



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 3, March 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.569

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)



# Role of Machine Learning in Water Quality Assessment

Mahima Chaurasia<sup>1</sup>, Nidhi Prasad<sup>2</sup>, Sanjeev Kumar Srivastava<sup>3</sup>

Assistant Professor, Applied Sciences And Humanities, I.E.T. Dr. Rammanohar Lohia Avadh University, Ayodhya, UP, India<sup>1</sup>

Assistant Professor, CSE, I.E.T. Dr. Rammanohar Lohia Avadh University, Ayodhya, UP, India<sup>2</sup>

Assistant Professor, Departmental of Environmental Sciences, Dr. Rammanohar Lohia Avadh University, Ayodhya, UP, India<sup>3</sup>

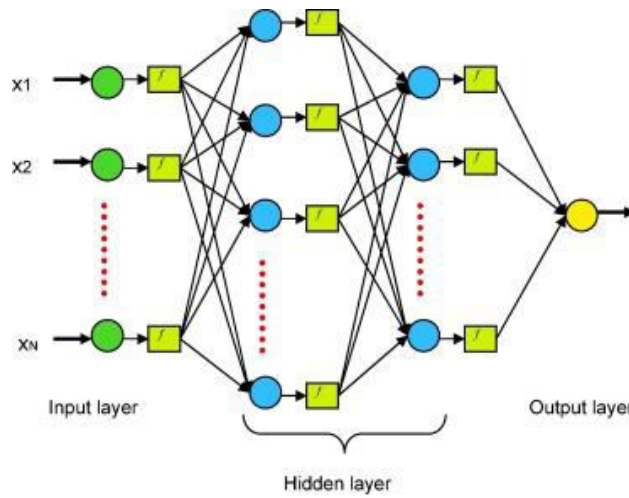
**ABSTRACT:** This study investigates the performance of artificial intelligence techniques including artificial neural network (ANN), group method of data handling (GMDH) and support vector machine (SVM) for predicting water quality components .. To develop the ANN and SVM, different types of transfer and kernel functions were tested, respectively. Reviewing the results of ANN and SVM indicated that both models have suitable performance for predicting water quality components. During the process of development of ANN and SVM, it was found that tansig and RBF as transfer and kernel functions have the best performance among the tested functions. Comparison of outcomes of GMDH model with other applied models shows that although this model has acceptable performance for predicting the components of water quality, its accuracy is slightly less than ANN and SVM. The evaluation of the accuracy of the applied models according to the error indexes declared that SVM was the most accurate model. Examining the results of the models showed that all of them had some over-estimation properties. By evaluating the results of the models based on the DDR index, it was found that the lowest DDR value was related to the performance of the SVM model.

