BOD MODELLING USING ARTIFICIAL NEURAL NETWORK

Roshni R[1], Elizabeth C Kuruvila[2]

[1] MTech Student, Civil Engineering, KMCT College of Engineering for women, Kerala, India
[2] Professor, Civil Engineering, KMCT College of Engineering for women, Kerala, India

ABSTRACT

Water quality modelling is required for proper water quality conservation and management. Limitation of fresh water sources suggests the need for water quality protection because it influences the lives of millions people. Water quality in the superficial waters has started to degenerate as a result of wastewater being let go to the receiving surface water without any control. The need for increased accuracies in modelling water quality has motivated the researchers to develop innovative models. Artificial neural networks (ANN) are capable of identifying the complex nonlinear relationships between input and output data. Biochemical Oxygen Demand (BOD) is an important parameter for usage conditions of surface waters. The aim of this study is to predict one of the most important water quality parameters, BOD, with the help of water temperature, hardness, dissolved oxygen and electrical conductivity data. This study involves the application of ANN based on monthly BOD modelling of Chaliyar river. Chaliyar river is the fourth longest river in Kerala and its water quality conservation is vital. ANN based study using feed forward back propagation neural network is used in this study. Sensitivity analysis is carried out on ANN model developed. Nine Artificial neural network models are presented in this paper. The performance of the model increased when T, EC, H, DO were given as input and it is the best model among other models. From sensitivity analysis, it can be concluded that EC has a significant role in BOD prediction. ANNs were able to capture the hidden relationships among the input variables and output variable and gave accurate results.This study involves the application of ANN based on monthly BOD modelling of Chaliyar river. Results from ANN models are presented. The results obtained in this study suggest that the ANN technique provide accurate results.

Keyword : - *Biochemical oxygen demand, Artificial Neural Networks, Chaliyar river, Water quality modelling*

---

1. INTRODUCTION

Biochemical Oxygen Demand (BOD) is an important parameter for usage conditions of surface waters. It is an approximate measure of the amount of biodegradable organic matter presented in a water sample. It is defined by the amount of oxygen required for the aerobic microorganisms present in the sample to oxidize the organic matter to a stable organic form. Excessive BOD results in low Dissolved Oxygen (DO) concentration and unsuitable living conditions for flora and fauna within. BOD determination requires 5 days, with data collection and evaluation occurring on the last day. Performing the test for BOD requires significant time and commitment for preparation and analysis. A test is used to measure the amount of oxygen consumed by these organisms during a specified period of time (usually 5 days at 20 ^oC). This is referred to as a BOD_5 measurement. ANN is an efficient tool for modelling and forecasting. ANN is a machine that is designed to model the way in which the brain performs a particular task or function of interest. It resembles the brain in two aspects that knowledge is acquired by the network from its environment through a learning process and inter neuron connection strengths, known as synaptic weights are used to store the acquired knowledge. The application of artificial neural network consists of two phases; training and testing. The procedure used to perform the learning process is called a learning algorithm, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective. The modification of synaptic weights provides the traditional method for the design of neural networks. However it is also possible for a neural network to modify its own network topology, which is motivated by the fact that neurons in the human brain can die and that new synaptic connections can grow. If training is successful, the network is able to apply what it had learned successfully to predict outputs corresponding to other input cases that have not been seen during training. The process of evaluating the performance of the network on a set of values that were not used for training is known as testing of the network. Areerachakul [5] compared predictive ability of ANFIS model and ANN models and estimated BOD of Saen Saep canal in Bangkok, Thailand. Abyaneh [8] predicted dissolved oxygen, with the help of water pH, runoff, water temperature and electrical conductivity data. Areerachakul et al. [4] used neural network for the prediction of dissolved oxygen in canals. Results of ANN models have been compared with the measured data on the basis of correlation coefficient. Zalaghi et al. [6] predicted the water quality parameters of Karoon River (Iran) by artificial intelligence-based models. Nine input water quality variables were employed for the models. Results showed that the computed values of DO, BOD and COD using both ANN and ANFIS models were in close agreement with their measured values in the river water.

