SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V10I5P117

Volume 10 Issue 5, 178-186, May 2023

Original Article

Multiparameter Prediction of Water Quality using Edge Intelligence

Aleefia A. Khurshid[1], Anushree Mrugank Minase[2], Ashlesha Bonkinpelliwar[3]

[1, 2, 3]*Department of Electronics Engineering, Shri Ramdeobaba College of Engineering and Management, Maharashtra, India.*

[1]*Corresponding Author : khurshidaa@rknec.edu*

Received: 21 March 2023 Revised: 23 April 2023 Accepted: 19 May 2023 Published: 31 May 2023

Abstract *- The ecosystem and public health are significantly threatened by water contamination. Contaminated water can have harmful consequences; therefore, monitoring water quality has become a pressing environmental concern. This work proposes a multiparameter water pollutant prediction model to ensure a green environment. The contribution is towards the use of fewer inputs to predict multiple parameters. Here Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and nitrate levels in the water are predicted using the Extra Trees algorithm. A modular network is implemented to ensure enhanced efficiency and independent training, while Principal Component Analysis (PCA) aids in reducing the data load and improving the response time. The learning algorithm uses minimal sensed parameters, such as temperature, pH, and DO conductivity and is cost-effective in computation and simple to integrate into an IoT hardware system, thus reducing the need for expensive online sensors and environment dependency. The results indicate enhanced efficiency with a maximum error of 10%. Nash-Sutcliffe Efficiency (NSE) for the proposed model is 0.9. The pruned model is also implemented on Raspbian OS to integrate the developed soft sensor in an IoT environment, and the prediction timing is 5.48 seconds with a CPU clock speed of 1.2GHz*

Keywords *- Water quality prediction, Machine learning, Extra trees algorithm, Soft sensors, Principal component analysis.*

1. Introduction

For all living things, water is an important natural resource. Its quality must be monitored closely to ensure its safety for consumption and use. Several factors, including organic and inorganic contaminants, determine water quality. Utilizing the Biochemical Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) characteristics, it is possible to determine the presence of organic contaminants in water. These parameters provide information on the quantum of organic pollutants in water and their potential impact on the environment and human health. Industrial waste, agricultural activities, and sewage discharge often contaminate rivers, lakes, and groundwater, accumulating pollutants such as nitrates, ammonia, and chloride. Because of its detrimental impact on human health and the ecosystem, nitrate contamination in water is one of these. Nitrate is a form of nitrogen found in many fertilizers and can enter water sources through agricultural runoff or wastewater. Nitrate can also promote the growth of harmful algae blooms in water bodies, depleting oxygen levels and harming aquatic life. When nitrate levels in drinking water exceed safe limits, it can lead to fatal health issues. Therefore, monitoring nitrate levels in water sources to ensure their safety for human consumption is crucial.

The traditional methods for BOD, COD, and nitrate detection involve laboratory analysis of water samples, which is time-consuming. Alternatively, monitoring systems can be designed using in-situ sensor technology. Some of these systems feature pricey instrumentation but a high sensitivity with a broad detection range. Others, where the suggested detection approach uses a portable sensor-based sensing device, have a restricted sensitivity at a lower cost.

To address this issue, machine learning techniques have gained attention in recent years for predicting pollutant levels in the water to develop a low-cost soft sensor. Regression methods and a limited set of remotely observed factors, including pH, DO, temperature, EC, and ORP, are used to predict the secondary parameters BOD, COD, and nitrate level, and different machine learning models have been applied.

This study aims to utilize edge intelligence to develop a multiparameter prediction soft sensor, reducing the use of costly online sensors. This study proposes using the extra trees algorithm to predict values of the water's BOD, COD, and nitrate levels. This combination of techniques for learning different tasks is carried out through modular networks. The proposed model has demonstrated effectiveness in predicting mentioned pollutants in water and utilizes it as an

implementation of the Multilayer Network (MLN). To improve the efficiency of the prediction model and lessen the load of high-dimensional data, Principal Component Analysis (PCA) is used.