**KEYWORDS:** water quality, machine learning, hydrometry, pollution, mechanisms, model

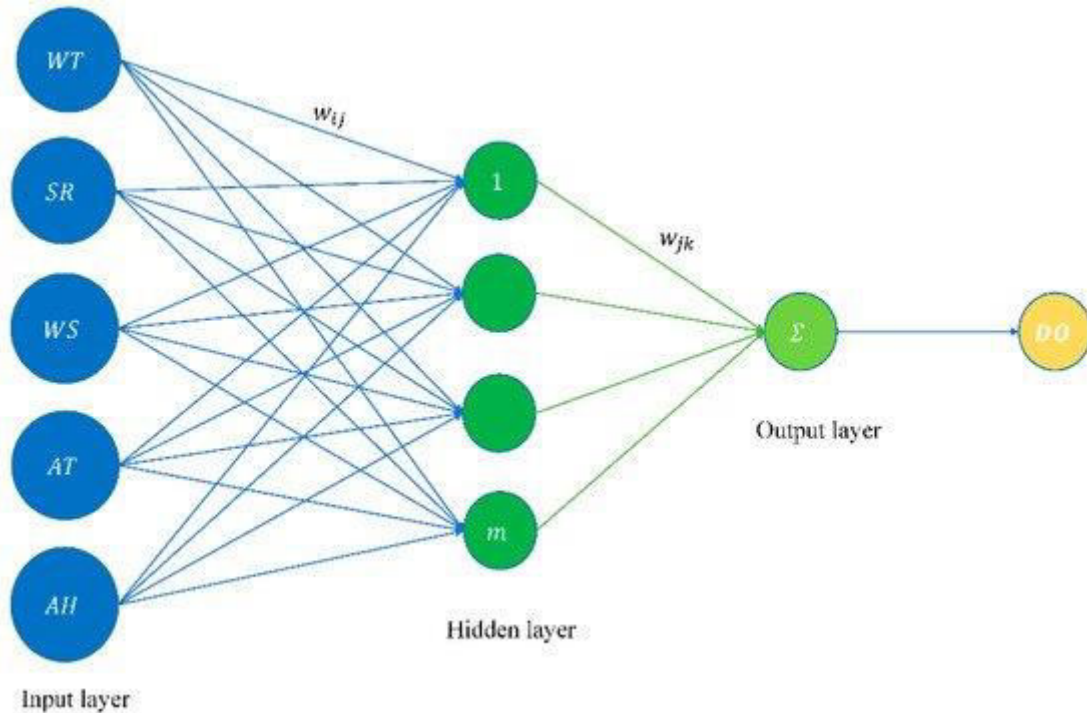
## I. INTRODUCTION

Water quality has a direct impact on public health and the environment. Water is used for various practices, such as drinking, agriculture, and industry. Recently, development of water sports and entertainment has greatly helped to attract tourists .Among various sources of water supply, due to easy access, rivers have been used more frequently for the development of human societies.[1,2]

Using other water resources such as groundwater and seawater sometimes assisted with problems. For example, using groundwater without suitable recharge will lead to land subsidence and using seawater is usually associated with pollution transmission . Therefore, the use of rivers has attracted attention. Several investigations related to rivers around the world have been conducted and a field of engineering named river engineering has been proposed. In river engineering, studies on morphological changes, sediment transport, water quality, and pollution transmission mechanisms are very important



.Flow structure, sediment transport and morphology of rivers are investigated in the hydraulics of rivers in river engineering .The study of water quality of rivers is a common theme in earth sciences. To evaluate the quality of rivers two approaches are considered, including measuring the water quality components and defining the mechanism of pollution transmission . Among water quality components, measuring the dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD), electrical conductivity (EC), pH, temperature, K, Na, Mg, etc. have been proposed .To this end, governments have constructed hydrometry stations along rivers that cross from urban areas, agro-industrial projects, industrial estates, and rivers that join dams' reservoirs . In hydrometry stations, the water quality components are measured and the stage-discharge relation is defined.[3,4]