The main objective of this study is to predict one of the most important parameters, BOD, with the help of some easily measured physical and chemical variables of the River Chaliyar using neural network. A further aim is to evaluate the results obtained and to apply sensitivity analysis to them, in order to determine which input variable(s) played a significant role in the prediction of output.

2 METHODOLOGY

For the purpose of BOD modelling artificial neural network approach is used. Monthly water quality data of Chaliyar River is employed for model development. Out of 19 years monthly data 70% data is used for training, 15% data for testing and 15% data for validation. 1997 to 2016 years data is collected for model development. Temperature, pH, Electrical conductivity, Hardness, COD, Dissolved oxygen, total coliform are the input parameters and BOD is the output parameter. Sensitivity analysis is done to find the relative significance of each input variable.

2.1 Artificial neural network

ANN is a computational model based on the structure and functions of biological neural networks. Artificial Neural Networks are massively parallel distributed data processing systems consisting of a large number of highly interconnected artificial neurons with performance characteristics resembling biological neurons of the human brain. The processes in developing the neural network model can be mainly divided into 2 phases, the training phase and forecasting phase. For this purpose a training algorithm is made use of. The type of algorithm used in the study is feed forward back propagation algorithm. Figure 1 shows the basic configuration of an Artificial Neural Network.

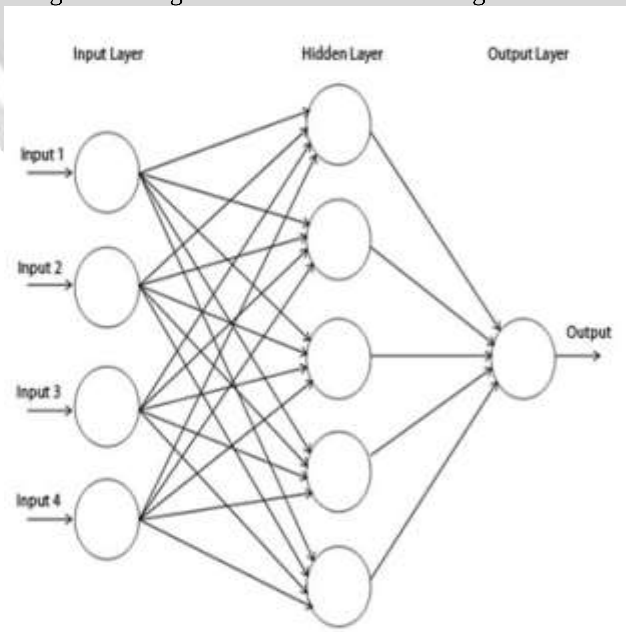

Fig -1: Basic configuration of an Artificial Neural Network

2.2 Feed forward back propagation training

Back propagation is the most commonly used supervised training algorithm training algorithm in the multilayer feed forward networks. The objective of a back propagation network is to find the weight that approximate target values of output with a selected accuracy. The network weights are modified by minimizing the error between a target and computed outputs. The error between the output of the neural network and the target outputs are computed at the end of each forward pass. If an error is higher than a selected value, the procedure continues with a backward pass, otherwise the training is stopped. The weights are updated continuously until minimum error is achieved. The least mean square error method is used to optimize the network weights.
For making use of the algorithm certain parameters has to be set, which include the training parameters such as epochs, learning rate, momentum constant, gradient and goal error. The epochs represent the maximum number of iterations possible in training. The learning rate lr is multiplied times the negative of the gradient to determine the changes to the weighs and biases. If the learning rate is set made too large, the algorithm becomes unstable. If the learning rate is set too small, the algorithm takes a long time to converge. The lr varies from 0 to1. Momentum allows a network to respond not only to local gradient, but also to recent trends in the error surface. Acting like a low pass filter, momentum allows the network to ignore small features in the error surface. Without momentum, a network may get stuck in a shallow local minimum. With momentum a network can slide through such a minimum. Momentum can be added to back propagation learning by making weight changes equal to sum of a fraction of the last weight change and the new changes equal to the sum of a fraction of the last weight change and the new change suggested by back propagation rule. The magnitude of the effect that last weight change is allowed to have is mediated by a momentum constant, mc, which can be any number between 0 and 1. When the momentum constant is 1, the new weight change is set equal the last weight change and the gradient is simply ignored. The gradient is computed by summing the gradients calculated at each training example, and the weights and biases are only updated after all training examples have been presented. The mean square error, epochs, learning rate, momentum, goal and transfer functions, the learning function and the parameters are initialized and the network is trained. The transfer function used for the hidden layer is The Tan-Sigmoid Transfer Function. The output layer uses a linear transfer function called purelin. The network was trained till any one of the stopping criteria mentioned below is met with. The criteria considered are i) maximum number of epochs is reached ii) performance has been assigned iii) momentum (mu) rate exceed mu-max. In the second phase (forecasting phase), the forecasting was done with the best neural network obtained in the first phase. 'MATLAB' software is to be used for developing the networks. The accuracy of the model goes on increasing as more data are available to neural network. For the neural network model, water quality parameters data is used as input and BOD is the output.