2. Related Work

The relevant literature is reviewed in the context of various machine learning techniques, COD, BOD, and nitrate prediction. These algorithms can successfully handle challenging non-linear issues because they learn from the dataset rather than relying on pre-established rules based on prior knowledge [1]. To help farmers make better decisions, machine learning (ML) algorithms have been proposed to simulate the changing patterns of water-quality indicators [2]. Data from frequently measured variables like pH, DO, and ammonia was examined using multilayer networks, random forests, and regression models.

Gradient-boosted decision trees are suggested by Zifei Wang et al. [3] as a method of predicting COD load in wastewater. This algorithm is memory and time-consuming. To estimate COD, the authors in [4] created an ANN model with remarkable accuracy utilizing eight input parameters. Regression techniques are used in this study [5] to calculate findings, and six essential remotely sensed factors are used to forecast COD.

To build the model, the weighted regression model is utilized. By using ten water variables as input to the model, according to the results of machine learning algorithms, the K-Nearest Neighbor is suitable for the COD measurement concerning the reaction time and error metrics for the water samples taken from the Ganga River, according to Arunima Pattanayak et al. [6].

In another study [7], the authors predict the suitability of surface water quality for drinking in the Shivganga River basin using ANN and MLR modelling approaches. The study finds that all parameters except total dissolved solids, electrical conductivity (EC), (TDS), TH, Mg and Ca, were within an acceptable range. The WQI method was used to evaluate its suitability for drinking, and only one sample showed poor quality due to the input of domestic and agricultural waste. ANN model provided good predictions of WQI for both seasons.

In addition, the authors of [8] present the use of an ANN model to forecast the level of nitrate in groundwater in Turkey's Harran Plain. The study uses easily quantifiable variables like temperature, electrical conductivity, groundwater level, and pH as input factors. In order to optimize the model, the back-propagation algorithm and optimal neuron numbers were chosen. The model closely tracks the experimental data, with an R-value of 0.93, making it a cost-effective and more accessible way to manage water resources.

Furthermore, the study [9, 10] compares the effectiveness of three machine learning models, including ANN, SVM, and XGB, for groundwater quality assessment. The study assesses the models' performance in classifying the amounts of nitrate and pesticide pollution for 303 wells in 12 US states. The ANN model showed the best accuracy for nitrate readings with an R2 score of 0.53. The article [11] also reviews fuzzy logic, ANN, and SVM for predicting nitrate contamination in water. It assesses the models' performance, identifies future research directions, and suggests combining models with evolutionary algorithms to improve accuracy.

The article [12, 13] also describes the investigation of nitrate pollution in groundwater using multivariate statistical techniques, including factor analysis and sparse principal component analysis. The study collects 156 groundwater samples from various sources and forecasts the nitrate contents at untested areas using empirical Bayesian modelling. According to the study, 5% of the samples had nitrate levels higher than the World Health Organization set permitted threshold. The study suggests that multivariate statistical tools and geostatistical models can help investigate groundwater geochemistry and improve nitrate pollution management in similar regions. The review [14,15,16,17] discusses the performance of ML in analyzing various types of water. Moreover, the document emphasizes the benefits of widely adopted ML algorithms for achieving accurate and efficient water quality evaluation.

The researchers in [18] discuss using machine learning algorithms to quantify nitrate and phosphorus in water bodies, aiming to tackle the issue of high levels of water pollution resulting from industrialization and population growth. For five distinct streams with differing land utilization practices, the authors assess the performance of various ML algorithms, such as LR, KNN, RT, ANN, RF, SVM, GPR, and BO. The findings suggest that the proposed methodology and results could assist users in predicting nitrate and phosphorus concentrations and developing efficient water pollution control plans.

In order to assess the underground water's susceptibility to nitrate contamination, Venkatramanan Senapathi et al. [19] compare machine learning algorithms, concentrating on correcting the drawbacks of the DRASTIC model, which is frequently utilized. The authors examine three ML models, namely RBNN, SVR, and RF, and propose a modified DRASTIC model. The experiments indicate that implementing ML models enhances groundwater vulnerability analysis's accuracy and reliability.

