International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V70I3P222
Volume 70 Issue 3, 195-200, March 2022

Original Article

Predicting BOD of Greywater using Artificial Neural Networks

Samir Sadik Shaikh[1], Rekha Shahapurkar[2]

[1]*Research Scholar, Computer Science and Engineering, Department, Oriental University, Indore, Madhya Pradesh, India*
[2]*Associate Professor, Computer Science and Engineering, Department, Oriental University, Indore, Madhya Pradesh, India*

[1]samirshaikh81@gmail.com, [2]rekhashahapurkar@gmail.com

Abstract - *The performance of an artificial neural network (ANN) model in evaluating the quality of water measures, such as BOD for greywater, is investigated in this article. Representative criteria for greywater quality include chemical oxygen demand (COD) and biochemical oxygen demand (BOD), along with indirect organic matter indicators. Mean square error (MSE) measurements were used to assess the ANN models' performance. The ANN model outperformed with MLR model in terms of performance, according to the results. MSE = 0.1299. Comparative indices of the improved ANN using temperature (T), pH, total suspended solids (TSS), chemical oxygen demand (COD), also total solids (TS) as input variables for BOD prediction were MSE = 0.1299. The ANN model was shown to be effective in predicting greywater BOD and COD levels. Furthermore, sensitive research findings revealed that the pH parameter has a greater impact on BOD and COD when compared to other factors.*

Keywords — *ANN, BOD, Greywater, Keras model, MSE.*

I. INTRODUCTION

Environmental preservation initiatives, particularly in the areas of water and air quality, have gotten a lot of press lately. Many academics have been tracking the steady collection of brief environmental quality data over the last few decades [1]. Predictions of environmental quality have gotten a lot of attention lately since it's crucial for agricultural and aquaculture operations' control, management, and planning [2]. One of the most important factors of environmental management to consider is the standard of the water, and it became the most popular limited resource in the long-term development of huge areas throughout the world. The water quality index is defined differently in different locations (WQI). Individual areas or nations' WQIs, such as Malaysia's National Water Quality Standard (Interim), WQI of British Columbia, The Water Quality Index for Canada (CWQI), in addition, the National Sanitation Foundation's WQI, are suggested by their respective authorities (NSF WQI). Each WQI is unique in terms of the factors and parameters that go into its computation.

A fundamental difficulty over water quality monitoring, according to Boyacioglu (2006) [3], seems to be the difficulty involved by means of examining the vast number of variables. Artificial intelligence (AI) techniques are most often used in water and environmental quality research as predictive tools [4]. Feedforward artificial neural networks (FANN) or Artificial neural networks (ANN), which remain inspired by and designed to replicate the human neurological system, are one of the most prominent AI approaches. When it comes to modelling and prediction of highly nonlinear systems, such as water class forecast scenarios [5], Artificial neural network (ANN) has demonstrated amazing effectiveness. In certain contexts, ANN modelling is integrated with certain other statistical analysis tools for improving the performance of the model, including Cho et al. (2011) [6], in which ANN would be integrated by principal component regression (PCR) for forecast groundwater arsenic substance, with the suggests it strategic however that PCR-ANN improved the predictive model. Other researchers have used this mix of techniques for ANN model prediction to improve the ANN model's accuracy [7-10].

The capacity of a neural network to train a system using previous data is its strongest asset. It has shown to be a more flexible, less assumption dependent, and adaptable approach in environmental-related disciplines such as air and water quality monitoring, lakes as well as dam modelling, hydrological forecasting, among others. The major benefit of neural network-based modelling techniques, according to Rabiatul and Zainal (2012) [11], implies that they will be simple to construct. Such features are extremely valuable to representing complicated processes that necessitate the development of comprehensive models.

Although the seas contain about 97 per cent of the total worldwide supply of water, just 3 per cent of it has been safe for direct consumption purposes. Kitchens basins, showering, laundering or washing machines, air - conditioning system outlets, dishwashers, and other similar appliances discharge greywater. Greywater generation ranges from 39 and 85 per cent in different nations, according to statistics. Non-potable water is treated and reused for toilet flushing, gardening, car cleaning, and floor washing, among other purposes. As a result, before utilizing greywater, it is critical to establish its purity. Biochemical oxygen demand (BOD) tests take five days to complete, but Chemical oxygen demand (COD) measurements only take a few hours. Because of certain issues with the identification as well as water quality testing parameters, for example, COD and BOD, the first and foremost objective in this research is to construct a prediction model based on an artificial neural network complicated water quality outcomes (BOD) based on temperature, pH, TS, TSS, and COD values.

