Sigma J Eng & Nat Sci 35 (1), 2017, 157-179

Sigma Journal of Engineering and Natural Sciences
Sigma Mühendislik ve Fen Bilimleri Dergisi

Review Paper / Derleme Makalesi

A COMPREHENSIVE REVIEW FOR ARTIFICAL NEURAL NETWORK APPLICATION TO PUBLIC TRANSPORTATION

Engin PEKEL*, Selin SONER KARA

Yıldız Technical University, Industrial Engineering Department, Beşiktaş-İSTANBUL

Received/Geliş: 13.06.2016 Revised/Düzeltme: 24.10.2016 Accepted/Kabul: 16.01.2017

ABSTRACT

This paper presents a comprehensive review of research studies related to the application of artificial neural networks (ANNs) to public transportation (PT) since 2000. PT applications with ANNs have a great prominence because it provides an opportunity of prediction, comparison and evaluation in PT. A short introduction for applied studies in public transportation based on NN is included to guide the unfamiliar readers and a detailed review table has been presented in the paper. More than a thousand studies have been viewed, however, 72 studies of PT are related to ANN. It is observed that multi-layer feed forward network with gradient descent training has been commonly used by now. In contrast, the other less known methods are prone to increase. This paper guides future research directions and presents the methods to be exerted in PT for input determination.
Keywords: Artificial neural network, multi-layer perceptron, public transportation, radial basis function..

TOPLU TAŞIMA ARAÇLARINDA UYGULANAN YAPAY SİNİR AĞI İÇİN KAPSAMLI BİR LİTERATÜR TARAMASI

ÖZ

Bu çalışma, 2000 yılından itibaren toplu taşıma (TT) alanında yapay sinir ağları (YSA) ile yapılan çalışmaların kapsamlı bir taramasını sunmaktadır. Yapay sinir ağı ile yapılan TT uygulamaları büyük bir öneme sahiptir çünkü YSA, araştırmacılara etkin bir şekilde tahmin, karşılaştırma ve değerlendirme yapabilme imkanı vermektedir. YSA ile uygulama yapmamış olan okuyucular için, YSA ile yapılan TT araçlarındaki çalışmaların kısa bir açıklaması çalışmaya dahil edilmiş ve detaylı bir tarama tablosu çalışma kapsamında sunulmuştur. Binden fazla çalışma incelenmiş, ancak YSA ve TT ile ilgili sadece 72 çalışmanın mevcut olduğu görülmüştür. İniş eğimli eğitimi ile çok katmanlı ileri beslemeli ağ yapısının daha çok kullanıldığı görülmüştür. Bu durumun aksine, diğer az bilinen yöntemler artma eğilimindedir. Bu çalışma, gelecek araştırma yönleri için rehberlik etmekte ve TT' de girdinin belirlenmesi için uygulanacak yöntemleri sunmaktadır.
Anahtar Sözcükler: Çok katmanlı algılayıcı, sinir ağı, toplu taşıma, radyal bazlı fonksiyon.

[*] Corresponding Author/Sorumlu Yazar: e-mail/e-ileti: pekelc@hotmail.com, tel: (212) 383 28 87

1. INTRODUCTION

PT within the transportation is of a vital prominence for the expanding metropolitan regions. Because the overall aspect in PT tenders influences on mobility that complies with a more common interpretation of sustainability that covers economic, social and environmental factors [1]. PT contains a number of modes such as buses, subways, bus rapid transit (BRT), light rails, tramways and ferries. These PT modes deal with the fuel consumption, traffic congestion and carbon emission within a city.

The area of PT studies has not highly drawn attention in ANNs since 2000. This situation can be realized from the research studies in this field which are less than 72 papers. ANN applications in the field of PT are interested in this paper since the ANN for PT is regarded as prominent by the researchers. Mainly, the importance of ANN for PT is seen from the papers that are applied for forecasting, evaluation, comparison and minimization problems. Papers, which have been applied in diverse fields, yield to present a literature paper. Moreover, there has not been any published review in the field of PT which is integrated with ANN. The objective of this paper is to provide a detailed information such as hidden layer number, hidden neuron number, the kind of ANN that is used and the kind of training method that is applied by presenting the studies in an integrated area of both PT and ANN. In addition to presentation of the detailed information, it aims to provide the further progress to the researchers. We reviewed the literature on ANNs approaches that applied to PT. The databases covered are ScienceDirect, Wiley and IEEE Xplore. A total of 72 papers are reviewed ranged from 2000 to 2016.

The main contributions of the paper are as follows: it is aimed to increase the interest in PT based on ANN, since any improvement in PT provides economic opportunities and reduces both gasoline consumption and carbon emissions. Moreover, positive interests to PT affect public health through a variety of mechanism [2].

After this chapter, introduction to ANN is presented by focusing especially on both feed-forward and recurrent neural network architecture in chapter 2. Chapter 3 brings to light the all-ANN application in reviewed PT studies. Chapter 4 clarifies the concepts reviewed in this paper via presenting brief explanations. Lastly in chapter 5, conclusions of review and future directions are presented.

2. ANN METHODS IN PT

Researchers from many scientific fields are devising ANNs to bring solution to the difficult problems such as optimization, prediction and pattern recognition [3]. The ANN is actually a novel computer architecture and a novel architecture relative to traditional computers by imitating biological neural networks. It allows using very simple computational operations such as fundamental logic elements to deal with complex, mathematically vague, nonlinear or stochastic problems [3]. ANNs can be considered as weighted directed graphs where directed edges are connections between input and output. Besides, artificial neurons are nodes [3]. There are many types of ANN model such as dynamic, static and memory ANN. However, a number of ANN models, which are mainly performed by the researchers, are presented in this paper. Network architecture can be divided into two groups; feed-forward network and feed-back networks.

Feed-forward neural networks (FFNNs) are the most general and the most broadly used models in many practical applications. The FFNN is the first and simplest sort of ANN that is designed. The information flows in only one course from the input nodes to the output nodes without creating loops. In this paper, single layer perceptron, multi-layer perceptron and radial basis function (RBF) networks are considered in FFNNs, because they are mainly performed by the researchers. Feed-back neural networks (FBNNs) is the network that contains at least one feed-back connection. This feedback feature is in discrete cycles of weight computation. Hopfield network and Boltzmann machine are included in FBNNs [3] in this paper, because they are mainly performed by the researchers. Figure 1 illustrates both feed-forward and recurrent ANN.

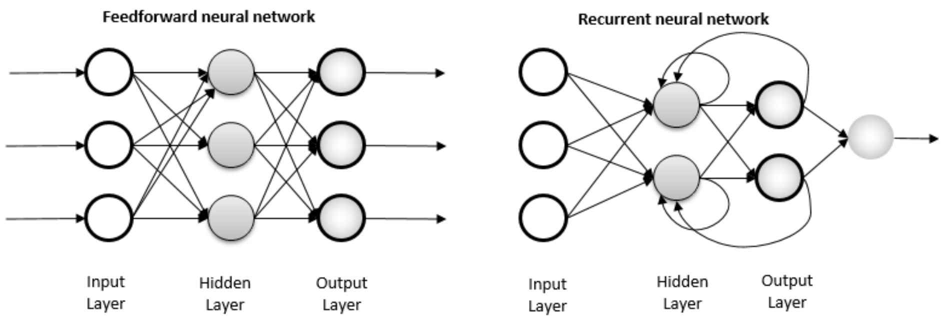

Figure 1. Feed forward and recurrent ANN architecture.

The activation function indicates the extent of a neuron's activation. The reaction of the neurons is shaped with respect to the activation function. Commonly performed activation functions are heaviside, hyperbolic tangent and logistic function. Another important part of neuron reaction is a threshold value because the activation function of a neuron behaves especially near the threshold value.

A key element for reaching a successful learning is to use a bias neuron. The output value of a bias neuron is always equal to 1. Bias neurons in an ANN that allows researchers to change the value of the activation function.

3. DETAILED ANALYSES OF THE LITERATURE

Detailed review of ANN applications in PT is presented in Appendix A. However, problem classifications, analysis of modeling approach, solution methodologies and descriptive analysis are explained in this chapter.

3.1. PT applications

PT within the transportation is of a vital prominence for the expanding metropolitan regions. Because the overall aspect in PT tenders influences on mobility that complies with a more common interpretation of sustainability that covers economic, social and environmental factors [1]. PT contains a number of modes such as buses, subways, bus rapid transit, light rails, tramways and ferries. In PT, the planning and scheduling of trips have a key role. Thus, it is seen that most of the researchers deal with bus and BRT trips and passenger flow. It can be stated that the highway public transportation modes, such as conventional bus and BRT systems are more common than rail public transportation modes, such as light rails and tramways due to higher flexibility and lower capital cost of the former.

3.2. Problem classifications

In the reviewed papers, the researchers have focused on four main problem types. At 69% of the reviewed papers, the researchers are more interested in forecasting and prediction. Evaluation and analysis, minimization and maximization, and comparison have proportions of: 25%, 4%, and 2%, respectively.

To give detailed information, sub-problem types are presented for four main problem types.

- **Forecast and prediction.** Forecast and prediction have been applied in various areas as follows: bus arrival time [5,6,15,16,17,20,32,40,45,50,80,89], traffic flow [7,25,27], bus average travel time and variance of travel times [10,13,39], route choice [11,12,48,86], decision making units [21,78], trip flow [22,34,35,43,44,74,83], passenger flow [28,29,33,37,38,46,51,84,85,90] and passenger waiting time [41,42].