The research in [20] uses easily observable and affordable data from the Kopaidian Plain to create an ANN model to assess nitrate groundwater contamination. Based on geography

and land use data, the study shows that ANNs offer a potential approach for forecasting pollution levels, producing a lower RMSE value.

In the review article [21], researchers systematically analyze the various models for estimating nitrate contamination resulting from farming activities. The study analyzes the models used, input requirements, evaluation metrics, and challenges encountered during modelling. The paper suggests that computer models represent valuable tools for estimating nitrate contamination, and the most appropriate model suitable for specific research objectives can be chosen.

The study in [22] analyses the efficacy of various models for estimating nitrate pollution in Iran's quasi-intensive farmland region, employing spatial and quasi-spatial regression techniques. The results suggest that spatial regression methods are effective in water pollution control decision-making, and the approach employed in this study could assist in delineating groundwater pollution in the study area.

Ouedraogo et al. [23] created a geographic information system database with 13 geographical variables relating to land use, soil kind, terrain, climate science, area, and nitrogen fertilizer use rate to anticipate underground nitrate contamination. The study indicated that Random Forest Regression (RFR) has high predictive capability than multiple linear regression (MLR) methods. The authors in [24, 25] investigate the predictive modelling power for nitrate pollution using RF regression. The study compares the RF approach to logistical regression (LR) using various efficiency criteria to ensure their generalization ability. The findings demonstrate RF's capacity to create precise models with potent prediction powers.

According to the authors of [26], who analyzed the IBK technique's efficacy for BOD prediction, the root mean square error is 0.1994, with an edge response time of just 0.15 s. For BOD prediction, the authors of [27] describe a modular neural network (MNN) with mean root mean square error for the testing as 0.0079.

Principal component analysis, also known as PCA, lessens the complexity of high-dimensional information while keeping patterns and trends intact [28-30]. The studies analyzed in this review demonstrate that machine learning models can effectively predict BOD, COD, and nitrate concentrations in water sources. They provide insights to draft efficient plans to tackle water pollution. Using artificial neural networks (ANNs) for modeling contaminants in water sources has shown promising results. While several independent models have been developed to estimate mentioned pollutants, process-based models are predominantly used. These models require high-quality data, more than six input parameters, and computational resources, which can be challenging in some areas.

Therefore, there is a need for more straightforward and cost-effective models that can be used in data-scarce regions. The necessity for a multiparameter prediction model suited for integration in an IOT context that can give an adequate level of accuracy utilizing edge intelligence while needing the fewest possible parameters along with minimum edge reaction time is another element that is underlined in this analysis.

3. Materials and Methods

3.1. Dataset

The "Namami Ganga" project's live data streaming compiles the database (30,000 samples), including crucial variables for six months to incorporate variations due to weather changes. Table 1 illustrates an example of this dataset, which contains dependent parameters, such as pH, dissolved oxygen, BOD, COD, conductivity, nitrate, temperature, Station Type Drain, and River.

3.2. Data Preprocessing

Preprocessing is done to increase data quality. Data is normalized using the linear scaling method, and after cleaning, around 23,000 data samples were collected.

3.3. Data Partition

The dataset is divided into testing and training subsets, with 80% allocated for training and 20% for testing. The different trees algorithm is then applied to generate a COD, BOD, and Nitrate prediction model.

3.4. Principal Component Analysis

In order to obtain a more comprehensive understanding of the data, PCA is used. PCA is a statistical tool that can simplify high-dimensional data by reducing it into fewer dimensions that capture the trends and patterns in the data. It identifies the most critical variables that serve as feature summaries.

In this study, two prediction models with and without PCA are experimented with the Nitrate parameter. Both models use the extra trees algorithm, and PCA is performed on five components. The accuracy of the models is assessed through testing, and scatter plots are generated to visualize the relationship between the predicted and actual values of COD, BOD, and Nitrate from the test dataset samples.