Greywater reuse might eliminate the need for potable water for non-potable uses such as flushing toilets and gardening by correctly matching water quality to water requirements. For example, many houses have a line of tubes that takes in drinking water for numerous uses also another line of tubes that removes wastewater. Throughout the approach, completely water-using equipment, in addition, applications are using a separate type of water: extremely processed water that is safe to consume. In regions where wastewater treatment occurs, such water will be utilized once then reaches a drainage system where it is delivered and processed repeatedly [12]. Most contemporary wastewater systems dispose of treated discharge of wastewater further into the sea and perhaps other water bodies, negating possibility for reuse of such wastewater that has been cleaned. For some locations, wastewater that has been used might perhaps be dumped straight into the environment. For failing to match the quality of water to its intended usage, such a system wastes water, energy, and money.

A greywater solution, on the other hand, collects water that has been consumed for a particular function but has not been contaminated by significant levels of pollution, such as food waste and sewage. Such water is possibly utilized with a number of options. Freshwater which had previously been used in a bath, laundry basket, dishwasher, and bathroom faucet, for example, could be redirected for irrigation outside [13]. Therefore, herein scenario, the use of freshwater resources for outside irrigation is being decreased, as are the wastewater shower, washing machine, and sink all create streams. Because when systems are appropriately planned and executed, population health problems associated with the use of varied water quality may be resolved. The population health consequences of reusing water must be considered when Increasing the use of greywater solutions throughout regions wherever water recycle rules they're not rigorously controlled [14].

Semi-supervised regression (SSR) prototype with a co-training process created constructed on support vector machines (SVM), from remote sensing data which saves water quality variables. The nonlinear connection among water quality parameters and, therefore, the SPOT5 spectrum is defined by 2 (two) support vector regression (SVR) prototypes. Recognition of the 2 (two) SVR model from a semi-supervised co-training procedure. The method is used to identify 4(four) descriptive water superiority parameters of the river. The outcomes display the novel system takes improved presentation as compared to the arithmetic regression method. Over mixing 2(two) SVR prototypes also by nonclassified examples, a working technique when matching examples are inadequate is found. Integrating methods of machine learning with remote sensing, this one delivers a dynamic method going to remote sensing water quality retrieval. The system marks the routine of mutually classified and nonclassified examples. Also, 2 support vector regressors advance the regression correctness takes excessive benefits dissimilarity to out-of-date regression methods while lacking matching examples [15]. But, additional readings are required; for example: collecting additional facts also does further space and time analysis as well as validation; pool 2 dissimilar regressors and outline classifying confidence through new appropriate calculations to acquire improved learning results.

Various models have been developed to assess the Quality of water, noticeable clarity of the water, Dissolve Oxygen, Chemical Oxygen Demand, as well as Biochemical Oxygen Demand bounds using Artificial Neural Networks, Support Vector Machine, Deep Neural Networks, K-Nearest Neighbour, other approaches. Researchers have also proposed IOT-based solutions for water quality testing [16]. Though traditional ways of examining the nature of greywater have existed for several years, there was far less focus for greywater analysis applying AI processes. The motive of this study is to assess the water standards.

The uncertainty and volatility in the wastewater treatment system are caused by the complexity of usual circumstances, influent shock, also wastewater treatment skills. Such uncertainties cause changes in effluent water quality and operation expenses and an increase mostly in ecological risk for getting waterways. Artificial intelligence has evolved into a formidable tool for reducing the difficulties and issues associated with treating wastewater [17]. Perspectives on future perspectives research frontiers with the use of AI in wastewater treatment plants that target pollution removal, cost reduction, water reuse, and management issues in really complicated situations.

It is vital to determine the composition of water before using it. Biochemical oxygen demand (BOD) estimates to take 5 days to process, whereas Chemical oxygen demand (COD) estimates take just a few hours. Improved methods for measuring water quality is necessary, but a more organized process is also becoming increasingly common. Almost all of these representations need an extensive range of statistics that's not always easily accessible, creating it a costly plus time-consuming process due to various challenges within enlisting using estimating contained in water quality borders such BOD and COD [18].

