a.m. and 8:00 a.m. in the morning, 18:00 and 19:00 pm in the evening with the absent of precipitation. A typical camcorder was used for data recording located at an elevated point with respect to road. The recordings were used for evaluating the traffic data with direct observation of cam recordings are shown below:

• Vehicle types,
• Drivers Gender,
• Vehicles speeds (km / h),
• Car behaviors as lane-changed or stayed on lane.

Geometric properties of the road are given in Table 1 and deformation points on road surface of the road are shown in Figure 4.

Table 1. Geometrical properties of the road and deformation point.

Width (W_d) (m)	Length (L_d) (m)	Depth/Height (D_d/H_d) (cm)	Road Width (B_r) (m)	Def. Position on Road surface (P_d) (m)	Deformation Type (S_d)
2	1.5	6.7	9.6	(2-4)	Stripping, Raveling and Deflection

Figure 4. Deformation points on Erzurum-Bingöl road surface.

The lane changing record data consisted of lane changing behavior of 825 vehicle drivers. For the ANN application the data were divided into two sets as training (n=700) and validation (n=125). Each data point belonging these sets consists of specific driver characteristics such as approaching velocity, type of the vehicle, gender of the driver and the lane changing decision of the vehicle while approaching to a surface deformation. The driver decision of lane changing can be described in Figure 5.
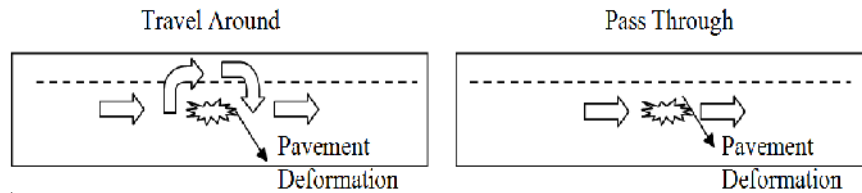

Figure 5. Driver behavior while approaching to a road deformation.

As the gender characteristics of the data are described are given in Table 2. 29 female drivers with a percentage of 3.2% among all drivers were observed. The overall percentage of female drivers among the car drivers was found as 4% since no female drivers recorded for commercial vehicles. Furthermore, the lower frequency of female drivers could be resulted due to road characteristics as it is located outside of the urban area.

Table 2. Number of drivers according to driver's gender and vehicle type.

Gender	Car	Bus	Lorry	Minibus	Truck
Male	655	31	29	30	53
Female	27	0	0	0	0
Total	682	31	29	30	53

The driving behaviors of male and female drivers were found to be different. In Table 3, lane changing behaviors with the approach characteristics are shown. Particularly, the numbers of vehicles belonging to each vehicle group which either changed their lane or pass through the deflection are summarized with mean approach velocities and corresponding standard deviations. It was seen that the mean approach velocities of each type of vehicle was higher for the drivers who remained on their lanes with respect to drivers who changed their lanes.

Table 3. Number of drivers according to lane changing decisions of each vehicle type

Vehicle Type	Lane Changed			Pass Through		
	Number	Mean	SD	Number	Mean	SD
Car	432	82.67	8.26	250	87.57	12.21
Bus	4	83.10	4.69	27	78.16	10.14
Tır	1	64.68	NA	28	70.29	9.41
Minibus	23	82.73	7.71	7	82.09	10.57
Truck	4	79.49	12.66	49	73.16	8.31

The ANN and LR models were constructed to predict the lane changing behaviors of a specific vehicle approaching a pavement deformation. Since, there existed considerably small number of female drivers and female drives are only recorded for cars, only the male drivers were considered for both the ANN and LR for the data homogeneity. For both models vehicle type and approach velocities were used as dependent variables and lane changing decisions were considered as independent or target variable which was going to be predicted.