Papers [66,67,68], in the area of both bus arrival time and bus travel times, have been published without ANN. The corresponding papers may have utilized ANN technique to benefit from the efficiency of ANN.

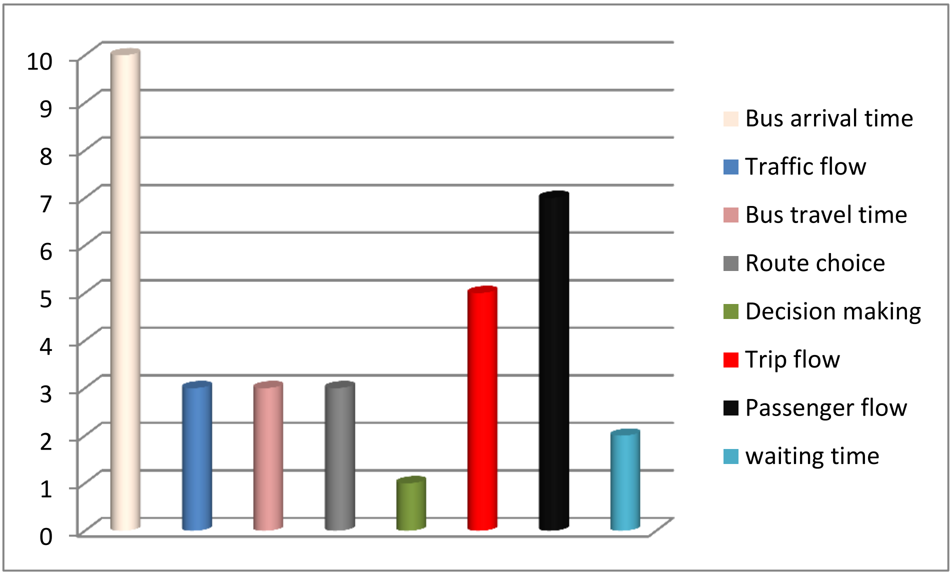

Figure 2. Distinct problem types for forecast and prediction

Figure 2 shows the different problem classifications in area of forecast and prediction. Bus arrival time has been the mostly interested problem type in this area. Following bus arrival time, researchers dealt with passenger flow and trip flow.

- **Comparison.** In this area, [35] is presented to compare the RBF with FFNN in the area of PT trip flow.
- **Evaluation and analysis.** This problem type has been applied in various areas as follows: measurement of the pollution impact of BRT [8], road safety [14], bus classification [18], air quality of PT [19,81,92], decision making units [21,48,49], rail capacity [23], performance of PT services perceived by the passengers [24,36,71,72,75,77], forecaster variables [26], and green vehicle distribution model [31].
- **Minimization and maximization.** In this area, minimization of the traveling time variance [9] and vehicle delay [47] are presented.

3.3. Analysis of modeling approach

The construction of ANN is a crucial element to achieve a successful result in a given problem type such as minimization, maximization and prediction. In literature, most of the papers utilized multi-layer feed forward (MLFF) and it is used in 44 papers. Other ANN constructions; RBF, support vector machine (SVM), and ANFIS are used in 3, 1, and 2 papers respectively.

All required descriptions of the parameters are presented in the Table 1.

Table 1. The detailed parameter descriptions used for methodologies

Parameters	Descriptions
d^i	Desired value for factor i
E	Error for delta rule
f_a	Approximation function in learning process
G	Gaussian function
H	Output in the hidden layer
L	Norm
m	Distance
M_k	Measurement matrix
o^i	Output value for factor i
p_k	Process noise (random process)
S	State
F	Sign factor
S_k	Approximation function in learning process
v_k	Measurement noise
w_j	Weights for input variable j
x_j^i	Input variable j for factor i
y	Response variable
α	Learning rate ($0<\propto<1$)
ω	Hidden layer weight
η	Sign factor

➢ **Multi-layer perceptron (MLP).** MLP includes multiple layers and each layer is connected to the next layer and MLP transforms the inputs to a set of outputs. MLP can solve non-linear and stochastic problems by back-propagation algorithms. The performance of the algorithm can be generally carried out by the SSE between the output and the desired values. As per usual, SSE value can be minimized by using error descent method. The learning stage is cut when the difference between O and d is zero. MLFF are applied in the papers [5-22, 25-29, 31, 33-39, 41- 51,82,88] and distinct areas such as prediction and forecast, comparison, evaluation, and minimization. The learning algorithm of MLP is shown in Eq. (1). MLP architecture is clearly shown in Figure 4.

$$\frac{\partial E}{\partial w_j} = \propto x_j^i.\left(o^i - d^i\right) \tag{1}$$

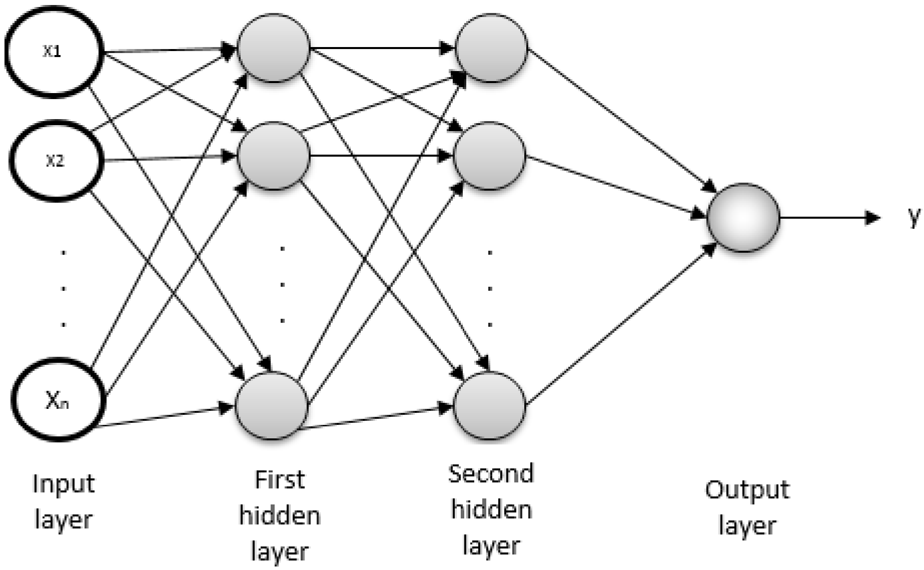

Figure 4. A visual architecture of multi-layer perceptron.

Back-propagation (BP) is a widely used method of training algorithm. The BP training method can be exerted in the networks where desired outputs are available, even the outputs of the intermediate layers are absent. This is usually regarded as a supervised learning method. The BP algorithm aims to minimize SSE by calculating the steep descent (gradient). Gradient descent learning rule is a delta rule, which is a special type of the back propagation algorithm. This rule embarks on the SSE in the learning phase. This rule has a great applicability that can be performed in both binary and continuous neurons [4].The delta rule is derived by attempting to minimize the error in the output of the ANN through gradient descent. The calculation of delta rule is illustrated in Eq. (2) and Eq. (3).

$$E = \sum_i E^i = \frac{1}{2}\sum_i \left(d^i - o^i\right)^2 \tag{2}$$

$$\frac{\partial E}{\partial w_j} = -\sum_i \left(d^i - f\left(\sum_j w_j x_j^i\right)\right).f'\left(\sum_j w_j x_j^i\right).x_j^i \tag{3}$$

To obtain the weight update rule, the error descent rule ($\Delta w_j = -\propto \frac{\partial E}{\partial w_j}$) is used and the change that is occurred in the weights can be calculated in Eq. (4) as follows:

$$\Delta w_j = \alpha \frac{\partial E}{\partial w_j} x_j^i \tag{4}$$

First, E, with respect to the weights, is found and then weights should be updated as shown in Eq. (5).

$$\Delta w_{ij} = -\propto \frac{\partial E}{\partial w_{ij}} \tag{5}$$

- **Radial basis function neural network (RBFNN).** RBFNN is a feed-forward network that is trained by a supervised training algorithm. Radial functions are a special type of function. It is a highly interesting that their response increases or decreases monotonically with distance from a central point [53]. RBFNNs have several superiorities with respect to back propagation networks. RBFNNs can be trained in two steps. In the first step, center vectors (c_j) are determined in a hidden layer. In the second step, Function estimation, which is a common objective function, is fit with coefficients to the outputs. The calculation of the output in RBFNN is shown in Eq. (6), Eq. (7) and Eq. (8).

$$O^j = w_{0j} + \sum_{k=1}^{N} w_{kj} h_k \tag{6}$$

$$h_k = G(d_k, \sigma_k) \tag{7}$$

$$O^j = w_{0j} + \sum_{k=1}^{N} w_{kj} G_k\left(\left|\left|x_j - c_j\right|\right|\right) \tag{8}$$

where O^j is the approximation function. Euclidian distance (h_k) is represented by $G_k\left(\left|\left|x_j - c_j\right|\right|\right)$. w_{kj} is the weights between layers.

RBF are performed in three papers: first paper [35] presented a comparison between MLFF and RBF. Second paper [19] used RBF to evaluate air quality of PT. Prediction are applied in the third paper [17]. Studies that have been exerted in RBF are less than MLFF.