2.3 Data normalization

Due to the nature of sigmoid function used in the back propagation algorithm, it is necessary to normalize all the input and output values to fall in range [0,1]. Without normalization, large input values in to an ANN would require extremely small weighing factors which may cause floating point errors and training would require more time. The normalized values are obtained as:

$$Rn = (Ri-Rmin) / (Rmax- Rmin) \qquad \text{……………(1)}$$

Where Rn = normalized value
Ri = Real value
Rmax = Maximum value of all data
Rmin = Minimum value of all data

2.4 Performance evaluation

The data's are divided into two sets, for training and testing. The criterion for error assessment is mean square error. Root mean square error (RMSE) is adopted as an index to evaluate the performance of ANN models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}} \qquad \text{…………..(2)}$$

Where Xi is the output calculated, Yi is the output in testing set and N is the total no: of data.

3. ANALYSIS OF THE PROBLEM

Monthly water quality data of 19 years of Chaliyar river is used for model development. Data is normalized using equation (1) to fall in range [0,1]. 286 data sets are used for model development.

3.1 Data Used

Monthly water quality data (1997-2016) of Mundayathodu, Vadapuram and Kuravanpuzha stations are collected and used for model develpoment. Chaliyar river is the fourth longest river in Kerala and is 169 km in length. It is also known as Beypore River and originates in the Western Ghats range at Elambalari Hills of Wayanad. The Chaliyar river is shown in Figure 2.

Fig -2: Chaliyar river

4. RESULTS AND DISCUSSIONS

The training set is used for learning of the hypothesis. Then testing set is used for selection of the network model based on the performance parameter MSE and correlation coefficient R.

4.1 Multilayer perceptron network model development

The training results of ANN with all inputs are shown in Figure 3 and the scatter plot between observed BOD and ANN predicted BOD is shown in Figure 4. Results of sensitivity analysis are shown in Table 1.