The suitable construction of ANN models remained found with a series like modelling training and testing stages. ANN models produced accurate predictions. Mean absolute error, and root mean square error, as well as mean absolute percentage error for Chemical Oxygen Demand, Suspended Solids and Mixed Liquor Suspended Solids, suggest that the model is used efficient manner [19]. Overall, the results recommend the ANN demonstrating technique takes a high probability for adoption in simulation, exact performance prediction, as well as process managing of wastewater treatment facilities.

II. NEURAL NETWORK

Supervised learning is a machine learning technique that takes as input a feature vector and a target pattern to estimate the parameters. This model may be used to identify new patterns and assign them a target. Classification (e.g., classifying players based on their behaviours during a game) and regression are two examples of supervised learning applications (e.g. predicting household prices according to features)

In machine learning and data mining, classification is a critical challenge. It has been widely used in a variety of application scenarios. To create a classifier, a user must first gather a collection of training examples/instances that have been labelled with specified classifications. After that, a classification method is used to the training data to create a classifier, which is then used to assign the predetermined classes to test examples (for evaluation) or future instances (for application).

For training the ANNs also relating input data towards output data, a supervised learning approach was utilised. The training is designed to estimate, validate, and forecast the parameters using error function reduction. Its prediction proved good, with an R-squared, a sum of square error (SSE), a root-mean-square error (RMSE), as well as a mean squared error (MSE). The ANNs models were discovered to be a reliable technique towards forecasting wastewater treatment plant performance. Predictive techniques will be used to protect the environment as well as other technological advances [20].

The Artificial Neural Network model has been designed to assess as well as forecast specified water quality characteristics during any site within the domain of interest in a timely manner. These input parameters are data evaluated at different places. Salinity, temp, dissolved oxygen, as well as chlorophyll-alpha are the factors of interest [21]. ANN is an algorithm that can effectively classify water quality and produce accurate results [22].

A neural network's fundamental structure is made up of three major layers with interconnected artificial neurons: a layer of input, a hidden layer, and a layer of output. As processing components of a neural network, interconnected neurons bring the burden of the network. Fig. 1 illustrates the interconnection of neurons from various levels. These processing elements are comparable to inputs in terms of synaptic strength, activation output, and partial [23].

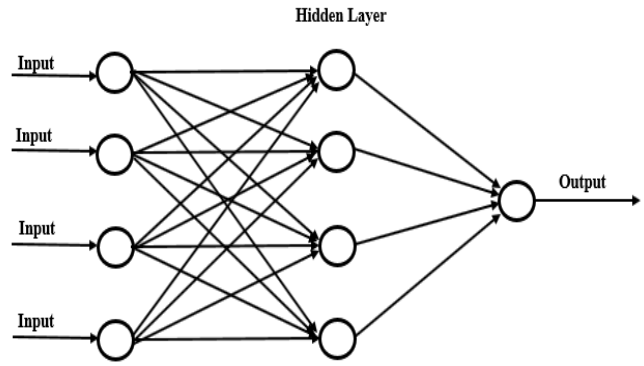


Fig. 1 Layered structure of ANN model

These neurons are relatively basic processing units that integrate bits of a larger issue. Every neuron generates one output by employing an activation function that takes into account the weighted sum of all its input data. The logistic sigmoid function is the most commonly used function of activation:

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

Here f(x) denotes a neuron's output, and x represents the weighted sum of a neuron's inputs. There are numerous varieties of neural networks, the very well of which is the multi-layer perceptron.

To train an Artificial Neural Network (ANN), the virtual network feeds data from the data set into the first layer's neurons. Such neurons, dubbed nodes, have two primary functions: adjusting the node weight through adaptive learning from incoming data and converting it into an impulse and transmitting that impulse to certain other nodes. Once the entire system goes through the training process, each neuron gets input from the neuron in the previous layer and sends its output to the neuron in the next layer in the network. As the network learns, the accuracy of the outputs generated by the network during the training phase improves. This method is repeated until a stopping condition (in this case, 200 epochs) is satisfied.

III. PROPOSED MODEL

Fig. 2 represents the ANN model for predicting BOD. The data set for this investigation was gathered over the course of 11 months using the kitchen sink (May 2020-March 2021), as shown in Table 1. Factors(variables) have been identified based on the observed values of various variables then their correlative analyses such as pH, water temperature (T), total suspended (TS), total suspended solids (TSS), and COD that impact water quality (BOD) was discovered and finally chosen for model development.

