FUNAI Journal of Science & Technology,
2(1), 2016, 104-114

MODELING STUDENTS' ACADEMIC PERFORMANCE USING ARTIFICIAL NEURAL NETWORK

O. C. Asogwa[1] and C. G. Udomboso[2]

[1]*Department of Maths/Comp. Sc./Stats./Info., Federal University,Ndufu-Alike, Ikwo, Nigeria*
[2]*Department of Statistics, University of Ibadan, Ibadan, Nigeria*
(Received:20 June, 2016; Revised:15 August,2016; Accepted:17 August, 2016)

Abstract

Artificial Neural Network has been discovered as a better alternative to traditional models and that is why a model based on the Multilayer Perceptron algorithm was developed in this study. The appropriate number of hidden neurons that best modeled the academic performance of students was determined by the developed Network algorithm. Test data evaluation showed that Network Architecture 17 - 80 – 1 was chosen among the numerous developed network architectures because of its model performances. The chosen network architecture gave the minimum value of Mean Square Error (*MSE* = 0.0718), minimum value of Network Information Criteria (*NIC* = 0.0743), maximum value of *R*- Square (R^2=0.8975) and maximum value of Adjusted Network Information Criteria (*ANIC* = 0.8931). It was equally observed that there were patterns in the movement of hidden neurons against the model evaluation criteria. As the number of the hidden neurons appreciates the value of both *MSE* and *NIC* decreases down the plot, while that of *R*-Square and *ANIC*values appreciate down the plot. The network was able to model the research problem with acceptable values judging from the model checking criteria considered in this work. Also the order of contribution of the predictor variables to the model was determined.

***Keywords**: Modeling, academic performance, hidden neurons, Artificial Neural Network and model selection criteria.*

1. Introduction

The main product of universities and higher institution of learning is students. Upon graduation, the students may either continue their studies into the postgraduate programme or become the manpower for the industry, government and private sectors. Thus, modeling of the students' academic performances are critical in ensuring that the supply chain is sustained and that is why some higher institutions of learning have developed interest in predicting the paths of students, thus identifying which student will require assistance in order to graduate at the stipulated time or maintain their studies.

Moreover, there has been a tremendous growth in the interest of application of the Artificial Neural Networks (ANNs) in the modeling of academic performance of students since the 1990s. This is because Artificial Neural Networks are usually assumed to be powerful tools for functional relationship establishment or nonlinear mapping in various applications. The history of neural networks begins with the earliest model of the biological neuron given by McCulloch and Pitts in 1943. They are described as a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs. Artificial Neural Networks are typically organized in layers. Layers are made up of a number of interconnected 'nodes' which contain an 'activation function'. Patterns are presented to the network via the 'input layer', which communicates to one or more 'hidden layers' where the actual processing is done via a system of weighted 'connections'. The hidden layers are then link to an 'output layer'. Artificial Neural Networks, which are of many types (FFNN - Feed forward neural network, RBF- Radial basis function, Kohonenself-organizing, etc) have the ability to recognize complex patterns quickly with a high degree of accuracy, it makes no assumptions about the nature and distribution of the data and they are not biased in their analysis. In addition, Artificial Neural Networks have non-linear tools and as such are good at predicting non-linear behaviors. Artificial neural networks form a broad category of computer algorithms that solve several types of problems, including mapping relationship, pattern classification, functions approximation, pattern completion, pattern association, filtering, optimization and automatic control (Mohaghegh *et al.*, 2000). The structure of the Artificial Neural Network is defined by the interconnection architecture

architecture between the processing elements. Information processing within an Artificial Neural Network occurs in the processing elements that are called neurons which are grouped into layers. Signals are passed between neurons over the connecting links as shown in Figure 1. In Fig. 1, there are n- inputs into the network, therefore, there are n- input neurons in the first layer; there is only one output from the network, consequently, there will be one output neurons in the output layer. The number of middle layers and the number of neurons in these layers a

performance: Karamouzis and Vrettos, (2009) used Artificial Neural Network model to predict community college graduation outcomes as well as the results of applying sensitivity analysis on the Artificial Neural Network parameters in order to identify the factors that affected successful graduation outcome. The need for disability services, the need for support services and the student's age at the time of application to the college were identified as the three factors that contributed most to student's graduation outcome. Naik and Ragothaman (2004) developed a model to predict MBA student