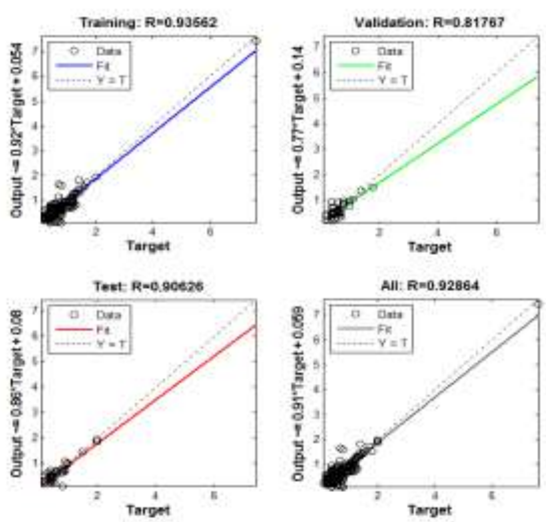

Fig -3: Training results of ANN with all inputs

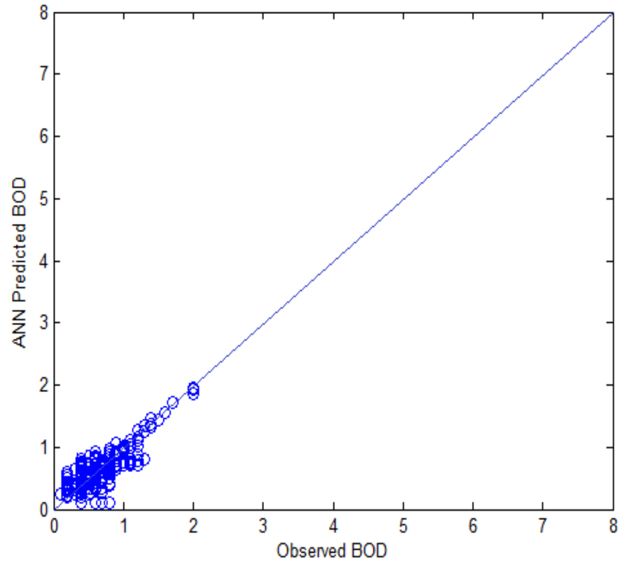

Fig -4 : Scatter plot between Observed BOD and ANN Predicted BOD with all inputs

Table -1: Sensitivity analysis

COMBINATION	RMSE	R
All	0.1014	0.92862
Eliminate T	0.1272	0.89352
Eliminate pH	0.1118	0.90224
Eliminate EC	0.1532	0.84526
Eliminate H	0.1385	0.88960
Eliminate COD	0.1044	0.92670
Eliminate DO	0.1456	0.86817
Eliminate TC	0.1232	0.91045

The training results of ANN with T, EC, H, DO as inputs is shown in Figure 5 and the scatter plot between observed BOD and ANN predicted BOD is shown in figure 6.

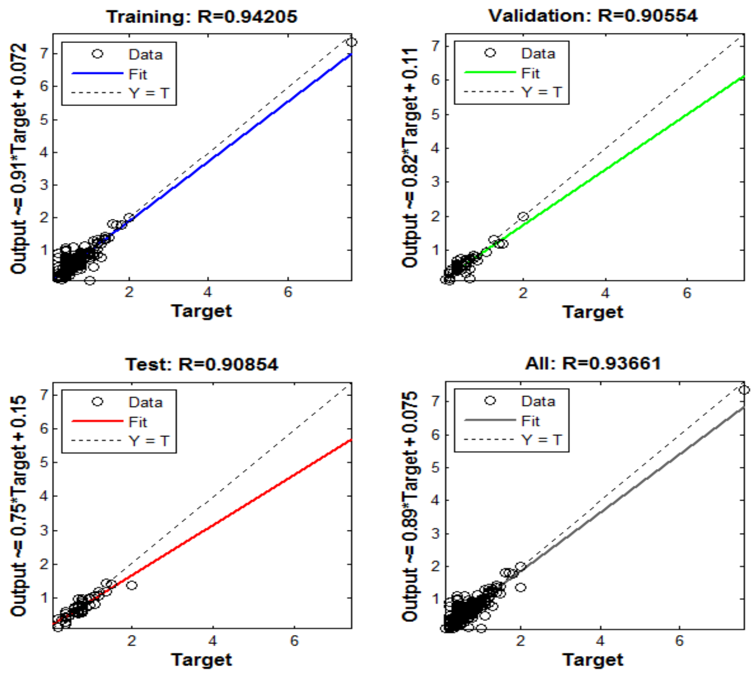

Fig -5: Training results of ANN with T, EC, H, DO as inputs

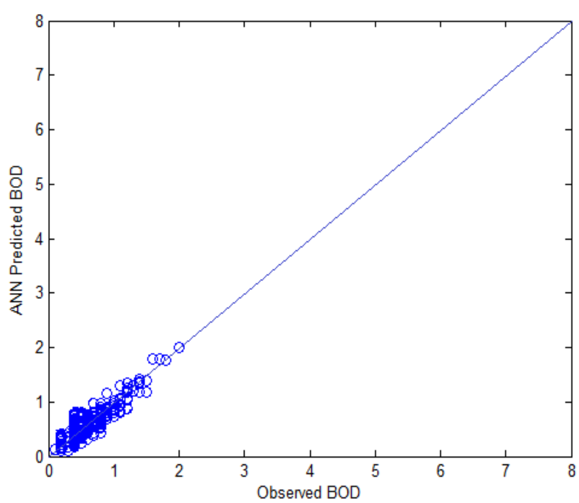

Fig -6 : Scatter plot between Observed BOD and ANN Predicted BOD with T, EC, H, DO as inputs