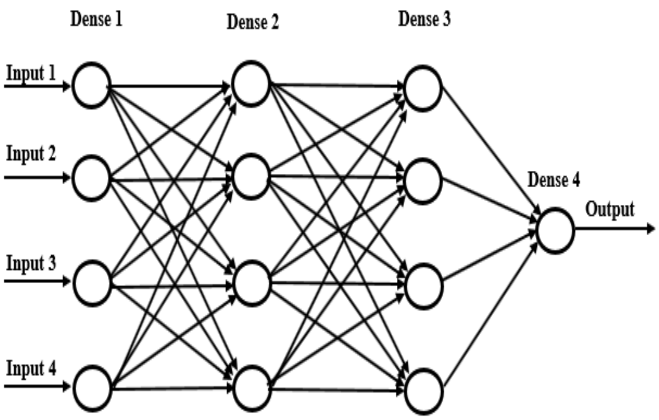


Fig. 2 ANN model for predicting BOD

Table 1. Cumulative water quality report

Sample No.	Temp (ºC)	pH	TS (mg/L)	TSS (mg/L)	COD(mg/L)	BOD(mg/L)
SS/R/01 /20	25.5	8.3	795	407	380	204
SS/R/02 /20	23.7	7.9	646	246	258	159
SS/R/03 /20	25.9	8.4	833	447	411	216
SS/R/04 /20	24.1	8	684	286	289	171
SS/R/05 /20	24.6	8.1	721	327	319	182
SS/R/06 /20	26.4	8.5	870	488	441	227
SS/R/07 /20	22.4	7.7	585	203	213	132
SS/R/08 /20	21.7	7.6	554	181	191	118
SS/R/09 /20	20.5	7.4	492	139	146	91
SS/R/10 /20	21.1	7.5	523	160	169	105
SS/R/11 /20	19.8	7.3	462	118	124	78
SS/R/12 /20	18.5	7.2	400	75	80	50
SS/R/13 /20	26.8	8.6	9.7	528	472	238
SS/R/14 /20	27.3	8.7	944	568	502	249

This needs a technique to quantify how well a probabilistic learning approach's prediction matches the observed statistics with the intention of evaluating its act arranged a particular statistic. To put it another way, need to measure how near the projected for a response value particular observation is to the real for a response value that observable. Almost majority of often-used metrics in regression remains the mean squared error (MSE), which is provided by

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(yi - f(\quad)ix)^2 \qquad (2)$$

in which $f(x_i)$ is f's forecast for the i^{th} opinion. Mean Square Error (MSE), if expected, will be minor and real answers stay extremely nearby and big if the predicted and true answers diverge significantly for some of the observations.

To put it another way, assume statistical learning training data to fit the statistical learning approach (x_1, y_1), (x_2, y_2), …, (x_n, y_n) then get the approximation f. Then calculate $f(x_1)$, $f(x_2)$, …, $f(x_n)$. The uncertainty they remain close towards y_1, y_2, …, y_n, at that time, the training MSE provided via (2) remains minimal. Furthermore, do not concern with maybe $f(x_i) \approx y_i$, rather, like to see if $f(x_0)$ is almost equivalent to y_0, wherever (x_0, y_0) is an unobserved test opinion that was aren't used to train statistical learning algorithms framework.

V. RESULTS

The experiment's model is made up of four tightly linked layers. The first layer is the input layer, which remains followed by two hidden levels and, eventually, an output layer. The code was developed in Python, using the Keras model tool and the sequential model imported. The sequential function aids in the production of hidden layers in a sequential manner, i.e. one after the other. Temperature, pH, TS, TSS, and COD data were obtained from laboratory readings over an 11-month period and fed into the first dense layer. "Relu" function was accustomed to normalising the data, particularly the '0' and '1' ranges. Because the data set available for training was minimal, the total amount of neurons in the hidden layer, as well as iterations (i.e. epochs), were greatly increased to achieve the lowest possible error the difference between the actual and anticipated values. 80% in terms of data was utilized for training the model, while the remaining 20% was used for testing. An error function with the mean squared (MSE) was used to compute the loss. Figure 3 depicts the output of the sequential model, which displays the projected value of BOD as well as the relative inaccuracy, i.e. loss caused by the model. Fig. 3 and Fig. 4 clearly shows that the loss is 0.1299 after 200 iterations, indicating that the model efficiency is approximately 88%.