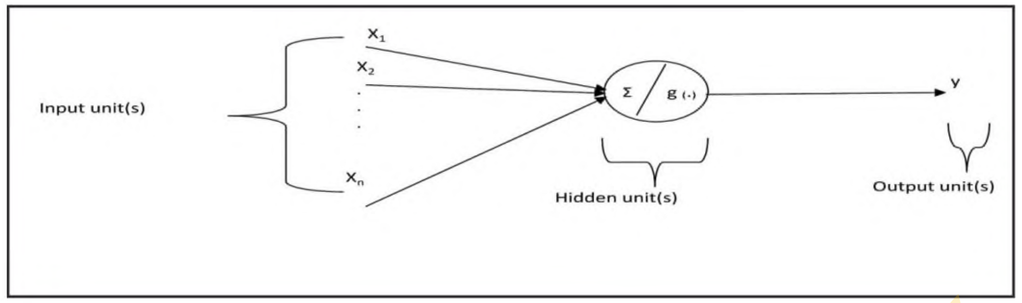

Fig. 1. The structure of the Artificial Neural Network

study is to evaluate the accuracy of the Artificial Neural Network Model formulated using its Mean Square Error (*MSE*), Network Information Criteria (*NIC*), *R*-Square and Adjusted Network Information Criteria (*ANIC*). A number of researches have been conducted on the students' academic

performance using Logistic Regression, Probability Analysis and the Artificial Neural Network. Their result indicated that Artificial Neural Network model performs better than the other considered models. In addition, other scholars have equally carried out a scholarly study on modeling academic

performance of students: Walczak and Sincich, (1999) compared the results of Artificial Neural Networks model to that of Logistic Regression Analysis for modeling student enrollment decision-making. Result of the work concluded that the level of performance of the Artificial Neural Network is not significantly higher than that of the other model, but performed better off. Arinzeet *al.*(2000) compared Artificial Neural Network model to Regression model in the prediction of groups of MBA student's performance. They concluded that Artificial Neural Network modelperformed better than Regression model with a lower bias. Asogwa, and Oladugba (2015a) used multilayer (FFNN) Artificial Neural Network model topredict academic performance rates of university students. The result showed about 97% accuracy prediction. Asogwa and Oladugba (2015b) compared a multilayer Artificial Neural Network and Multinomial Logistic Regression. Multilayer Artificial Neural Network model performed better than the other. Kanakana and Olanrewaj (2011) utilized a Multilayer Perception Neural Networks to model students' performance. The result showed that Artificial Neural Networks-based model is able to model students' performance with high accuracy. Lykourentzou *et al.* (2009) used a Multiple Feed-Forward Neural Networks to model students' final achievements. The result was grouped into good and poor performances. This work is actually an extension of some of the works reviewed in this paper, since it included some of all the possible factors that were not seen in some of the reviewed works that affects students' academic performance. Some of the literatures used parent's education, parent's occupation as a single factor whereas some didn't consider a factor of disability.

1. Methodology

Through extensive review of literatures, a number of socio-economic, biological, environmental, academic, and other related factors like pre-admission requirements that are considered to have influence on the academic performance of university students were identified and reconciled with the information we got in the students' admission records. Some of these factors were: Gender, Parents marital status, father's education, mother's education, father's occupation, mother's occupation, O' level results, Age at entry, time delayed before admission, type of secondary school attended, location of secondary school attended, post UTME, and disability. The data used in this study were secondary data and was collected from the admission record of students of the faculty of Veterinary Medicine, University of Nigeria, Nsukka. The data were collected from 2008 -2014 set in the faculty at their penultimate year. Only Students' records with complete information were accessed in this study and that amount to a total of 450 students' record reviewed. These factors that were collected from the students' records considered the output variable as the performance of students at their penultimate year (CGPA) and others as the predictor variables. With the help of MatlabR2009a as a statistical tool, neural codes were generated for the variables and were properly analyzed under Artificial Neural Network.