➢ **Support vector machine (SVM).** SVM, in general, is used to analyze data for classification and regression analysis with regard to structural risk minimization (SRM). Instead of minimizing the absolute value of an error square, SVM applies SRM [59]. SRM is a function that minimizes some risk function $R(w)$, which is also called the expected loss, and it is calculated as in Eq. (9), Eq. (10) and Eq. (11).

$$R(w) = \int L(y, o) dP(x, y) = \int L(y, f_a(x, w)) dP(x, y) \tag{9}$$

$$L(y, o) = |y - f_a| \tag{10}$$

$$o = y = \begin{cases} f_a(x, w) = \sum_{i=1}^{N} w_i \varphi_i(x), if\ the\ network\ is\ RBF \\ f_a(x, w, \omega) = \sum_{i=1}^{N} w_i \varphi_i(x, \omega_i), if\ the\ network\ is\ MLP \end{cases} \tag{11}$$

The expected loss function $L(y, f_a(x, w))$ is computed on the training set $\mathrm{D}(x_i, y_i)$ and $\varphi_i(x, \omega_i)$ is a number of functions such as sigmoidal or tangent [59]. Only a single paper [15] is presented which applied SVM in order to predict bus arrival time.

- **Adaptive neuro fuzzy inference system (ANFIS).** ANFIS is an ANN which is based on the Takagi-Sugeno inference system. It is used to model high level nonlinear functions. ANFIS allows to establish a number of fuzzy if-then rules with suitable membership functions to form the stipulated input-output pairs [60]. ANFIS has five layers which contains distinct operations and these operations is summarized as below:

Layer 1: Determine the membership functions ($\mu_A, \mu_B, ...$) i.e. bell-shaped, by considering linguistic expressions

Layer 2: Perform the rules to produce the weights in the intermediate layer and the calculation of the weights is shown in Eq. (12).

$$w_i = \mu_{A_i}(x) * \mu_{B_i}(y) \tag{12}$$

Layer 3: Calculate the normalized firing strengths to scale between 0 and 1
Layer 4: The output of this layer is considered as consequent parameter
Layer 5: Calculate the overall output

Two papers, which used ANFIS, are presented as follows: first paper [31] dealt with green vehicle distribution model in the field of evaluation and analysis. Second paper [48] dealt with a signalization problem in the area of decision-making unit.

3.4. Solution methodologies

- **Resilient propagation (Rprop).** Resilient propagation applies a direct adaption of the weight step based on local gradient information. The Rprop consider only the sign of the partial derivative and moves on each weight with regard to η^{+} or η^{-} factor. Individual value (Δ_{ij}) for each weight is shown as in Eq. (13) [56].

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^{+} * \Delta_{ij}^{(t-1)}, if \ \frac{\partial E^{(t-1)}}{\partial w_{ij}} * \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \eta^{-} * \Delta_{ij}^{(t-1)}, if \ \frac{\partial E^{(t-1)}}{\partial w_{ij}} * \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ \Delta_{ij}^{(t-1)}, else \\ where \ 0 < \eta^{-} < 1 < \eta^{+} \end{cases} \tag{13}$$

When the individual update value for each weight is determined, the change in the weights is calculated as

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, if \ \frac{\partial E^{(t)}}{\partial w_{ij}^{(t)}} > 0 \\ +\Delta_{ij}^{(t)}, if \ \frac{\partial E^{(t)}}{\partial w_{ij}^{(t)}} < 0 \\ 0, else \end{cases} \tag{14}$$

$$\Delta w_{ij}^{(t+1)} = w_{ij}^{t} + \Delta w_{ij}^{(t)} \tag{15}$$

The update values and the weights are becoming different every time the entire pattern set is offered to the network [56]. Peters et al. exerted resilient propagation in order to develop a system that makes a timetable optimization and obstructs time delay.

- **Genetic-algorithm-based neural network.** Genetic algorithms (GAs) belong to the larger class of evolutionary algorithms that offer solutions to the problems which are difficult to deal with. GA is an iterative search procedure keeping on a population of structures that are possible solutions to specific domain challenges. GAs are search algorithms based on the natural selection process [57]. A typical cycle of GA connection weights is applied as following steps:

1. Generate a corresponding ANN with weights by determining each genotype in the current generation with a number of connection weights
2. Calculate the mean squared error (MSE), SSE or mean absolute percentage error (MAPE) between desired and output values and a correction process should be included in the fitness function to give a penalty to large weights
3. Choose parents for reproduction with regard to fitness function
4. Execute crossover or mutation to parents in order to produce the next generation [58].

Some authors [19,87] performed a hybrid GA-ANN in order to model air quality inside a PT or determine signal priority.

- **Supplemental methods**
  - **Kalman Filter.** Kalman filter provides a recursive solution to the linear optimal filtering problems. Kalman filter both bypasses the need for holding the past output data and provides more efficient estimation [54]. The Kalman filtering enables adjustment of estimates for a particular purpose. The researchers interested in Kalman filter may reach the detailed gist from the book of "Kalman Filtering and Neural Networks" [54]. The Kalman filter is formulated in Eq. (16) and Eq. (17).

$$s_{k+1} = F_{k+1,k} s_k + p_k \tag{16}$$

$$o_k = M_k s_k + v_k \tag{17}$$

Kalman filter, which estimates the future states of dependent variables, is applied by the researcher in order to predict bus arrival time at bus stop [15].

  - **Bootstrap method.** The Bootstrap is a type of a larger class of methods that resample from the existing data set with replacement. It is so difficult to determine the standard error or estimate a parameter without any parametric assumptions [55]. In that case, the bootstrap offers a way to determine the standard error. Bootstrap method is used in an ANN to evaluate the level of error. The readers interested in bootstrap method may reach the detailed gist from the book of "Bootstrap Methods: A Guide for Practitioners and Researchers" [55]. The researchers [18] proposed a different approach in order to classify bus lines. Moreover, different ANN structures are tested by bootstrap method.

- **K-nearest neighbors (k-NNs).** K-NNs algorithm, used for classification or regression, is a nonparametric classification method. It finds the k closest points to a query point with regard to the Euclidean distance and selects the majority. Its ease of usage enables applicants to implement it to large scale complex problems [61].The efficiency of the k-NN is increased by assigning a weight to each of k-NNs [62]. Another way to improve the efficiency of the k-NN is to exert the AURA k-NN. The Aura k-NN neural network technique is able to solve large problems for classification faster than traditional k-NN [63]. The researchers [15] applied k-NNs to predict bus arrival times at a bus stop for different routes.
- **Pruning algorithm.** Pruning algorithm is commonly used in order to determine the size of hidden neuron or layer in ANN. This algorithm set the weights to zero and evaluates the change in the error (SSE). If the change is higher than the previous error, the action to be applied is removed. This evaluation process should be repeated until reaching the least error with respect to the threshold [64]. The researchers interested in pruning algorithm can reach the detailed gist from the study of "Pruning Algorithms-A Survey" [64].
- **Growing algorithm.** Growing algorithm begin with a small network to determine the sufficient layers in ANN instead of pruning algorithm. It grows by allowing hidden units with regard to improvement in the error [65]. The researchers interested in pruning algorithm can reach the detailed gist from the study of "A Function Estimation Approach to Sequential Learning with Neural Networks" [65].

Pruning and growing algorithm applied in [10] to determine the optimal hidden neuron number in the area of bus arrival time prediction.

3.5. Descriptive analysis

The distribution of journals, which covers the published papers, indicates their desires in ANN applications with PT. The distribution of the journals with respect to publication years and the number of papers published are presented in Table 2.

Table 2. Distribution of literature based on the source of publication

Publication	Year of publication											
	2002-2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Total
Transportation Research Part C	-	-	-	-	-	-	2	1	1	-	-	-	4
Journal of Transportation Eng.	1	-	-	-	-	-	-	2	1	-	-	-	4
Journal of Intelligent Trans. Sys.	-	-	1	-	-	-	1	-	-	-	1	-	3
IEEE Transactions. On Int. Trans. Sys.	1	-	-	-	-	-	-	-	1	-	-	-	2
Mathematical and Computer Mod.	-	1	1	-	-	-	-	-	-	-	-	-	2
Journal of Public Transportation	1	-	-	-	-	-	-	-	-	1	-	-	2
Expert systems with Applications	-	-	-	-	-	-	-	-	-	2	-	-	2
Journal of Advanced Transportation	1	-	-	-	-	1	-	-	-	-	-	-	2
Neural Computing and applications	-	-	-	-	-	-	-	-	-	2	-	-	2
Applied Mechanics and Materials	-	-	-	-	-	-	-	-	2	-	-	-	2
Others (only 1 study)	6	2	1	2	3	1	2	2	7	8	7	6	47
Total	10	3	3	2	3	2	5	5	12	13	8	6	72

Table 2 reveals that the subjects of ANN applications in PT are considered by many journals. Moreover, the journals including more than two publications are illustrated in Table 2. This also clarifies the huge area of review in this study. Among the journals, two are clearly more active than the others in publishing ANN on PT: Transportation Research Part C (4 papers in various subjects) and Journal of Transportation Engineering (4 papers).