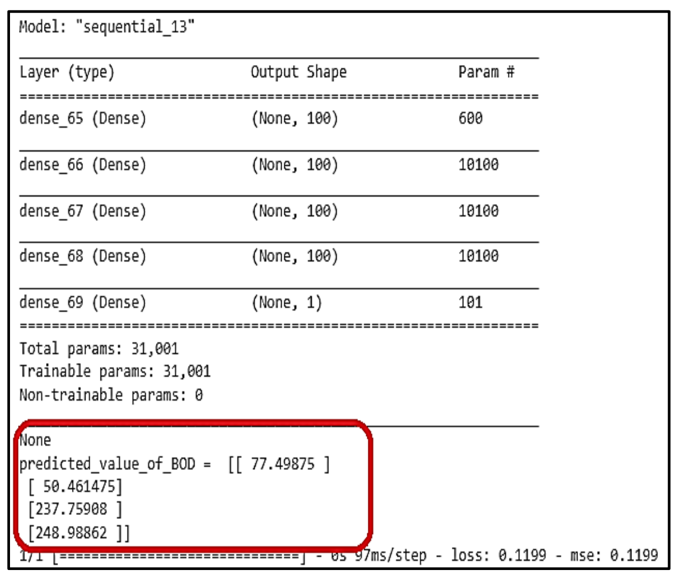
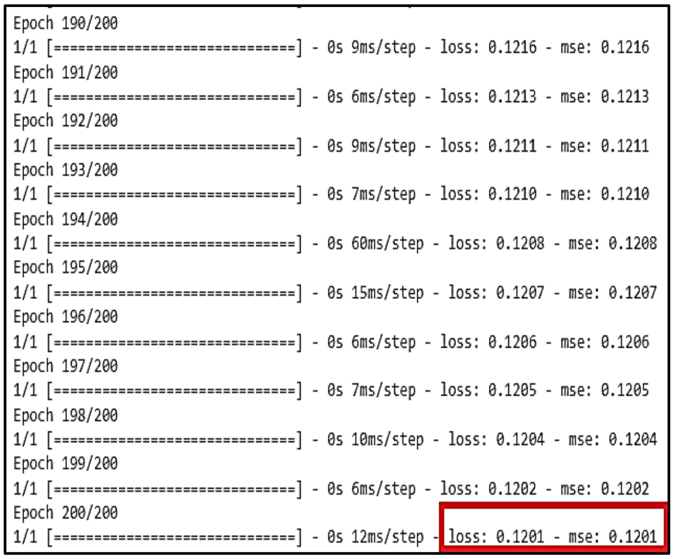


Fig. 3 Predicted value of BOD



```
Epoch 190/200
1/1 [==============================] - 0s 9ms/step - loss: 0.1216 - mse: 0.1216
Epoch 191/200
1/1 [==============================] - 0s 6ms/step - loss: 0.1213 - mse: 0.1213
Epoch 192/200
1/1 [==============================] - 0s 9ms/step - loss: 0.1211 - mse: 0.1211
Epoch 193/200
1/1 [==============================] - 0s 7ms/step - loss: 0.1210 - mse: 0.1210
Epoch 194/200
1/1 [==============================] - 0s 60ms/step - loss: 0.1208 - mse: 0.1208
Epoch 195/200
1/1 [==============================] - 0s 15ms/step - loss: 0.1207 - mse: 0.1207
Epoch 196/200
1/1 [==============================] - 0s 6ms/step - loss: 0.1206 - mse: 0.1206
Epoch 197/200
1/1 [==============================] - 0s 7ms/step - loss: 0.1205 - mse: 0.1205
Epoch 198/200
1/1 [==============================] - 0s 10ms/step - loss: 0.1204 - mse: 0.1204
Epoch 199/200
1/1 [==============================] - 0s 6ms/step - loss: 0.1202 - mse: 0.1202
Epoch 200/200
1/1 [==============================] - 0s 12ms/step - loss: 0.1201 - mse: 0.1201
```

Fig. 4 Calculation of Mean Square Error (MSE)

The model's performance was assessed using mean squared error statistics (MSE). Fig. 5 shows the graphical representation of how MSE is decreased with each epoch.