1.1 The Artificial Neural Network Model

Artificial Neural Network (ANN) model proposed by Anders (1996) was used in his research work. The model is given as:

$$y = f(X,W) + e_i$$

$$y = f(X,W) + e_i = \alpha X + \sum_{h=1}^{H} \beta_h \, g\Big(\sum_{i=0}^{I} \gamma_{h_i} x_i\Big) + e_i$$

$$y = f(X,W) + e_i = \alpha X + \sum_{h=1}^{H} \beta_h \left(1+e^{-x}\right)^{-1} \Big(\sum_{i=0}^{I} \gamma_{h_i} x_i\Big) + e_i \qquad (1)$$

$$g(.) = \frac{1}{1+e^{-x}} = \left(1+e^{-x}\right)^{-1}$$

y is the ouput variable

X is the input variables

where:

$X = (x_0 = 0, x_1, \ldots, x_I)$

$W = (\alpha, \beta, \gamma)$

α is the weight of the input unit(s)

β is the weight of the hidden unit(s)

γ is the weight of the output unit(s)

$g(.)$ is the logistic transfer function

e_i is the error term

2.2 The network architecture and design

Multilayer Perceptrons (MLPs) are layered feed forward networks typically trained with static back propagation. Therefore, given the computational capabilities of a multilayer perceptron a three-layered Feed Forward Neural Network was developed (Algorithm) in this study in line with the Anders (1996) model. The first layer (input level) comprised of 17 neurons (processing elements) - one for each profile parameter (input). The third layer(output level) comprised 1 neurons denoting academic students' performance ordinary level result was initiated as four profile parameters. This explains why in the table1 there were 17 neurons (processing elements) in the input layer, a varying number of

neurons in the hidden layer and 1 neuron in the output layer. This is denoted by 17 – (varying neurons) – 1, as used in this work. However, based upon recommendations from Cybenko (1989) and Hornik *et al.*, (1989) that one hidden-layer network is sufficient to model any complex system, the designed network model used only one hidden layer. Also, one of the best ways to determine the number of hidden neurons in the hidden layer is by trial and error, which incorporates the rule of thumb. The rule of thumb is to start with the smallest size possible for a given problem to allow for generalization, then to increase the size of the hidden neurons, until the optimal results are achieved, Zhang and Trimble (1996).In this study, it was decided to fix different number of neurons in the hidden layer and through evaluation of their model performance, pick the one that gives the optimal performance among others. We started by choosing one neuron in the hidden layer, followed by two neurons, etc. Back-propagation learning algorithm was used for training the network. The logistic activation function was used at the hidden layer and the identity activation function was used at the output layer. The network architecture was run for about 100 epochs. The number of hidden neuron cannot be limited to 100 neurons. It can exceed 100 neurons but this work only considered from 1neuron to 10neurons and 10neurons to 100neurons on 10unit interval.

2.3 Model selection criteria

The model checking criteria used in this study were the one partaking to Artificial Neural Network modeling, apart from the commonly known model selection criteria. The model selection criteria used in this study were the Coefficient of Determination (*R*-Square), the Mean Square Error (*MSE*), the Network Information Criteria (*NIC*), and the Adjusted or Alternative Network Information Criteria (*ANIC*). We are so familiar with the first two and they are commonly used in OLS (ordinary least squares) modeling, but the last two are model selection criteria popularly used in Artificial Neural Network modeling. There are more other model selection criteria, but these ones were considered in this work, although they were not chosen on a criterion. Both MSE and NIC are declared acceptable (optimum) when their values are low (close to 0). Also, ANIC and R – square are optimum when their values are high (close to 1). All their values lie between 0 and 1.

3. Results

The Table containing the developed network architecture with the evaluated model evolution criteria is shown below:

Table 1: Table of model architecture with model evaluation criteria

S/N	Networks at different hidden neurons	*MSE*	*NIC*	*ANIC*	*R-square*
1	17 – 1 – 1	0.4760	0.4695	0.3203	0.2916
2	17 – 2 – 1	0.3980	0.2345	0.2920	0.2350
3	17 – 3 – 1	0.3592	0.1791	0.4873	0.4657
4	17 – 4 – 1	0.3082	0.5264	0.5602	0.5416
5	17 – 5 – 1	0.3521	0.1930	0.4974	0.4762
6	17 – 6 – 1	0.3844	0.3920	0.4514	0.4282
7	17 – 7 – 1	0.3077	0.2857	0.5608	0.5423
8	17 – 8 – 1	0.3646	0.2902	0.4796	0.4576
9	17 – 9 – 1	0.2139	0.1956	0.6947	0.6818
10	17 – 10 – 1	0.2997	0.2877	0.5723	0.5542
11	17 – 20 – 1	0.1520	0.1612	0.7831	0.7739
12	17 – 30 – 1	0.1607	0.1433	0.7707	0.7610
13	17 – 40 – 1	0.1951	0.2130	0.7216	0.7098
14	17 – 50 – 1	0.1466	0.1518	0.8687	0.8622
15	17 – 60 – 1	0.0920	0.0930	0.7907	0.7819
16	17 – 70 – 1	0.1353	0.1394	0.8069	0.7987
17	**17 – 80 – 1**	**0.0718**	**0.0743**	**0.8975**	**0.8931**
18	17 – 90 – 1	0.1195	0.1476	0.8294	0.8222
19	17 – 100 – 1	0.1453	0.1464	0.7926	0.7838