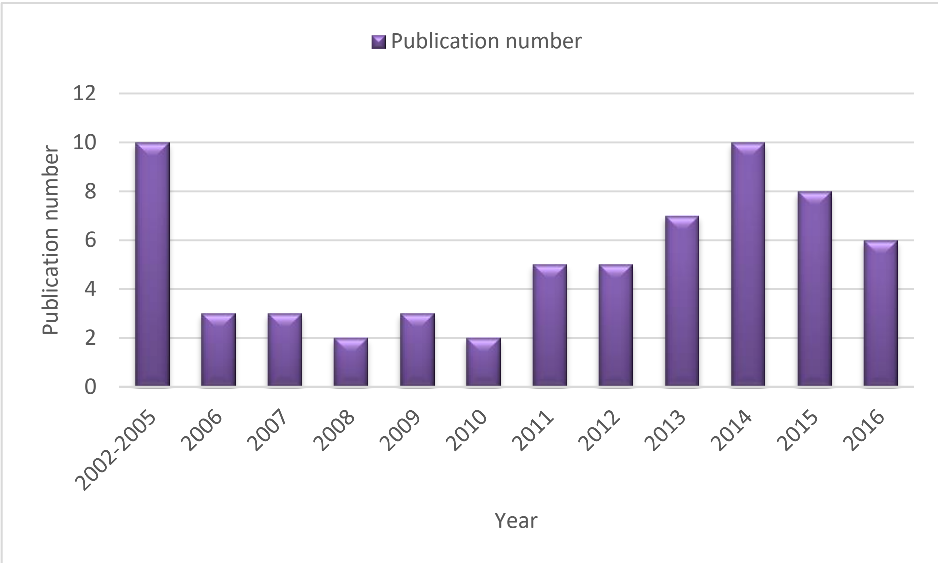

Figure 5. Number of neural network applications in PT.

ANN applications in PT tend to increase especially in the last five years. The raise of the interest is seen from the Figure 5. Figure 5 also presents the two moving average that is based on two sequential period.

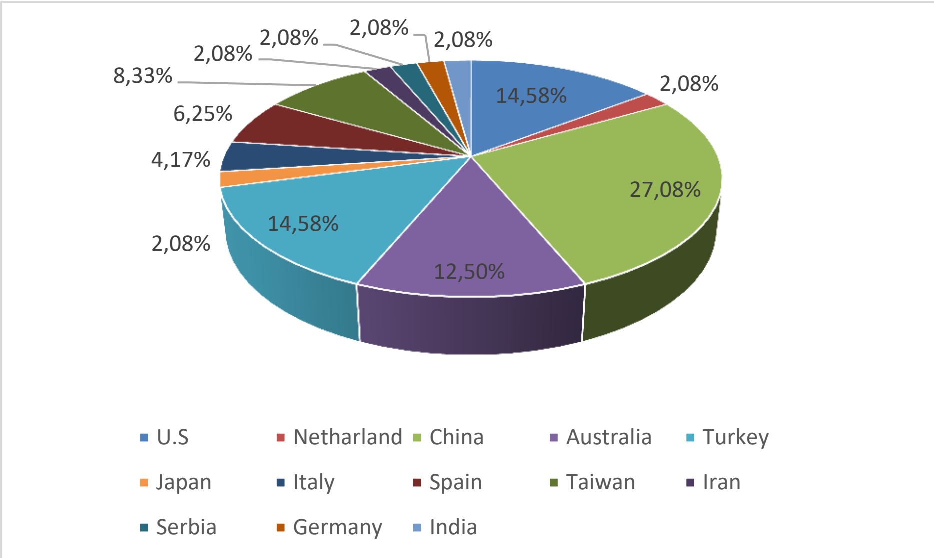

Figure 6. Research interest with respect to country.

Figure 6 presents that the highest interested country is China.

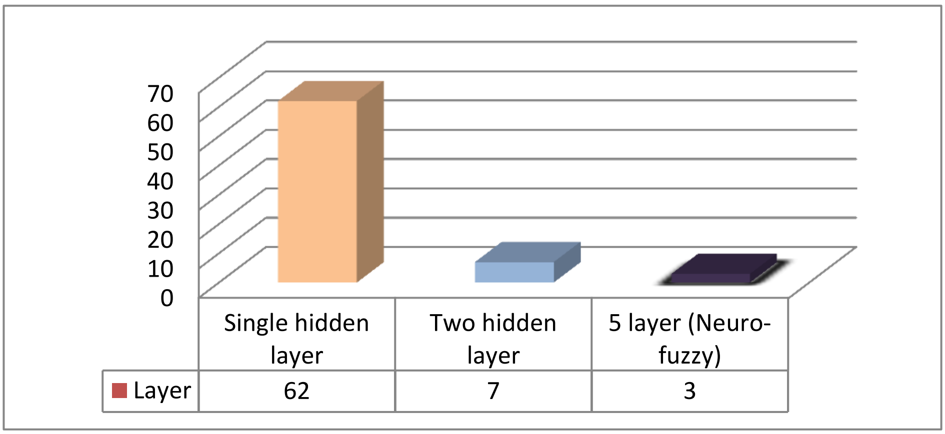

Figure 7. Distribution of hidden layer numbers

Distribution of hidden layer numbers, which is used in ANN, is shown in Figure 7. The most of the papers, (62), used a single hidden layer. Several papers [6, 19, 25, 37, 46, 50,76], (7), choose two hidden layers. Except above layers, three of the papers [31, 48,71] used five hidden layers which mean ANFIS.

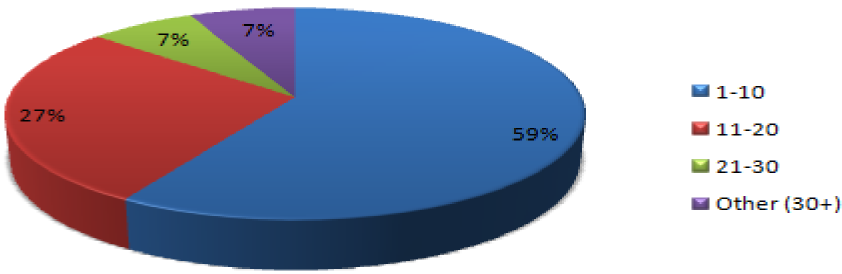

Figure 8. The number of hidden neuron

The number of hidden neurons is grouped into four main ranges that are shown in Figure 8. 1-10 neuron group are mostly chosen in 59% of the papers. Researchers choose 11-20 neuron group in 27% of the papers which is less than 1-10 neuron group.

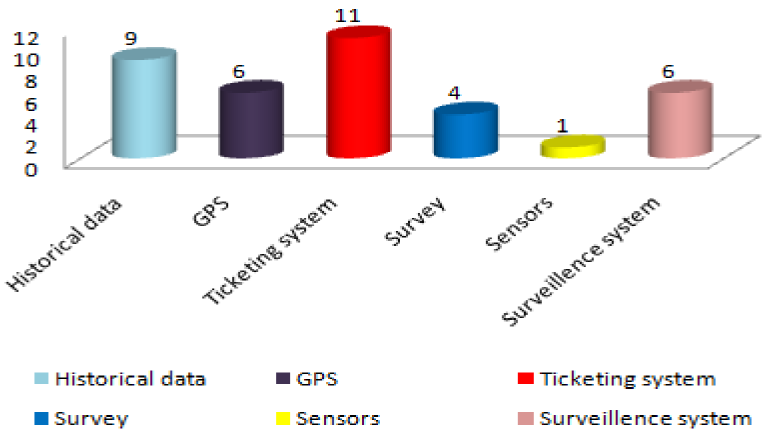

Figure 9. The methods of data collection

It is seen that six main data collection methods are present in Figure 9. The most chosen data collection methods are the ticketing system [9, 11, 18, 21, 22, 26, 34, 37, 42, 43, 50] and historical data [7, 15, 17, 19, 27, 28, 29, 46, 51], 11 and 9, respectively. Moreover, GPS [6, 13, 16, 25, 39, 45] and surveillance system [8, 10, 20, 28, 33, 47] chosen commonly to collect the data.

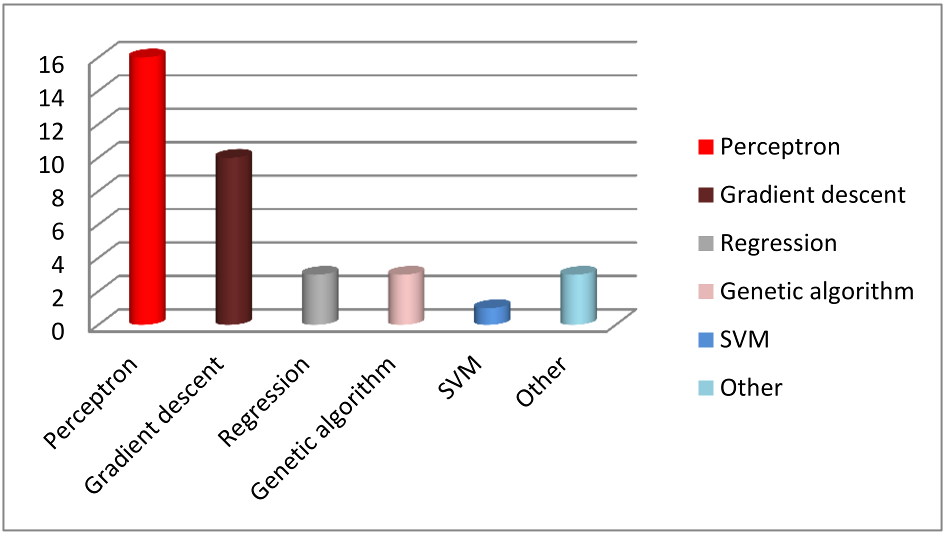

Figure 10. The number of different learning algorithms

In the light of Figure 10, the most used learning algorithms are perceptron [5, 8, 11, 13, 16, 18, 22, 25, 26, 28, 29, 33, 36,82,86,88] and gradient descent [6, 12, 17, 20, 35, 43, 44, 46, 50]. GA [19, 30,87], Regression [14, 37] and SVM [15] are used 3, 2 and 1 time, respectively.