The correlation coefficient and RMSE values when all inputs are given are 0.92862 and 0.1014 respectively. After performing sensitivity analysis the significant input parameters are found as T, EC, H and DO. By giving these four parameters as input the correlation coefficient and RMSE values are 0.93661 and 0.09 respectively. The correlation coefficients obtained when Temperature is eliminated (R=0.89352) and pH is eliminated (R =0.90224) are shown in Figure 7 and Figure 8 respectively.

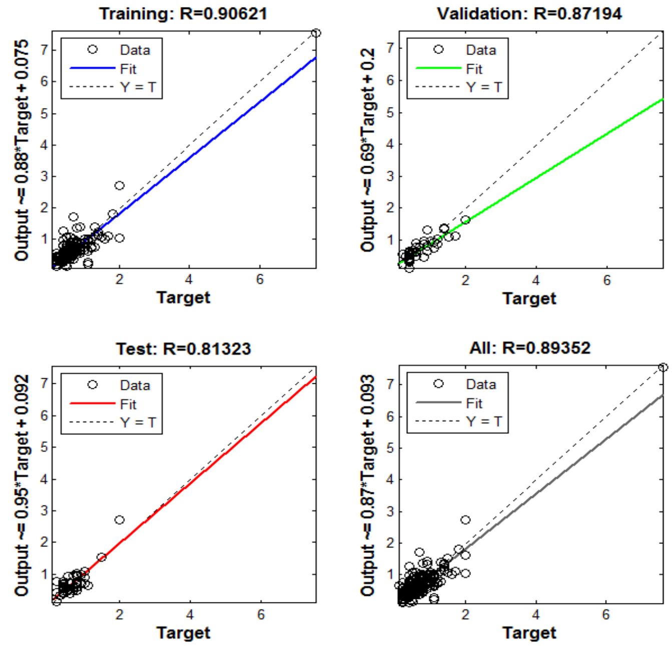

Fig -7 : Correlation coefficient when Temperature is eliminated

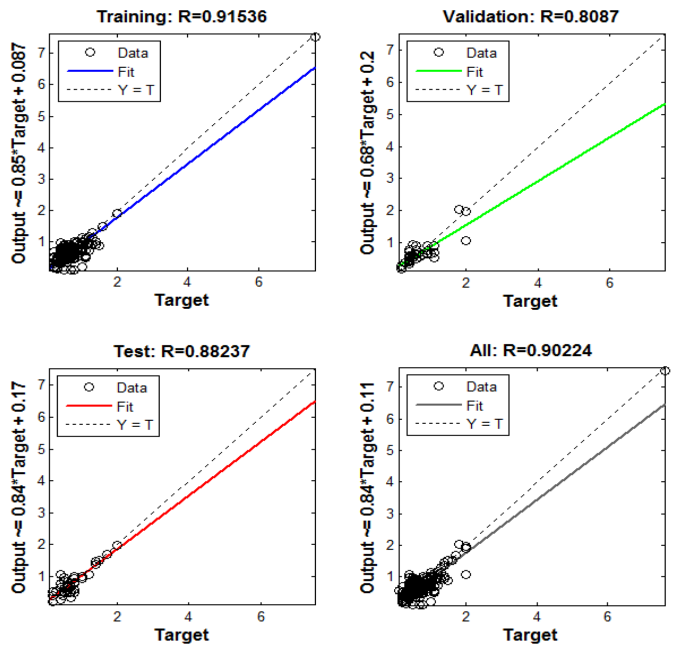

Fig -8 : Correlation coefficient when pH is eliminated

The correlation coefficient obtained when EC is eliminated (R=0.84526) is shown in Figure 9.