From the Table 1 above, it was observed that as the number of hidden neurons increase, the value of *MSE* decreases generally. Based on the need to come with a better architecture that will best model the study problem, several model architectures were evaluated in accordance with the formulated algorithm. The architecture 17 – 80 – 1 gave the minimum value of *MSE* among others. It can also be pictured in the same table above that the Network Information Criteria decreases with an increase in the number of hidden neurons. At the architecture 17 – 80 – 1, the value of the *NIC* was the minimum. The behavior of the *R* – Square (R^2) and the Adjusted Network Information Criteria (*ANIC*) were opposite of both the *MSE* and the *NIC*. Column 5 in the table one which represented the R- Square showed that as the number of hidden neurons increases down the table, the value of *R* – Squared increases considerably. It was observed that the architecture 17 – 80 – 1, evaluation contained the maximum value of *R* – Square. The sixth column in the same table one showed the computed values for *ANIC* at different number of hidden neurons. Adjusted Network Information Criteria (*ANIC*) showed a steady

increase down the table as the number of hidden neurons increases. The maximum value of *ANIC* was given by the network architecture 17 - 80 – 1.In addition, the pattern or the behavior of the model evaluation criteria across the number of hidden neurons can equally be justified by the use of graph plots. The graph plots which described the behaviour is illustrated in Fig. 2.

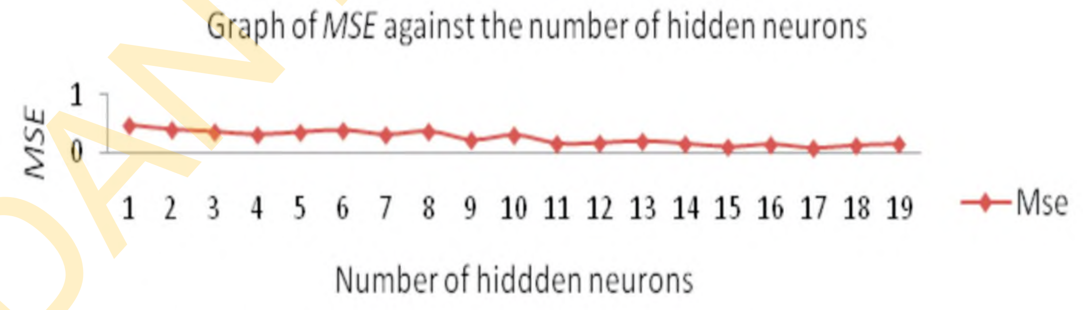

Fig. 2. Graph of *MSE* against the number of hidden neurons

The justification of the patterned movement of the *MSE* down the table as the number of hidden neurons considerably increase from 1 to 10 on a 1 interval unit and from 10 to 100 on a 10 interval units can be viewed in Fig 2 above. The value of *MSE* decreases as the number of hidden neurons increases. Such movement can be seen in Fig. 2 above;