3.6. Performance measurement

The evaluation of the output of the ANN has been conducted in terms of SSE, MSE and root mean square error (RMSE).

The RMSE is used to evaluate how well an ANN learns a given model. The RMSE aims to represent the standard deviation between desired values and output values. The calculation of RMSE is presented in Eq. (18) as follows.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O^i - d^i)^2}{n}} \tag{18}$$

MSE and SSE formulation has been illustrated in Eq. (19) and Eq. (20) as follows

$$MSE = \frac{\sum_{i=1}^{n}(O^i - d^i)^2}{n} \tag{19}$$

$$SSE = \sum_{i=1}^{n}(O^i - d^i)^2 \tag{20}$$

Sometimes, the output even reaches an optimal epoch, it may not meet the expected values or pattern. It is beneficial to add a parameter that is called momentum since any performance value may have decreased. The momentum parameter is widely used to avoid the system from coalescing to a local minimum point. When the learning rate tends to increase, instability into the learning rule bring out that wild oscillations is prone to unearth.

4. DISCUSSION

MLFF network, which can solve non-linear and stochastic problems with back-propagation algorithm, has been commonly practiced in the field of PT. Multi-layer feed forward covers the most of the ANN architectures that are approximately equal to 88%. However, the recurrent neural network, which is a network whose neurons send feed-back signals to each other, has been scarcely performed by the researchers [30]. RBF, ANFIS and SVM are chosen at the rate of 6%, 4%, and 2%, respectively. It is obvious that MLFF comparing to other neural networks has been more preferred in the field of PT.

% 69 of the applications have aimed the prediction or forecasting by the researchers. Brief explanations of forecast and prediction papers are presented in the section of problem classification. Bus arrival time, passenger flow and trip flow in the field of prediction and forecasting generally are preferred by the researchers. While the most of the researchers has dealt with the prediction and forecasting, the rest of the researchers have practiced the minimization and maximization problems [9, 47], evaluation and analysis [8, 14, 18, 19, 21, 23, 24, 26, 31, 36, 48, 49], and comparison [35] fields. The researchers, who conducted evaluation and analysis, dealt with common decision making units and performance of PT services. The researchers, who conducted minimization and maximization, studied in traveling time variance and vehicle delay. The researchers, who conducted comparison, [35] compared RBF and FFNN in the area of PT trip flow.

A single hidden layer, which consists of a single layer of output nodes, has fed the inputs directly to the outputs via a series of weights. A single hidden layer selection, which is approximately equal to 84% among all layer types, has been frequently employed to construct ANN. Two hidden layers have been rarely practiced by the researchers [6, 19, 25, 37, 46, 50] while a single hidden layer has a common utilization. Besides above layers, two of the papers [31, 48] used five hidden layers that means ANFIS. The selection of hidden neuron number is as important as the determination of hidden layer number. When the analysis of hidden neuron number in the papers are made, it is obvious that 59% of the papers preferred 1-10 hidden neuron group. 27% of the papers preferred 11-20 hidden neuron group and the rest of the papers selected 21-30 hidden neuron and 30 and over, at the same rate: 7%.

The feed-back structure should be brought forth when the established ANN model has been run. Mainly, perceptron rule (42%) and gradient descent method (29%) are exerted in the most of the papers that is mostly related to the prediction and forecasting fields. Besides the two main methods above, regression and GA are performed at the rate of 9.6% and 6.45%, respectively. Moreover, the other methods are used in ANN.

When the established ANN model is executed, it should be evaluated with regard to the specific performance measurements to choose the best ANN model. SSE, MSE and MAPE have been mostly preferred in order to evaluate the performance of the chosen methods by the researchers.

Reviewing papers reveals that the subjects of ANN applications on PT are considered by many journals. Besides, the journals with more than two publications are illustrated in Fig 6. Among the journals, two of them are clearly more active than the others in ANN applications with PT: Transportation Research Part C (4 papers in various subjects) and Journal of Transportation Engineering (4 papers). While total number of papers in other journals has prominence, each of them published only one paper in this field.

In the light of papers, interest of countries in the research field revealed that the highest interested country is China with 27.08 %. U.S and Turkey, which have the similar interest with 14.58 %, following China. It can be concluded that China, U.S and Turkey have the most influence in this research field.

It is clear that ANN has both advantages and disadvantages [73,93]. The advantages of ANN are presented as follows:

- ANN can be trained with less formal statistical data, while most of statistical methods are parametric model that need higher background of statistic
- ANN can clearly uncover complex non-linear relationships between input and output parameters
- ANN can uncover all possible interactions between predictor variables
- ANN can be performed by using multiple different training algorithms

- ANN can generate its own organization or representation of the information that is received during learning time

The authors [80] applied an algorithm to predict bus arrival time in their paper and they asserted that the applied algorithm clearly outperforms ANN and K-NN alone in both accuracy and efficiency of the algorithm. While ANN has a lot of advantages that are presented above, it has a number of disadvantages as well. ANN disadvantages are presented as follows:

- ANN is considered as a black box and it is really hard to explicitly identify possible relationships
- ANN may need greater computational time
- ANN can be trapped to over-fitting

5. CONCLUSION

Since PT based on ANN has drawn great attention from the researchers for a long time, this brief literature is produced. The overall findings of this paper are summarized in the following.

i. First, most of the applications have aimed the prediction or forecasting. It can be said that data is an important part for prediction and forecasting in the phase of the training. Based on detailed analysis in the PT literature, it is easily seen that historical data and GPS have been used for obtaining the data in most of the studies.

ii. Second, a single hidden layer has mostly been employed to construct ANN. In some of the papers, pruning [64] and growing algorithm [65] have been offered to determine the size of neuron in a hidden layer as an accurate tool in this field. However, ANN architectures such as hidden layers and neurons have been determined by trial and error in most of the papers.

iii. Third, MLFF has been applied in most of the ANN architecture. Perceptron and gradient learning rules have been mostly employed in PT studies. However, perceptron rule has been mostly exerted in the training phase. After training phase, output is required to be evaluated for the best result. SSE, MSE and MAPE have been adjusted for performance evaluation method.

iv. Fourth, it is obvious that the supervised learning algorithms have been performed in most studies.

v. Hopfield and RBF network should be employed more as ANN architecture by combining the chaotic condition. Chaos theory can be applied to determine ANN input parameters. Thus, joint practice of ANN and chaos theory can be explored deeply in the field of transportation such as PT and BRT for future direction because the optimal input parameter has a remarkable prominence in these fields.

vi. Some papers in field of travel mode choice [69] and evaluation of bus transport reliability [70] with respect to simulation are studied by researchers. However, these papers can be applied with a suitable ANN models, since the simulation has a complex application structure and takes more time than ANN.

vii. The authors [80] applied an algorithm to predict bus arrival time in their paper and they asserted that the applied algorithm clearly outperforms ANN and K-NN alone in both accuracy and efficiency of the algorithm. Integration of ANN and K-NN may be researched by comparing their algorithm with regard to efficiency and accuracy.

For further future research, AURA K-NN, which is faster than traditional k-NN, can be utilized in the field of classification applications in PT. Moreover, joint practice of ANN and chaos theory can be explored deeply in the field of transportation such as PT and BRT for future direction because the optimal input parameter has a remarkable prominence in these fields.

Acknowledgements / Teşekkür

There is no conflict of interests regarding the publication of this paper.

Appendix A

Comprehensive applications in public transport based on ANN.