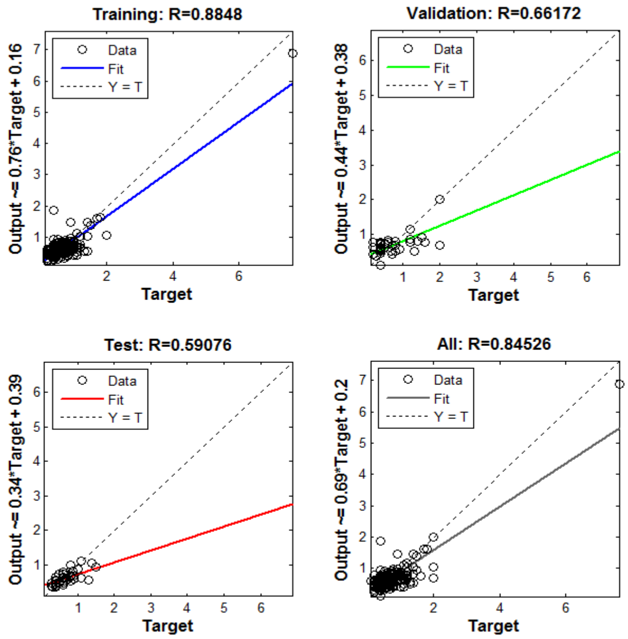

Fig -9 : Correlation coefficient when EC is eliminated

The correlation coefficient obtained when DO is eliminated (R=0.86817) is shown in Figure 10.

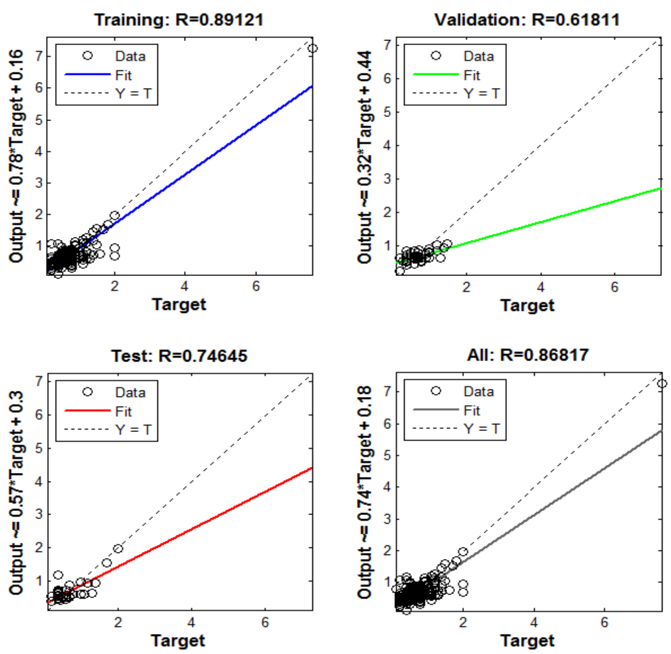

Fig -10 : Correlation coefficient when DO is eliminated

The correlation coefficient obtained when COD is eliminated (R=0.92672) is shown in Figure 11.

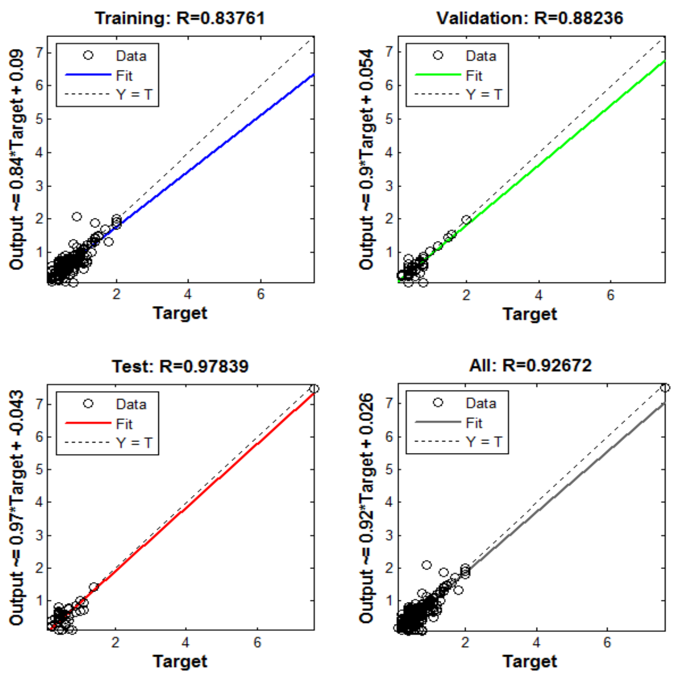

Fig -11 : Correlation coefficient when COD is eliminated

The correlation coefficient obtained when TC is eliminated (R=0.91045) is shown in Figure 12.