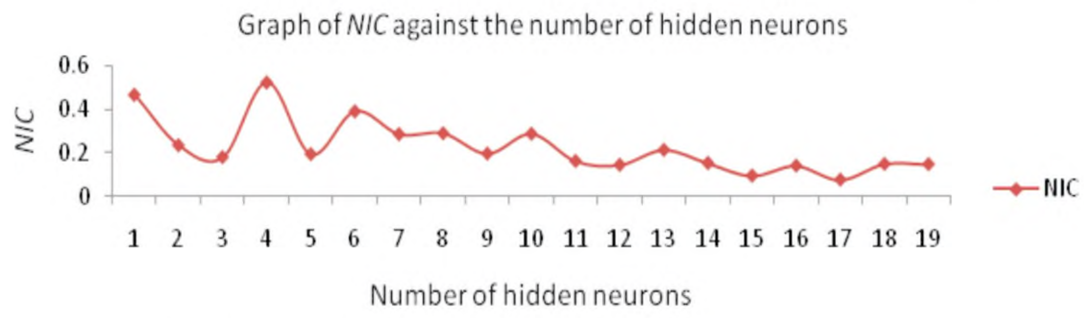

Fig. 3. Graph of *NIC* against the number of hidden neurons

The graph plot in Fig. 3 showed the pattern movement of the Network Information Criteria across the number of the hidden neurons. The movement of the Network Information Criteria (*NIC*) was such that as the number of the hidden neurons increases, the value of the Network Information Criteria decreases. The graph plot in Fig. 2 has the same nature as the one in Fig. 3, which show that both the Mean Square Error and the Network Information Criteria of the Network Architecture formulated helped in minimizing the error due to the network as the numbers of hidden neurons were increased. Therefore, Network Information Criteria (*NIC*) decreases in value as the number of the hidden Neurons increases. However, another justification was that of the *R* – Square which is also called coefficient of determination. The plots were such that as the number of the hidden neurons increases down the plot, the coefficient of determination *R* – Square value increases as well. It maximizes the error in the network and that's why the network architecture with the maximum value of *R* – Square was the best in tackling our study problem. From Fig. 3, it could be observed that it is opposite of both the *MSE* and the *NIC*. This is because the value of the *R* – Square appreciates as the number of the hidden neurons appreciates; the reverse is the case for both the *MSE* and the *NIC*. The graph is seen below.

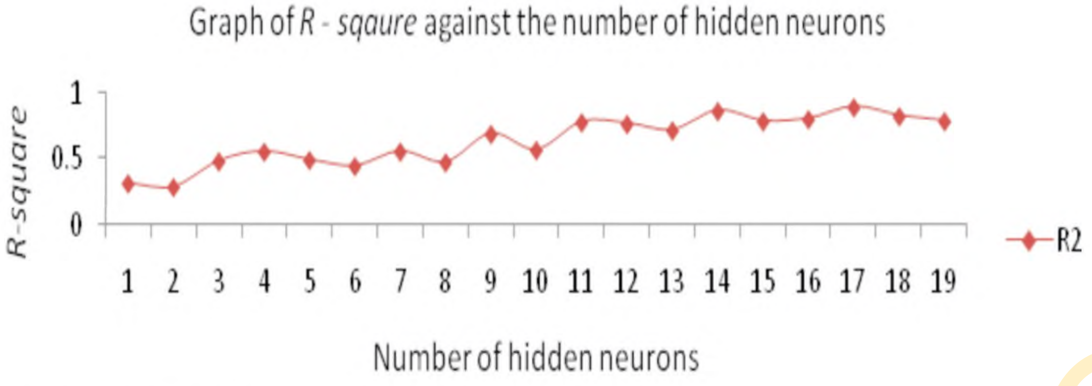

Fig. 4. Graph of *R* – square against the number of hidden neurons

Consequently, Fig. 5 below, depicted the graph plot of the Alternative Network Information Criteria (*ANIC*) against the number of the hidden neurons. So Fig. 5 also followed the pattern of Fig. 4 since the movement is in the same direction. That is, as the number of hidden neurons increases, the value of the Alternative Network Information Criteria (*ANIC*) appreciates as well. This justification was equally very important in other to detect why the considered network architecture was used to solve our study problem. Fig. 5 can be observed below.