Authors	Objective of studies	Type to reach data	Number of input variables	Number of hidden layer	Number of hidden neurons	ANN model	Performance evaluation	Learning rule	Training method	Activation function
Chien et al. (2002)		Traffic surveillance system	5	Single	6	Multi layer Feed forward (MLFF)	SSE	Gradient descent	Backpropagation (BP)	Sigmoid
HU et al. (2002)	Forecast	Historical data and survey	20	Two	50,-	MLFF	SSE	Gradient d.	Backpropagation	Sigmoid*, purelin
Guan et al. (2002)		Historical data	14	Single	4	MLFF	-	-	Backpropagation	Tansig
Chen et al. (2004)	Forecast	-	4	single	-	MLFF	-	Perceptron	Backpropagation	*
Jeong et al. (2004)		GPS receiver	3	single	15	MLFF	-	Perceptron	Levenberg-Marquardt (BP)	*
Çelikoğlu et al. (2005)		Ticketing system	4	single	3	MLFF	SSE	perceptron	Backpropagation	*
Peters et al. (2005)	Prediction	Train stations	-	single	63	MLFF	-	-	Resilient (BP)	*
Çelikoğlu et al. (2005)		-	4	single	25	MLFF	-	Gradient d.	Radial basis function (BP)	*
Stella et al. (2006)	Construct opt. model	Train location database	6	single	-	MLFF	-	A new defined rule	BP	*
(Celikoglu, 2006)	Calibrate	-	n	single	-	MLFF	SSE, MSE	Gradient	BP	*
(Celikoglu, 2006)	Forecast	Ticketing system	4	single	-	MLFF	SSE	Regression	BP	*
Celikoglu et al. (2007)	Compare	Electronic ticketing system	4	Single	-	MLFF, RBF	SSE	Gradient	BP	*
Celikoglu et al. (2007)	Forecast	Electronic ticketing system	4	Single	3	MLFF, GRNN	-	Gradient	Levenberg-Marquardt (BP)	*
Chen et al. (2007)	Prediction	Auto. passenger counter	4	Two	-	MLFF	MSE	Gradient	BP	Sigmoidal, tangent
(Jing, 2008)	Prediction	Railway management	15	Single	30	MLFF	-	-	(Weight calibration) BP	*
Liu et al. (2008)	Solve TP	-	-	-	-	Recurrent	-	GA, Hopfield	(Weight calibration) BP	*
Çelebi et al. (2009)	Prediction	15-min observation	4	Single	At least 20	MLFF	MAPE, MSE	Perceptron	BP	Hyperbolic tangent
Shen et al. (2009)	Maximization	Traffic	-	Single	5	MLFF	-	-	BP	*
Çetiner et al. (2010)	Prediction	Historical data	4	Single	-	MLFF	-	-	(Weight calibration) BP	*
Yu et al. (2010)	Prediction	Historical data	5	Single	3	SVM Network	-	SVM	(Kalman filter) BP	*
Khosravi et al. (2011)	Minimization	GPS	-	two	-	MLFF	-	Delta rule	(User-def. cost function) BP	*
Nugroho et al. (2011)	Evaluation	Monitoring station	19	Single	4	MLFF	-	Perceptron	BP	*
Mazloumi et al. (2011)	Prediction	Signal control sys.	6	Single	-	MLFF	-	-	(Bayesian regulazation) BP	*
Yu et al. (2011)	Prediction	Video survey	3-4	Single	5	MLFF	-	K-nearest neighbours, SVM	BP	*
Mazloumi et al. (2011)	Prediction under uncertainty	GPS	uncertain	single	2-3	MLFF	-	-	(Bayesian regulazation and Levenberg-Marquardt) BP	*

Appendix A (continuing)

Comprehensive applications in public transport based on ANN (continuing).

Authors	Objective of studies	Type to reach data	Number of input variables	Number of hidden layer	Number of hidden neurons	ANN model	Performance evaluation	Learning rule	Training method	Activation function
Azadeh et al. (2012)	Forecast and evaluation	Railway	Predicted	Single	18	MLFF	-	A flexible ANN	BP	*
Mazloumi et al. (2012)	Forecast	GPS	4	Two	5 7 1 7 6 5 3 8	MLFF	SSE	Perceptron	BP	Sigmoidal
Wei et al. (2012)	Prediction	Historical data	3	Single	-	MLFF	MAPE	Perceptron	BP	*
Özuysal et al. (2012)	Prediction	Stations	10	Two	15 5	MLFF	-	Regression	BP	*
Yuen et al. (2013)	Route choice	Stations	3	Single	-	MLFF	-	Perceptron	BP	*
Goh et al. (2013)	Road safety	Traffic Incident Man. System	-	Single	4	MLFF	-	Regression	(Gradient descent) BP	*
Kadiyala et al. (2013)	Air quality of a PB	Database	5	Single	7	RBF	-	GA	BP	*
Pei et al. (2013)	Analysis	Survey	5	Two	8 3	MLFF	-	-	BP	*
Lin et al. (2013)	Prediction	GPS	4	Single	16	Hierarchical ANN	-	-	BP	Sigmoidal
LIAO et al. (2013)	Analysis	Survey	6	Single	15	MLFF	-	-	BP	*
Gurmu et al. (2014)	Prediction	GPS	4	Single	-	MLFF	-	Perceptron	(Gradient descent) BP	*
Wang et al. (2014)	Prediction	Historical data	6	Single	-	RBF	-	Gradient	BP	*
Jimenez et al. (2014)	Classification	Municipal bus company	19	Single	9	MLFF	-	Perceptron	BP	*
Lai et al. (2014)	Evaluation	Simulation results	6	Single	14	MLFF	-	-	BP	*
Ona et al. (2014)	Reduce the instability	Database	12	Single	6	MLFF	-	Perceptron	(Gradient descent) BP	Sigmoidal
Xie et al. (2014)	Prediction	Historical data	4	Single	15	MLFF	-	Perceptron	(Divide and conquer method) BP	*
Jonavic et al. (2014)	Develop a green vehicle dist. model	-	8	5 layer	X	ANFIS	-	-	BP	Bell-shaped
Garrido et al. (2014)	Analysis	Satisfaction survey	12	Single	6	MLFF	MAPE	Perceptron	BP	Sigmoidal
Ma et al. (2014)	Prediction	Smart card sys. And historical	3	Single	10	MLFF	-	-	BP	*
Cheng et al. (2014)	Decide the signalization	Sensors	-	5 layer	X	ANFIS	-	-	(Gradient descent) BP	Sigmoidal
Kadiyala and Kumar (2015)	Evaluation	Database	-	Single	5	MLFF	-	-	BP	-
Islam et al. (2016)	Evaluation	Database	-	Two	-	MLFF	-	GRNN	-	-
Liang et al. (2016)	Prediction	Database	-	Single	3	Recurrent	-	-	-	-

REFERENCES / KAYNAKLAR

[1] Alan T. Murray, Rex Davis, Robert J. Stimson and Luis Ferreira. Public Transportation Access. Transportation Research Part D: Transport and Environment 1998; 3(5), 319-328, Doi: 10.1016/S1361-9209(98)00010-8.

[2] James F. Sallis, Lawrence D. Frank, Brian E. Saelens and M. Katherine Kraft. Active transportation and physical activity: opportunities for collaboration on transportation and public health research. Transportation Research Part A: Policy and Practice 2004; 38(4), 249-268, Doi: 10.1016/j.tra.2003.11.003.

[3] Anil K. Jain, Jianchang Mao and Mohiuddin K.M. Artificial neural networks: a tutorial. Computer 1996; 29(3), 31-44, Doi: 10.1109/2.485891.

[4] Colin Fyfe. Artificial Neural Networks. Department of Computing and Information Systems, 1996.

[5] Mei Chen, Xiaobo Liu, Jingxin Xia and Steven I. Chien. A Dynamic Bus-Arrival Time Prediction Model Based on APC Data. Computer-Aided Civil and Infrastructure Engineering 2004; 19(5), 364-376, Doi: 10.1111/j.1467-8667.2004.00363.x.

[6] Abbas Khosravi, Ehsan Mazloumi, Saeid Nahavandi, Doug Creighton and J.W.C Van Lint. A genetic algorithm-based method for improving quality of travel time prediction intervals. Transportation Research Part C 2011; 19(6), 1364-1376, Doi: 10.1016/j.trc.2011.04.002.

[7] B. Gültekin Çetiner, Murat Sari and Oğuz Borat. A Neural Network Based Traffic-Flow Prediction Model. Mathematical and Computational Applications 2010; 15(2), 269-278

[8] Sudarmanto Budi Nugroho, Akimasa Fujiwara and Junyi Zhang. An empirical analysis of the impact of a bus rapid transit system on the concentration of secondary pollutants in the roadside areas of the TransJakarta corridors. Stochastic Environmental Research and Risk Assessment 2011;25(5), 655-669, Doi: 10.1007/s00477-011-0472-x.

[9] Fabio Stella, Vittorio Vigano, Davide Bogni and Matteo Benzoni. An Integrated Forecasting and Regularization Framework for Light Rail Transit Systems. Journal of Intelligent Transportation Systems 2006; 10(2), 59-73,Doi: 10.1080/15472450600626240.

[10] Ehsan Mazloumi, Geoff Rose, Graham Currie and Majid Sarvi. An Integrated Framework to Predict Bus Travel Time and Its Variability Using Traffic Flow Data. Journal of Intelligent Transportation Systems 2011;15(2), 75-90, Doi: 10.1080/15472450.2011.570109.

[11] J.K.K. Yuen, E.W.M. Lee, S.M. Lo and R.K.K. Yuen. An Intelligence-Based Optimization Model of Passenger Flow in a Transportation Station. IEEE Transactions on Intelligent Transportation Systems 2013; 14(3), 1290-1300, Doi: 10.1109/TITS.2013.2259482.

[12] Hilmi Berk Celikoglu. Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modeling. Mathematical and Computer Modelling 2006; 44(7-8), 640-658, Doi: 10.1016/j.mcm.2006.02.002.

[13] Zegeye Kebede Gurmu and Wei (David) Fan. Artificial Neural Network Travel Time Prediction Model for Buses. Journal of Public Transportation 2014; 17(2), 45-65.

[14] Kelvin Chun Keong Goh, Graham Currie, Majid Sarvi and David Logan. Bus accident analysis of routes with/without bus priority. Accident Analysis and Preventation 2014; 65, 18-27, Doi: 10.1016/j.aap.2013.12.002.

[15] Bin Yu, William H.K. Lam and Mei Lam Tam. Bus arrival time prediction at bus stop with multiple routes. Transportation Research Part C 2011;19(6), 1157-1170, Doi: 10.1016/j.trc.2011.01.003.

[16] Ranhee Jeong and Laurance R. Rilett. Bus Arrival Time Prediction Using Artificial Neural Network Model. IEEE Intelligent Transportation Systems Conference 2004;988-993, Washington, USA, Doi: 10.1109/ITSC.2004.1399041.