In this study, a single hidden layer, feed forward neural network model was used with the back-propagation learning algorithm as it is sufficient for approximating complex nonlinear functions [10]. Determination of the optimum hidden layer size for the best performing model is considered as a difficult task. Despite the fact that, there is no reliable theory for selection, a trial-error methodology was implemented to select the number of neurons in the hidden layer. Before the application of ANN, the approaching velocity data as normalized to fall in range of 0.1 , and 1.0. In the study 9 ANN models were investigated with different hidden neuron numbers. A calibration data set (n=698) were introduced to the network consist of vehicle type and approaching velocity data. After calibration, each ANN model performance was evaluated with the validation data set (n=100). The neural network module of the commercial MATLAB[©] program was used in this study.

The variations of the MSE's for each ANN model are shown in Figure 6. It was seen that with the increasing number of hidden neurons, the MSE increases and the prediction performance of the model decreases down to a certain point. Particularly, ANN with 5 or 6 hidden nodes were found to be performing well and the differences of the performance with other ANN models are evident.
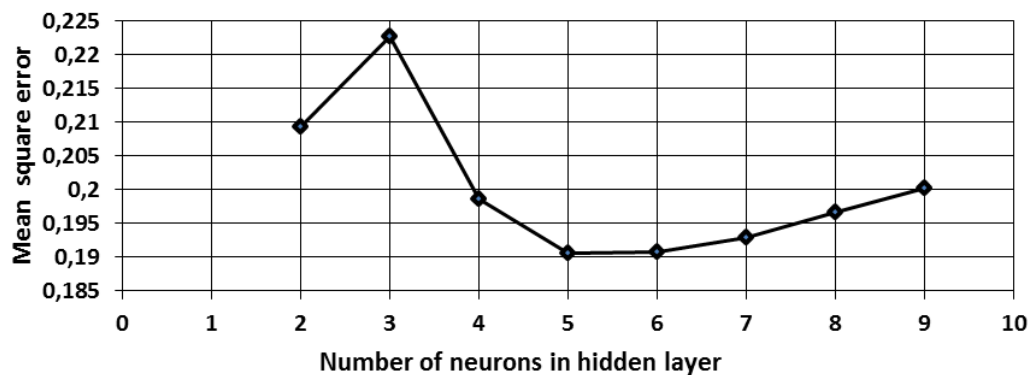

Figure 6. Mean square error's of models with different neuron numbers at hidden layer.

As similar to the ANN model, vehicle type data and approach velocities were introduced to the LR model to predict the target variable (lane changing behavior of the driver). Despite the forward entrance of the model variables, all considered variables were included simultaneously as there were relatively small numbers of variables. Hence, a single LR model was evaluated. As the vehicle type was a categorical variable with 5 cases (Car, Bus, Lorry, Minibus and Truck), 4 logical (1 or 0) dummy variables were proposed for the model. Variables considered in the LR model were shown in Table 4 and five-predictor logistic models were fitted to the data. The logistic regression analysis was carried out by the Logistic procedure in SPSS$^{©}$ version 20.

Table 4. Variables introduced to the LR model.

Name	Dependency	Variable Type	Conditions
Car	Independent	Dummy/Logical	1=Yes, 0=No
Bus		Dummy/Logical	1=Yes, 0=No
Tır		Dummy/Logical	1=Yes, 0=No
Minibus		Dummy/Logical	1=Yes, 0=No
Truck		Discarded	"Yes" if others are 0
Approach Velocity		Scale	NA
Lane Change Decision	Dependent/Target	Logical	1=Yes, 0=No

The prediction equation of the models was determined as below;

Log (lane changing decision)=3.432(Car)+0.801(Bus)-1.018(Lorry)+3.876(Minibus)-3.303(Approach Velocity)-1.227

According to the LR model, the log of the odds of changing the lane for a vehicle driver was negatively related to the approach velocity. Hence, the higher the approach velocity, the less likely it is that the driver would change his or her lane due to the surface deformation. As the vehicle types were considered, it was seen that the car and minibus drivers are more likely change their lanes with respect to the bus and lorry drivers. While truck drivers were represented by the null state of all dummy variables, negative constant of the model show that, truck drivers mostly passed through the deformation.