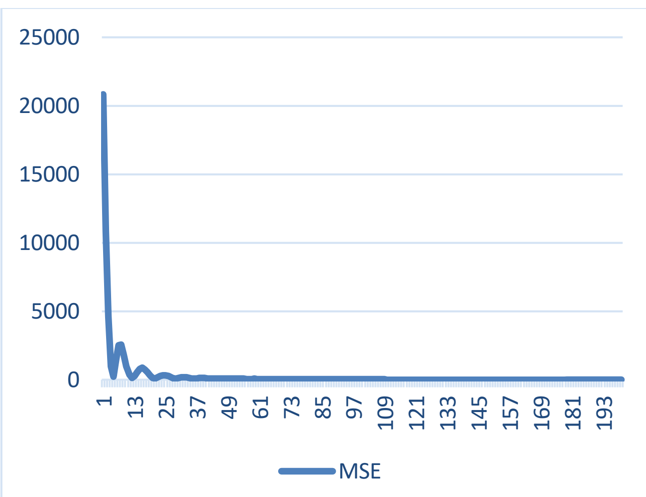


Fig. 5 Mean Square Error Vs. Epochs.

A comparison of predicted values of BOD and sample values from the dataset is shown in fig. 6. It shows a loss of 0.1201 means there is approximately 88% efficiency.

Table 2. Result of the testing set

No. of records	No. of correctly predicted	Accuracy percentage
200	176	88%

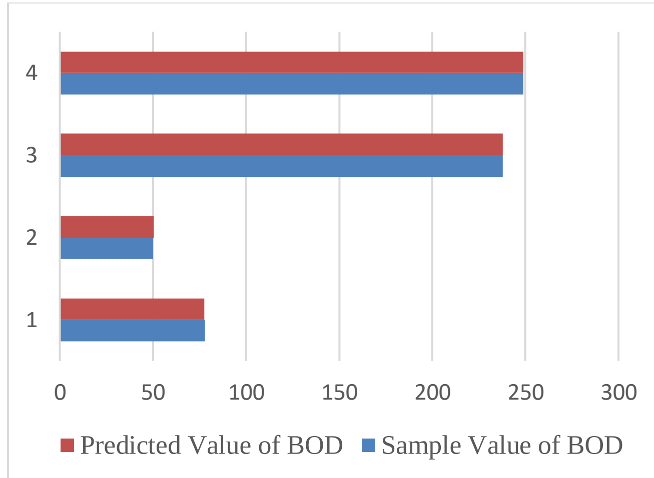


Fig. 6 Comparision between predicted values of BOD and sample values

VI. CONCLUSION

Greywater may be reused in several ways. Water that is consumed once in a bathroom sink, clothes washer, as well, and shower, for example, could be redirected outside for irrigation. As an example, suppose the need order to provide greywater towards external irrigation may be decreased, as well as wastewater streams generated by the shower, dishwasher, washing machine, and hand basin can be reused. Whenever the systems are appropriately planned and executed, any public health problems associated with the use of varied water quality may be addressed. As a result, it is critical to test the quality of water before utilizing that as well. BOD measurements require five days, but COD measurements only take a few hours. This article explores the usefulness of an artificial neural network (ANN) model in forecasting a water quality indicator, namely BOD used for greywater. The model's performance was assessed using mean squared error statistics (MSE). The findings showed that the ANN model with the smallest number of inputs such as pH, temperature (T), total suspended particles, total solids, and chemical oxygen demand could accurately predict BOD in greywater. The ANN model predicts efficiency of approximately 88% with a loss of 0.1201.

REFERENCES

[1] Antonopoulos V.Z, Papamichail D.M, and Mitsiou K.A, Statistical and Trend Analysis of Water Quality and Quantity Data for the Strymon River in Greece, Hydrology and Earth System Sciences, ASMA 2012, River Water Quality Monitoring. 5(4) (2012) 679–692.

[2] Dhalla P, et al., Quick and Reliable Estimation of BOD Load of Industrial Beverage Wastewater by Developing BOD Biosensor, Sensors and Actuators B: Chemical. 133 (2008) 478–483.

[3] Boyacioglu H, Surface Water Quality Assessment Using Factor Analysis, Water South Africa. 32(3) (2006) 389–39.