[17] Lei Wang, Zhongyi Zuo and Junhao Fu. Bus Arrival Time Prediction Using RBF Neural Networks Adjusted by Online Data. Procedia-Social and Behavioral Sciences 2014; 138, 67-75, Doi: 10.1016/j.sbspro.2014.07.182.

[18] Felipe Jimenez, Francisco Serradilla, Alfonso Roman and Roman Jose Eugenio Naranjo. Bus Line Classification Using Neural Networks. Transportation Research Part D 2014;30, 32-37, Doi: 10.1016/j.trd.2014.05.008.

[19] Akhil Kadiyala, Devinder Kaur and Ashok Kumar. Development of hybrid genetic-algorithm-based neural networks using regression trees for modeling air quality inside a public transportation bus. Journal of Air& Waste Management Association 2013;63(2), 205-218, Doi: 10.1080/10962247.2012.741054.

[20] Steven I-JyChien, Yuqing Ding and Chienhung Wei. Dynamic Bus Arrival Time Prediction with Artificial Neural Networks. Journal of Transportation Engineering 2002;128(5), 429-439, Doi: 10.1061/(ASCE)0733-947X(2002)128:5(429).

[21] AAzadeh, M Saberi, R Noorossana, Mohammad Saidi Mehrabad, M Anvari and H Izadbakhsh. Estimating efficient value of controllable variable using an adaptive neural network algorithm: Case of a railway system. Journal of Scientific and Industrial Research 2012;71(1), 45-50.

[22] H. Berk Çelikoğlu and Murat Akad. Estimation of Public Transport Trips by Feed Forward Back Propagation Artificial Neural Networks; a Case Study for Istanbul. Soft Computing: Methodologies and Applications Advances in Soft Computing 2005; 32, 27-36, Doi: 10.1007/3-540-32400-3_3.

[23] Yung-Cheng Lai, Yung-An Huang and Hong-Yu. Estimation of Rail Capacity Using Regression and Neural Network. Neural Computing and Applications 2014;25(7-8), 2067-2077, Doi: 10.1007/s00521-014-1694-x.

[24] Alvaro Costa and Raphael N. Markellos. Evaluating Public Transport Efficiency with Neural Network Models. Transportation Research Part C: Emerging Technologies 1997;5(5), 301-312, Doi: 10.1016/S0968-090X(97)00017-X.

[25] Ehsan Mazloumi, Sara Moridpour, Graham Currie and Geoff Rose. Exploring the Value of Traffic Flow Data in Bus Travel Time Prediction. Journal of Transportation Engineering 2012;138(4), 436-446, Doi: 10.1061/(ASCE)TE.1943-5436.0000329.

[26] Juan de Ona and Concepcion Garrido. Extracting the Contribution of Independent Variables in Neural Network Models: a New Approach to Handle Instability. Neural Computing and Applications 2014; 25(3-4), 859-869, Doi: 10.1007/s00521-014-1573-5.

[27] Teng Jing. Forecasting Railway Network Data Traffic: a Model and a Neural Network Solution Algorithm. 2008 Workshop on Power Electronics and Intelligent Transportation System2-3 August 2008; 203-208, place Guangzhou, Doi: 10.1109/PEITS.2008.23.

[28] Yu Wei and Mu-Chen Chen. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C 2012; 21(1), 148-162, Doi: 10.1016/j.trc.2011.06.009.

[29] Mei-Quan Xie, Xia-Miao Li, Wen-Liang Zhou and Yan-Bing Fu. Forecasting the Short-Term Passenger Flow on High-Speed Railway with Neural Networks. Computational Intelligence and Neuroscience 2014;2014,Doi: 10.1155/2014/375487.

[30] Junyan Liu, Honglian Yin, Wangling Qiu and Zhuofu Wang.GA-Hopfield Network for Transportation Problem. Wireless Communications, Networking and Mobile Computing 12-14 October 2008;1-4, place Dalian, Doi: 10.1109/WiCom.2008.1525.

[31] Aleksandar D. Jovanovic, Dragan S. Pamucar and Snezana Pejcic-Tarle. Green vehicle routing in urban zones- A neuro-fuzzy approach. Expert Systems with Applications 2014;41(7), 3189-3203, Doi: 10.1016/j.eswa.2013.11.015.

[32] Bin Yu, Zhong-Zhen Yang, Kang Chen and Bo Yu. Hybrid model for prediction of bus arrival times at next station. Journal of Advanced Transportation 2010;44(3), 193-204, Doi: 10.1002/atr.136.

[33] Dilay Çelebi, Bersam Bolat and Demet Bayraktar. Light Rail Passenger Demand Forecasting by Artificial Neural Networks. Computer & Industrial Engineering year 6-9 July2009; 239-243, place Troyes, Doi: 10.1109/ICCIE.2009.5223851.

[34] Hilmi Berk Celikoglu. Modeling Public Transport Trips with General Regression Neural Networks; A Case Study for Istanbul Metropolitan Area. Applications of Soft Computing: Advances in Intelligent and Soft Computing 2006;36, 271-280, Doi: 10.1007/978-3-540-36266-1_26.

[35] Hilmi Berk Celikoglu and Hikmet Kerem Cigizoglu. Modelling public transport trips by radial basis function neural networks. Mathematical and Computer Modelling 2007;45, 480-489, Doi: 10.1016/j.mcm.2006.07.002.

[36] Concepcion Garrido, Rocio de Ona and Juan de Ona. Neural networks for analyzing service quality in public transportation. Expert Systems with Applications 2014;41, 6830-6838, Doi: 10.1016/j.eswa.2014.04.045.

[37] Mustafa Özuysal, Gökmen Tayfur and Serhan Tanyel. Passenger Flows Estimation of Light Rail Transit (LRT) System in Izmir, Turkey Using Multiple Regression and ANN Methods. Promet-Traffic & Transportation 2012; 24(1), 1-14, Doi: 10.7307/ptt.v24i1.264.

[38] Zhenliang Ma, Jianping Xing, Mahmoud Mesbah and Luis Ferreira. Predicting short-term bus passenger demand using a pattern hybrid approach. Transportation Research Part C 2014; 39, 148-163, Doi: 10.1016/j.trc.2013.12.008.

[39] Ehsan Mazloumi, Geoff Rose, Graham Currie and Sara Moridpour. Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. Engineering Applications of Artificial Intelligence 2011; 24, 534-542, Doi: 10.1016/j.engappai.2010.11.004.

[40] Amer Shalaby and Ali Farhan. Prediction Model of Bus Arrival and Departure Times Using AVL and APC data. Journal of Public Transportation 2004; 7(1), 41-61.

[41] YU-LONG Pei, KAN Zhou and TING Peng. Prediction model of Passenger Waiting Time in High-Speed Rail Hub Based on BP Neural Network. Applied Mechanics and Materials 2013;321-324, 1903-1906, Doi: 10.4028/www.scientific.net/AMM.321-324.1903.

[42] Jan Peters, Bastian Emig, Marten Jung and Stefan Schmidt. Control and Automation 2005, Prediction of Delays in Public Transportation using Neural Networks. Computer Intelligence for Modelling28-30 November 2005;2, 92-97, place Vienna, Doi: 10.1109/CIMCA.2005.1631451.

[43] Hilmi Berk Celikoglu and Hikmet Kerem Cigizoglu. Public transportation trip flow modeling with generalized regression neural networks. Advances in Engineering Software 2007;38, 71-79, Doi: 10.1016/j.advengsoft.2006.08.003.

[44] Hilmi Berk Celikoglu. Radial Basis Function Neural Network Approach to Estimate Public Transport Trips in Istanbul. Soft Computing as Transdisciplinary Science and Technology Advances in Soft Computing 2005;29, 31-40, Doi: 10.1007/3-540-32391-0_11.

[45] Yongjie Lin, Xianfeng Yang, Nan Zou and Lei Jia. Real-Time Bus Arrival Time Prediction: Case Study for Jinan, China. Journal of Transportation Engineering 2013;139(11), 1133-1140, Doi: 10.1061/(ASCE)TE.1943-5436.0000589.

[46] HU Jian-ming, SONG Jing-yan, ZHANG Yi and YANG Zhao-sheng. Study on Automatic Creating Method of Public Transportation Dispatching From Based on BP Neural Network. The IEEE 5th International Conference on Intelligent Transportation