The performances of the ANN and LR models were evaluated with considering the null model which is a model without predictors and acts as a base line for performance evaluation. According to the null model all cases were predicted considering the largest outcome category [11]. In Table 5, results of models considering the validation data are shown. Particularly, the ANN model consist of 5 hidden neurons slightly performs better than LR model with successful prediction performance of 74%. Both the ANN and LR models perform well with respect to Null model without and predictors.

Table 5. Model results considering the prediction performances.

Observed		Predicted		Overall Percentage
		Yes	No	
ANN	Yes	47	3	74%
	No	25	25	
LR	Yes	48	2	70%
	No	28	22	
Null	Yes	50	0	50%
	No	50	0	

4. CONCLUSIONS

In this study, the potential of an ANN model for simulating the drivers behavior was evaluated with successive comparison of the model performances with the LR and Null models. While there existed a slight performance increment for the ANN model with respect to the LR model;

i). It is quite evident that, the ANN models can play an important role for providing the driver behavior approaching to a road surface deformation.

ii). With the investigation of coefficients of the LR model, it could be said that, approaching velocity plays an important role on the lane changing behavior of a driver. This can be criticized by the fact that, drivers with high approaching velocities preferred passing through the deformation to avoid the accidents by changing their lanes with a high speed. Additionally, it can be though that, the as the deformation can be noticed very late by the drivers, they are not able to react rapidly to change their lanes.

iii). According to the LR model, the log of the odds of changing the lane for a vehicle driver was negatively related to the approach velocity. Hence, the higher the approach velocity, the less likely it is that the driver would change his or her lane due to the surface deformation.

iv). According to the LR model, the log of the odds of changing the lane for a vehicle driver was negatively related to the approach velocity. Hence, the higher the approach velocity, the less likely it is that the driver would change his or her lane due to the surface deformation,

v). Particularly, the ANN model consist of 5 hidden neurons slightly performs better than LR model with successful prediction performance of 74%. Both the ANN and LR models perform well with respect to Null model without and predictors.

In the scope of this study, only recordings for dry weather condition were considered to have homogeneity in the data. Further studies can be performed considering the wet or snowy weather conditions to investigate those effects on the behaviors of drivers.

While the gender of a driver can influence the driving characteristics and lane changing behavior, the, gender of a driver was not considered in this study as there were not enough number of female drivers for commercial vehicles. While, observations showed the female drivers tend to drive slower, further studies required to investigate the driving characteristics by considering the gender.

It should be noted that, these results depend on only a limited data obtained from Erzurum-Bingöl road in Turkey. For further research, the numbers of observations and intersections can be increased.

REFERENCES

[1] E. Ozgan, "Karayolu araç Tipi ve Kaza ekli le Kaza Sonuçları Arasındaki li kilerin Analizi", Gazi Ünv., Müh. Mim. Fak. Dergisi, cilt 23, no:3,syf. 97-104, 2008.

[2] K. F Türe, "Trafik Kazaları ve Trabzon Bölünmü Sahil Yolu Örne inde Kaza Tahmin Modelinin Olu turulması", Y. Lisans Tezi, Fen Bilimleri Enstitüsü, Karadeniz Teknik Üniversitesi, Trabzon, 2008.

[3] A. Darçin and, B. Akba "A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics," Scientific Research and Essays vol. 5, no.19, pp. 2837-2847, 4 October, 2010

[4] D. Delen, R. Sharda and M. Bessonov, "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks," Accident Anal. Prev., vol.38, no.3,: pp.434-444, 2006.

[5] R.R. Hashemi, L.A. Le Blanc, C.T Rucks and A. Shearry, "A neural network for transportation safety modeling," Expert Syst. Appl., vol.9 no.3, pp.247-256. 1995.