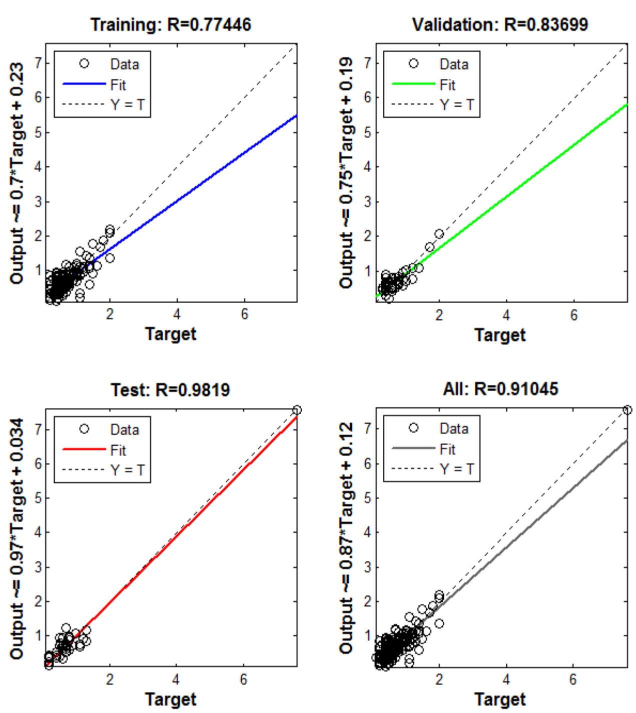

Fig -12: Correlation coefficient when TC is eliminated

5. CONCLUSIONS

Nine Artificial neural network models are presented in this paper. The performance of the model increased when T, EC, H, DO were given as input and it is the best model among other models. From sensitivity analysis, it can be concluded that EC has a significant role in BOD prediction and COD has least correlation with BOD. COD has lower correlation coefficient and high root mean square error whereas EC has high correlation coefficient and least root mean square error. ANNs were able to capture the hidden relationships among the input variables and output variable and gave accurate results. ANN has many advantages such as high accuracy, cost effective, fast monitoring, identify complex patterns, ability to capture nonlinear behaviours and are thus efficient tools for modelling BOD.

6. REFERENCES

[1]. A. Gamal El-Din, , D.W Smith and M. Gamal El-Din. Application of artificial neural networks in wastewater treatment, J. Environ. Eng Sci., pp.81-95, Jan 2004.
[2]. A. Jain and A. M. Kumar. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing 7: 585–592, 2007.
[3]. H. R. Maier and G. C. Dandy. Application of ANN for Reservoir Inflow Prediction and Operation. Journal of Water Resources Planning and Management. 125(5): 263-271, 1999.
[4]. S.Areerachakul and S.Sanguansintukul. A Comparison between the Multiple Linear Regression Model and Neural Networks for Biochemical Oxygen Demand Estimations, The Eight International Symposium on Natural Language Processing, Thailand, 2009.
[5]. S. Areerachakul. Comparison of ANFIS and ANN for estimation of biochemical oxygen demand parameter in surface water. Int. J. Chem. Biol. Eng. 6, 286–290, 2012.
[6]. S. Emamgholizadeh, , H. Kashi, , I. Marofpoor and E. Zalaghi. Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. International Journal of Environmental Science and Technology. Volume 11, Issue 3, pp 645–656, 2014.
[7]. S.H. Musavi, , and M.Golabi. Application of Artificial Neural Networks in the River Water Quality Modeling: Karoon River,Iran, Journal 0f Applied Sciences, Asian Network for Scientific Information, pp.2324-2328, 2008.
[8]. Z. Abyaneh. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters.J. Environ. Health Sci. Eng. 12 (40), 1–8, 2014.