Systems3-6 September 2002; 863-867, place Singapore, Doi: 10.1109/ITSC.2002.1041333.
[47] Guojiang Shen and Xiangjie Kong. Study on Road Network Traffic Coordination Control Technique with Bus Priority. IEEE Transactions on Systems, man, and Cybernetics-part c: applications and reviews 2009;39(3), 343-351, Doi: 10.1109/TSMCC.2008.2005842.
[48] Sheng-Tzong Cheng, Jian-pan Li, Gwo-Jiun Horng and Kuo-Chuan Wang. The Adaptive Road Routing Recommendation for Traffic Congestion Avoidance in Smart City. Wireless Personal Communications 2014;77(1), 225-246, Doi: 10.1007/s11277-013-1502-4.
[49] Yong LIAO and Tao Wu. The Analysis of Demand Characteristics of Passenger Transportation Based on BP Neural Network. Applied Mechanics and Materials2013;409-410, 1292-1295, Doi: 10.4028/www.scientific.net/AMM.409-410.1292.
[50] Mei Chen, Jason Yaw, Steven I. Chien and Xiaobo Liu. Using Automatic Passenger Counter Data in Bus Arrival Time Prediction. Journals of Advanced Transportation 2007;41(3), 267-283, Doi: 10.1002/atr.5670410304.
[51] Wei Guan, Jinsheng Shen and Wanping Wang. Using S-ANN Method to Forecast the Ridership of Beijing Public Transportation. International Conference on Traffic and Transportation Studies (ICTTS)23-25 July 2002; 877-882, place Guilin China, Doi: 10.1061/40630(255)122.
[52] Daniel Graupe. Principles of Artificial Neural Networks. Advanced Series in Circuits and Systems Vol. 6, 2rd Edt. Chicago, USA. Word Scientific Publishing Co. Pte. Ltd (Singapore), 1995.
[53] Mark J. L. Orr. Introduction to Radial Basis Function Networks. Centre for Cognitive Science, 1996.
[54] Simon Haykin. Signal Processing, Kalman Filtering and Neural Networks, a Wiley-Interscience Publication, John Wiley & sons inc., 2001.
[55] Michael R. Chernick. Wiley Series in Probability and Statistics, Bootstrap Methods: A Guide for Practitioners and Researcher, Second Edt., year 30 April 2007, John Wiley & SONS INC., Doi: 10.1002/9780470192573.
[56] Martin Riedmiller and Heinrich Braun. A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. IEEE International Conference on Neural Networks28 March-01 April 1993; 1, 586-59, place San Francisco, Doi: 10.1109/ICNN.1993.298623.
[57] Asif Ullah Khan, T.K. Bandopadhyaya and Sudhir Sharma. Genetic Algorithm Based Backpropagation Neural Network Performs Better than Backpropagation Neural Network in Stock Rates Prediction. International Journal of Computer Science and Network Security 2008;8(7), 162-166
[58] Xin Yao. Evolving artificial neural network. Proceedings of the IEEE 1999; 87(9), 1423-1447, Doi: 10.1109/5.784219.
[59] Vojislav Kecman. Neural Networks and Fuzzy Logic Models. Learning and Soft Computing: Support Vector Machines. 1st edition, The MIT Press, 2001.
[60] Jyh-Shing Roger Jang.ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Transactions on Systems, Man, and Ceybernetics 1993; 23(3), Doi: 10.1109/21.256541.
[61] Ting Liu, Andrew W. Moore and Alexander Gray. New Algorithms for Efficient High-Dimensional Nonparametric Classification. The Journal of Machine Learning Research 2006,7, 1135-1158, 2006.
[62] Nader Salari, Shamarina Shohaimi, Farid Najafi, Meenakshii Nallappan and Isthrinayagy Karishnarajah. A Novel Hybrid Classification Model of Genetic Algorithms, Modified k-Nearest Neigbor and Developed Backpropagation Neural Network. Plus One 2014;9(11), Doi: 10.1371/journal.pone.0112987.

[63] Victoria J. Hodge, Rajesh Krishnan, Jim Austin, John Polak and Tom Jackson. Short-term prediction of traffic flow using a binary neural network. Neural Computing and Applications 2014;25(7-8), 1639-1655, Doi: 10.1007/s00521-014-1646-5.

[64] Russell Reed. Pruning Algorithms-A Survey. IEEE Transactions on Neural Networks 1993;4(5), 740-747, Doi: 10.1109/72.248452.

[65] Visakan Kadirkamanathan and Mahesan Niranjan. A Function Estimation Approach to Sequential Learning with Neural Networks. Neural Computation 1993;5(6), 954-975, Doi: 10.1162/neco.1993.5.6.954.

[66] Chen, Mei, Xiaobo Liu, and Jingxin Xia. “Dynamic prediction method with schedule recovery impact for bus arrival time.” Transportation Research Record: Journal of the Transportation Research Board 1923 (2005): 208-217.

[67] Padmanaban, R. P., Lelitha Vanajakshi, and Shankar C. Subramanian. “Estimation of bus travel time incorporating dwell time for APTS applications.” Intelligent Vehicles Symposium, 2009 IEEE. IEEE, 2009.

[68] Vanajakshi, Lelitha, Shankar C. Subramanian, and R. Sivanandan. “Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses.” IET intelligent transport systems 3.1(2009):1-9.

[69] Bhatta, B. P., & Larsen, O. I. Errors in variables in multinomial choice modeling: A simulation study applied to a multinomial logit model of travel mode choice. Transport policy (2011); 18(2), 326-335.

[70] Ap. Sorratini, J., Liu, R., & Sinha, S. Assessing bus transport reliability using micro simulation. Transportation Planning and Technology 2008; 31(3), 303-324.

[71] Bilişik, Ö. N., Erdoğan, M., Kaya, İ., & Baraçlı, H. (2013). A hybrid fuzzy methodology to evaluate customer satisfaction in a public transportation system for Istanbul. Total Quality Management & Business Excellence, 24(9-10), 1141-1159.

[72] Cascetta, E., & Cartenì, A. (2014). A quality-based approach to public transportation planning: theory and a case study. International Journal of Sustainable Transportation, 8(1), 84-106.

[73] Nuzzolo, A., & Comi, A. (2016). Advanced public transport and intelligent transport systems: new modelling challenges. Transportmetrica A: Transport Science, 1-26.

[74] Moreira-Matias, L., Cats, O., Gama, J., Mendes-Moreira, J., & de Sousa, J. F. (2016). An online learning approach to eliminate Bus Bunching in real-time. Applied Soft Computing, 47, 460-482.

[75] Freitas, A. L. P. (2013). Assessing the quality of intercity road transportation of passengers: An exploratory study in Brazil. Transportation Research Part A: Policy and Practice, 49, 379-392.

[76] Islam, M. R., Hadiuzzaman, M., Banik, R., Hasnat, M. M., Musabbir, S. R., & Hossain, S. (2016). Bus service quality prediction and attribute ranking: a neural network approach. Public Transport, 8(2), 295-313.

[77] Shen, J., & Li, W. (2014). Discrete Hopfield Neural Networks for Evaluating Service Quality of Public Transit. International Journal of Multimedia and Ubiquitous Engineering, 9(2), 331-340.

[78] León, M., Mkrtchyan, L., Depaire, B., Ruan, D., & Vanhoof, K. (2014). Learning and clustering of fuzzy cognitive maps for travel behaviour analysis. Knowledge and information systems, 39(2), 435-462.

[79] Liang, V. C., Ma, R. T., Ng, W. S., Wang, L., Winslett, M., Wu, H., ... & Zhang, Z. (2016, May). Mercury: Metro density prediction with recurrent neural network on streaming CDR data. In Data Engineering (ICDE), 2016 IEEE 32nd International Conference on (pp. 1374-1377). IEEE.

[80] Dong, J., Zou, L., & Zhang, Y. (2013, June). Mixed model for prediction of bus arrival times. In 2013 IEEE Congress on Evolutionary Computation (pp. 2918-2923). IEEE.

[81] Kadiyala, A., & Kumar, A. (2016). Univariate time series based radial basis function neural network modeling of air quality inside a public transportation bus using available software. Environmental Progress & Sustainable Energy, 35(2), 320-324.

[82] Dadula, C. P., & Dadios, E. P. (2015, December). Neural network classification for detecting abnormal events in a public transport vehicle. In Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2015 International Conference on(pp. 1-6). IEEE.

[83] Segundo, F. R., e Silva, E. S., & Farines, J. M. (2014, October). Predicting journeys for DTN routing in a public transportation system. In 2014 IEEE 10th international conference on wireless and mobile computing, networking and communications (WiMob) (pp. 494-499). IEEE.

[84] Dou, M., He, T., Yin, H., Zhou, X., Chen, Z., & Luo, B. (2015, June). Predicting passengers in public transportation using smart card data. In Australasian Database Conference (pp. 28-40). Springer International Publishing.

[85] Zhou, C., Dai, P., Wang, F., & Zhang, Z. (2016). Predicting the passenger demand on bus services for mobile users. Pervasive and Mobile Computing, 25, 48-66.

[86] Omrani, H. (2015). Predicting Travel Mode of Individuals by Machine Learning. Transportation Research Procedia, 10, 840-849.

[87] Ghanim, M. S., & Abu-Lebdeh, G. (2015). Real-time dynamic transit signal priority optimization for coordinated traffic networks using genetic algorithms and artificial neural networks. Journal of Intelligent Transportation Systems, 19(4), 327-338.

[88] Düzenli, G. (2015). RFID card security for public transportation applications based on a novel neural network analysis of cardholder behavior characteristics. Turkish Journal of Electrical Engineering & Computer Sciences, 23(4), 1098-1110.

[89] Deng, L., He, Z., & Zhong, R. (2013, November). The Bus Travel Time Prediction Based on Bayesian Networks. In Information Technology and Applications (ITA), 2013 International Conference on (pp. 282-285). IEEE.

[90] Wang, S., Zhou, R., & Zhao, L. (2015). Forecasting Beijing Transportation Hub Areas's Pedestrian Flow Using Modular Neural Network. Discrete Dynamics in Nature and Society, 2015.

[91] Kadiyala, A., & Kumar, A. (2015). Univariate time series based back propagation neural network modeling of air quality inside a public transportation bus using available software. Environmental Progress & Sustainable Energy, 34(2), 319-323.

[92] Kadiyala, A., & Kumar, A. (2016). Univariate time series based radial basis function neural network modeling of air quality inside a public transportation bus using available software. Environmental Progress & Sustainable Energy, 35(2), 320-324.

[93] Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology, 49(11), 1225-1231.