[4] Li M, and Hassan R, Urban Air Pollution Forecasting Using Artificial Intelligence-Based Tools, Air Pollution. (2006) 195–220

[5] Khuan L.Y, Hamzah N, and Jailani R, Prediction of Water Quality Index (WQI) Based on Artificial Neural Network (ANN), In 2002 Student Conference on Research and Development Proceedings. Shah Alam, Malaysia. (2002) 157–161

[6] Cho K.H, et al., Prediction of Contamination Potential of Groundwater Arsenic in Cambodia, Laos, and Thailand using Artificial Neural Network, Water Research. 45 (2011) 5535–5544.

[7] Khan R.A, et al., Using Principal Component Scores and Artificial Neural Networks in Predicting Water Quality Index, Chemometrics in Practical Applications. (2001) 271–288.

[8] Faruk O.D, A Hybrid Neural Network and ARIMA Model for Water Quality Time Series Prediction. Engineering Applications of Artificial Intelligence. 23(4) (2010) 586–594.

[9] Han H.G, Chen Q.L, and Qiao J.F, An Efficient Self-Organizing RBF Neural Network for Water Quality Prediction, Neural Networks. 24 (7) (2011) 717–725.

[10] Xu L, and Liu S, Study of Short-Term Water Quality Prediction Model Based on Wavelet Neural Network. Mathematical and Comput Modelling. 58(3–4) (2013) 807–813.

[11] Rabiatul M.N, and Zainal A, Optimum Numbers of a Single Network for Combination in Multiple Neural Networks Modelling Approach for Modelling Nonlinear System Optimum, IIUM Eng Journal. 12(6) (2012) 45–58.

[12] (2010). Al-Beiruti S, Jordan: Greywater Treatment and Use for Poverty Reduction in Jordan (English) - Multiple-Use Water Services Group. [Online]. Available: http://www.musgroup.net/page/553.

[13] Al-Hamaiedeh H.D, and Bino M, Effect of Treated Greywater Reuse in Irrigation on Soil and Plants, Desalination. 256(1-3) (2010) 115-119. doi: 10.1016/j.desal.2010.02.004.

[14] Al-Jayyousi O.R, Greywater Use: Islamic Perspectives, In: Greywater Use in the Middle East, Mcilwaine and Redwood (Eds), IDRC. (10) (2010).

[15] X. Wang, L. Ma, and X. Wang, ―Apply Semi-Supervised Support Vector Regression for Remote Sensing Water Quality Retrieving, Int. Geosci. Remote Sens. Symp. (2010) 2757–2760.

[16] Shaikh SS, Shahapurkar R, Machine Learning-Based Quality Prediction of Greywater: A Review Information and Communication Technology for Competitive Strategies (ICTCS 2020), Lecture Notes in Networks and Systems, Springer, Singapore. 190 (2021) 337-347.

[17] Lin Zhao, Tianjiao Dai, Zhi Qiao, Peizhe Sun, Jianye Hao, Yongkui Yang, Application of Artificial Intelligence to Wastewater Treatment: A Bibliometric Analysis and Systematic Review of Technology, Economy, Management, and Wastewater Reuse, Process Safety and Environmental Protection. 133 (2020) 169-182.

[18] Shaikh SS, Shahapurkar R, Predicting COD and BOD Parameters of Greywater Using Multivariate Linear Regression, Advances in Parallel Computing, IOS Press. 39 (2021) 228-238.

[19] Güçlü D, Dursun Ş, Artificial Neural Network Modelling of a Large-Scale Wastewater Treatment Plant Operation. Bioprocess Biosyst Eng, Springer, Singapore. (33) (2010) 1051–1058.

[20] Matheri AN, Ntuli F, Ngila JC, Seodigeng T, Zvinowanda C, Performance Prediction of Trace Metals and COD in Wastewater Treatment Using Artificial Neural Network. Computers & Chemical Engineering. 149 (2021) 107308.

[21] Palani S, Liong SY, Tkalich P. An ANN Application for Water Quality Forecasting. Mar Pollut Bull. 56(9) (2008) 1586-1597.

[22] S. Khadijah, H. Lokman, A. Mohd, S. Mohd, G. Rozaida, Water Quality Classification Using an Artificial Neural Network, Conference Series: Materials Science and Engineering, IOP Publishing. 601(1) (2019) 1-5.

[23] Singh T.N, Sinha S, Singh V.K, Prediction of Thermal Conductivity of Rock through Physico-Mechanical Properties, Building and Environment. 42 (2007) 146-155.