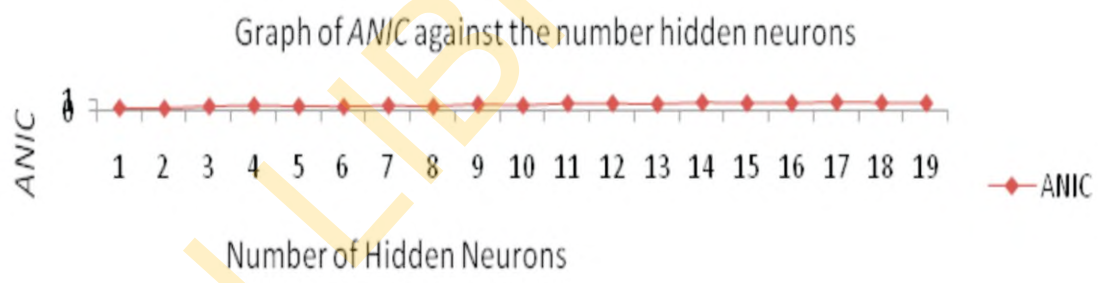

Fig. 5. Graph of *ANIC* against the number of the hidden neurons

4. Conclusion

From the analysis carried out, we noticed that an appropriate model architecture which tackled the study problem was developed. Among the numerous evaluated Network architectures, the recommended Network that well modeled our research problem very well was network architecture 17 – 80 – 1. The network is one among the 19 different network architectures considered. It was equally noticed that at this chosen network;17 – 80 – 1, the value of MSE was the minimum (0.0718).Moreover, its NIC realization was equally the minimum (0.0743)R – Square obtained the maximum value (0.8975) at the same network architecture 17 – 80 – 1. Likewise the ANIC value which appreciates as the number of the hidden neurons increases, showed the maximum value (0.8931) at the same network architecture 17 – 80 – 1. In addition, we noticed that the coefficient values of predictor variables like mother's education, father's occupation, O' level Mathematics, O' level Chemistry, O' level Biology, type of secondary school attended, location of secondary school attended and PUTME were positive, whereas the others were negative. In conclusion, we were able to see that the influencing factors (students' admission records) which contribute to students' academic performances can appropriately be modeled by the use of the formulated network model 17 – 80 – 1.

References

Anders, U. (1996). Model selection in neural networks, ZEW *Discussion Papers 96-21*. Retrieved from http://hdl.handle.net/10419/29449

Arinze, B., Subba Narasimha, P. N, & Anandarajam, M. (2000). The predictive accuracy of artificial neural networks and multiple regressions in the case of skewed data: Exploration of some issues. *Expert systems with Applications,* **19**, 117-123.

Asogwa, O. C. and Oladugba, A. V. (2015a)

Of Students Academic Performance Rates Using Artificial Neural Networks (ANNs). *American Journal of Applied Mathematics and Statistics, 2015, **3**, No. 4, 151-155.*

Asogwa, O. C. and Oladugba, A. V. (2015b).On The Comparison of Artificial Neural Network (ANN) and Multinomial Logistic Regression (MLR). *West African Journal of Industrial and Academic Research **13** No. 1.*

Cybenko, G. (1989).Approximation by superpositions of a sigmoidal function. *Journal Mathematics of Control, Signals, and Systems,***2** **(**4), 303-314.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feed forward networks are universal approximators, Neural Networks, vol. **2**, (5), 359-366.

Kanakana, G., & Olanrewaju, A. (2011).Predicting student performance in engineering education using an artificial neural network at Tshwane University of Technology, *Proceedings of the ISEM,* Stellenbosch, South Africa.

Karamouzis, S. T and Vrettos, A. (2009).Sensitivity Analysis of neural Network parameters for identifying the factors for college students' success. *World Congress on Computer Science and Information Engineering*. Los Angeles, USA, March 31th –April 2nd, 671-675.

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., and Loumos, V. (2009).Early and dynamic student achievement prediction in e-learning courses using neural networks. *Journal of the American Society for Information Science and Technology*, **60** (2), 372-380.

McCulloch, W.S. and Pitts, W. H. (1943). A logical Calculus of the ideas immanent in nervous activity. *Bulletin of mathematical Biophysical*, **5**, 115-133.

Mohaghegh, S., Reeves, S., and Hill, D., (2000).Development of an Intelligent Systems Approach for Restimulation candidate.*Proceedings of SPE/ CERT Gas Technology Symposium*, Calgary.**3-5**, 1-11.

Naik, B. & Ragothaman, S. (2004).Using Neural Networks to predict MBA student success, *College Student Journal*, vol. **38**, (1), 143-149.

Walczak, S., & Sincich, T. (1999).A comparative analysis of regression and neural works for university admissions. *Information Sciences*, **119** (1-2), 1-20.

Zhang, E.Y., P.E. and P. Trimble.(1996). Forecasting Water Availability by Applying Neural networks with Global and Solar Indices. *Proceedings of the Sixteenth Annual American Geophysical Union,* 1-10.