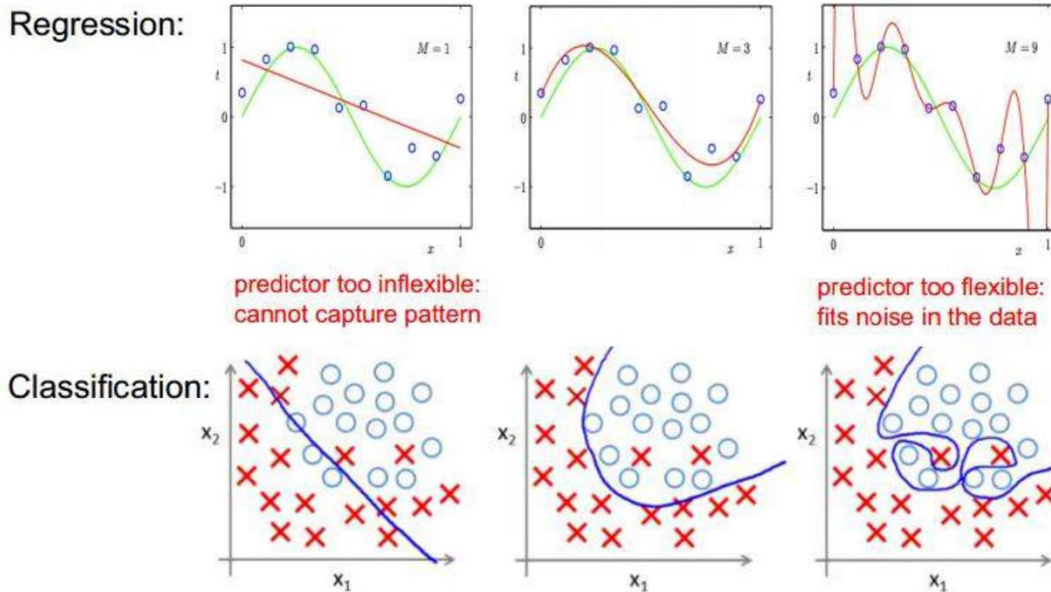


Obtained values from hydrometry stations contain basic information for feasibility studies and development of water conservation projects. Evaluation of water quality is a basic stage for development of agriculture projects in terms of determination of cropping pattern, type of irrigation system, and systems of water purification for industries . To investigate the mechanism of pollution transmission, in addition to field and laboratory experiments, advanced numerical methods such as computational hydraulic, image processing and GIS methods have been utilized .By reviewing the time history of water quality components, investigators have attempted to estimate future values. Nowadays, by advancing soft computing techniques in most areas of water and environmental engineering, researchers have attempted to accurately analyse time series of water quality components and their internal relation . In this regard,

a multilayer perceptron (MLP), radial basis network (RBF) and an adaptive neuro-fuzzy inference system (ANFIS) for water quality components.[5,6]

## II. DISCUSSION

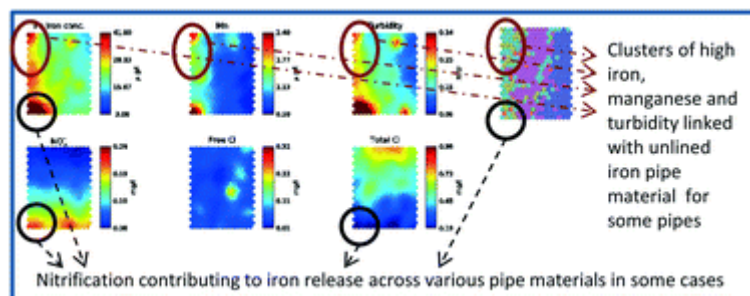
They stated that all applied models have suitable performance for prediction of water quality components; however, the MLP model was slightly more accurate. It was managed the water quality of a water supply system. They considered this an optimization problem and used modern optimization methods to solve it. It was introduced a new approach for water allocation.



They considered water quality as one of the main factors in their approach. It was developed a Probabilistic Support Vector Machines (PSVMs) model associated with GIS technique for planning the classification and distribution of surface and groundwater water. They stated that the use of these two methods would provide accurate information for feasibility studies of water conservation projects. We could utilize artificial neural networks for predicting the water quality components in several case studies. He stated that artificial intelligence techniques have suitable performance for modeling and predicting the internal relation between the water quality components and modeling their time series. Reviewing the literature shows that water quality assessment and prediction is an important factor for developing water conservation projects and, to this end, artificial intelligence techniques have been proposed. Hence, in this study the water quality components (were predicted using a support vector machine, article neural network and group method of data handling.[7,8]

## III. RESULTS

To develop an optimal model, an approach that was introduced by authors was considered. They stated that for developing the ANN, some steps should be considered to reduce the trial and error process.



They stated that for the initial design of ANN model, after dataset division, in the first step one hidden layer consisting of numbers of neurons equal to input features is considered. At this stage, the performance of different transfer functions is evaluated and the best ones are chosen. In the next step, the size of the network is modified to improve the



precision of the developed model. To this end, the numbers of neurons or number of hidden layers would increase. The last two stages of this approach are also applicable to the design of SVM. The main point relating to the design of the SVM is defining the kernel function. In this study, 80% of the dataset was used for training and the remaining 20% for testing. These two groups of datasets were used for developing the ANN, SVM and GMDH. For developing the GMDH model.