[6] A.P. Akgüngör and E. Do an, "Estimating road accidents of Turkey based on regression analysis and artificial neural network approach," Adv. in Transportation Stud. An Int. J. Section, A pp.16: 11-22., 2008.

[7] H. Altun, A. Bilgil, and B.C. Fidan, "Treatment of multi-dimensional data to enhance neural network estimators in regression problems," Expert Systems with Applications vol.32, pp.599–605., 2005

[8] D.E Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation. In Parallel Distributed Processing", D.E. Rumelhart, J.l. McClelland, and the PDP Research Group, Et ds., Vols. I and II, Bratford Books and MIT Press, Cambridge, MA., 1986.

[9] N. Karunanithi, W.J. Grenney, D. Whitley and K. Bovee, "Neural Networks for River Flow Prediction," J. Comp. Civil Eng., ASCE, vol.8, pp.201-220, 1994.

[10] G. Cybenko, "Approximation by superposition of a sigmoidal function," Math. Control Signals Syst., vol.2, pp.303-314, 1989.

[11] C. Yıng, J. Peng, K. L. Lee and G.M. Ingersoll. "An Introduction to Logistic Regression Analysis and Reporting," The Journal of Educational Research, vol.96, no. 1. 2002

Authors

Metin Mutlu Aydın

Metin Mutlu Aydın is currently working as research assistant at Gumushane University Civil Engineering Department Transportation Division. Prior to joining Gumushane University, Mr. Aydın worked as research assistant at Dokuz Eylul University. He is continuing his PhD study in Civil Engineering Department of Akdeniz University, Graduate School of Natural and Applied Sciences, Antalya, Turkey. Also he holds a master's degree in Transportation Engineering from Dokuz Eylul University and a bachelor's degree in Civil Engineering Department from Dokuz Eylul Un versity.

Mehmet Sinan Yıldırım

Mehmet Sinan Yıldırım is currently working as research assistant at Gediz University Civil Engineering Department Hydraulic Division. Prior to joining Gediz University, Mr. Yıldırım worked as research assistant at Gumushane University. He is continuing his PhD study in Civil Engineering Department of Katip Celebi University, Graduate School of Natural and Applied Sciences, Izmir, Turkey. Also he holds a master's degree in hydraulic engineering from Middle East Technical University and a bachelor's degree in Civil Engineering department from Middle Technical University.

His current research interests include lane utilization, lane changing, dwell time, bus blockage effect, signalized intersections, intelligent transportation vehicles and modal shift in transit system.His current research interests include hydraulic modeling, flow, Artificial neural network and lane flow.

Orhan Karpuz

Orhan Karpuz is currently working as assistant professor doctor at Gumushane University Civil Engineering Department Transportation Division. Prior to joining Gumushane University, Mr. Karpuz worked as research assistant at Karadeniz Technical University. Also he has finished Msc. and Phd. in Transportation Engineering from Karadeniz Technical University and a bachelor's degree in Civil Engineering Department from Karadeniz Technical University.

Kiarash Ghasemlou

Kiarash Ghasemlou is currently working as transportation planner at EMBARQ Turkey. He helps to advance sustainable urban transport research, project planning and implementation. Mr. Ghasemlou supports EMBARQ Turkey to planning, development and evaluation the performance of sustainable transportation projects in Turkish cities. Prior to joining EMBARQ Turkey, Ms. Ghasemlou worked as research assistant at University of Dokuz Eylul on "Investigation of Effect of Driver Behavior on the Capacity and Performance of Urban Intersection and Arterials" supported by TUB TAK Also he has done research works about impact of buses/heavy vehicles on signalized intersections and macroscopic modeling of traffic stream particularly by using Cell Transmission model.His current research interests include lane utilization, signalized intersections, concrete roads, porous concrete, rigid roads. He is continuing his PhD study in Istanbul Technical University in field of urban and regional planning. Also he holds a master's degree in transportation engineering from Dokuz Eylul University of Turkey and a bachelor's degree in civil engineering from Iran.