It is notable that the structure of ANN, after the trial and error process, consisted of two hidden layers where its first and second hidden layers included eight and three neurons, respectively. A comparison of three models in terms of predicting the Cl declared that the SVM model has the best performance. For predicting the Ec, all three models have suitable performance. For predicting the HCO<sub>3</sub>, the performances of SVM and ANN are close together and their precision is greater than the GMDH. This result was repeated for Mg. For predicting the Na, the SVM has the best performance. For estimation of SO<sub>4</sub>, the best performance was related SVM. For estimation of TDS, all three models have suitable performance and their accuracy is close together. For estimation of pH, the best performance is related to SVM. The accuracy of this model in the prediction of pH indicates that its accuracy has been reduced by a small amount. [8]

#### IV. CONCLUSIONS

In this paper, the performance of artificial intelligence techniques including GMDH, SVM and ANN were evaluated to predict the water quality components. To this end most dataset related well-known components, such as pH, SO<sub>4</sub>, Na, Ca, Cl, Mg, HCO<sub>3</sub> etc., were collected. Results indicated that the applied models have suitable performance for predicting water quality components, however, the best performance was related to the SVM. Reviewing the structure of SVM showed that the best accuracy was related to the RBD as kernel function. Results of ANN indicated that its accuracy is acceptable for practical purposes. The best performance of tested transfer function was related to tansig. The lowest accuracy of models was related to GMDH. The DDR index of results of applied models shows that all three models slightly over-estimate. Comparison of the performance of GMDH, SVM and ANN according to DDR shows that the data dispersion of SVM was less than the others. Although the accuracy of model GMDH is less than that of model SVM, their DDR indices are close together. Furthermore, comparison of the performance of applied models indicated that the outcomes of GMDH and SVM models were more reliable in comparison with ANN.[9,10]

#### REFERENCES

- [1] Ping Liu & Jin Wang & Arun Kumar Sangaiah & Yang Xie & Xinchun Yin, 2019. "Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment," Sustainability, MDPI, Open Access Journal, vol. 11(7), pages 1-14, April
- [2] K. Gopavanitha and S. Nagaraju —A low cost system for real time water quality monitoring and controlling using IoT, IEEE Conference paper, August 2017.
- [3] Manish Kumar Jha ; Rajni Kumari Sah ; M. S. Rashmitha ; Rupam Sinha ; B. Sujatha ; K. V. Suma —Smart Water Monitoring System for Real-Time Water Quality and Usage Monitoring, IEEE Conference paper, July 2018.
- [4] Asma Al Khaili, Aisha Al Mamari, Hoda Amer and Walid Ibrahim —An Affordable System for Remotely Monitoring Water Quality in Residential Water Tanks, IEEE paper, November 2018
- [5] Vaishnavi V. Daigavane, Dr. M.A Gaikwad. |Water Quality Monitoring System Based on IOT| 2017 Advances in Wireless and Mobile Communications, Nov 2017 ISSN 0973-6972
- [6] Neural Network Basics using Matlab, [http://staff.ttu.ee/~jmajak/Neural\\_networks\\_basics\\_s\\_.pdf](http://staff.ttu.ee/~jmajak/Neural_networks_basics_s_.pdf)
- [7] Thair S.K. , Abdul Hameed M. J. , and Ayad S. M., Prediction of water quality of Euphrates river by using artificial neural network model ,International Research Journal of Natural Sciences Vol.2,No.3,pp.25-38, September 2014
- [8] Nikhil Kumar Koditala ; Purnendu Shekar Pandey, | Water Quality Monitoring System Using IoT and Machine Learning, IEEE paper, August 2018
- [9] Najah, A. El-Shafie, O. A. Karim, Amr H. El-Shafie —Application of artificial neural networks for water quality prediction, May 2013, Volume 22, pp 187– 201.
- [10] —Resources, Resources — Internet of Things World Forum, [www.iotwf.com/resources/72](http://www.iotwf.com/resources/72)



Impact Factor:  
7.569



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details