3.5. Machine Learning Algorithms

According to the Examined Research, the RF model can reach satisfactory results compared to multivariate statistics or other machine learning techniques. In addition to the well-known machine learning structures in this study, the different

tree framework is used for the first time to model and predict multiple pollutants in river water.

3.6. Proposed Method

The extra trees approach predicts a low-cost model's BOD, nitrate, and COD values. The quality of the data used for simulation and testing can impact how effective learning algorithms are. Using the support of training data and randomized feature subsets, Extra Tree is an enhanced version of Random Forest that creates several decision trees. The fact that Extra Trees lessen variation is another benefit. Because nodes in the decision trees are divided randomly, the approach is not significantly influenced by particular dataset features or trends.

The algorithm's effectiveness is evaluated using performance metrics. To predict multiple parameters, a modular neural network approach is utilized. The intermediary processes the outputs of each module to create the network's overall output, resulting in reliable and independent training.

Four water parameters-pH, DO, EC, and temperature-form the model's input based on the correlation map's depiction of the interactions between various parameters and the relative weights of those interactions (Figure 1). The learning model is simulated using the variables with a correlation coefficient greater than 0.25. The categorical data also serves as the input to the trained model, streamlining data analysis.

Figure 2 illustrates the steps involved in developing a modular network. The flowchart depicts a procedure that starts with inputting the different parameters acquired from the Ganga River dataset. DO, temp, PH, and EC include categorical parameters. The dataset forms an input to the extra trees prediction model to predict the BOD and COD levels. These predicted values and other parameters are input for the Nitrate extra trees prediction model. This final model predicts the levels of nitrate in the water based on the combined dataset. Therefore, the process uses multiple prediction models to forecast the Ganga River's water quality parameters, Nitrate, COD, and BOD levels.

The BOD and COD prediction models capture the correlations between the four input variables in the current scenario. Nitrate levels in the water are affected due to changes in BOD values, and this dependency is captured using an independent network where predicted BOD and COD form an additional input to the extra trees model. The amalgamation of predictions improves the accuracy of the overall Nitrate prediction. Preprocessing the data through Principal component analysis condenses the data into fewer dimensions that serve as feature summaries.

4. Results

4.1. Performance Metrics of Multiparameter Prediction Model

Various performance metrics are computed, as indicated in Table 2, to evaluate the effectiveness and accuracy of a predictive model. These metrics provide a quantitative assessment of how well the model predicts the outcome of interest. NSE, or Nash-Sutcliffe Efficiency, gauges the model's capacity for prediction. The time taken by each model run on an Intel(R) Core(TM) i5 processor operating at 1.80 GHz speed is also given in Table 2.

Table 3 compares current results from the most recent literature to the outcomes of the suggested approach. As the table shows, the proposed method delivers better performance measures using fewer input variables. Moreover, previous studies were based on a prediction of single parameters, whereas this work is directed toward designing, simulating and implementing a multiparameter prediction system.

4.2. Performance Analysis of Extra Trees Algorithm

Performance metrics provide a quantitative measure of the accuracy of a forecasting model. In contrast, the plot of true vs predicted values visually represents the model's performance. Additionally, the plot can help identify outliers or other unusual data points affecting the model's performance. Figures 3 and 4 show the scatter plots for both models' accurate vs predicted nitrate values.

A comparison of both plots concludes that the model inclusive of PCA generates better and more accurate results than the other model. Figures 5 and 6 show the scatter plots for actual vs predicted values of BOD and COD using the Extra Trees Algorithm. The prediction error ranges from 5 to 10% of these samples. The training time for the model runs on an M1 processor operating at 3.2 Ghz is 722 ns, and the prediction time equals 48.7ns.

The pruned model is also implemented on Rasbian OS to integrate the developed soft sensor in an IOT environment, and the prediction timing is 5.58 seconds with a CPU clock speed of 1.2GHz. From the test results, it can be concluded that the model is best-suitable for predicting multiple parameters in water bodies, specifically storage tanks, rivers, and lakes. The model can deliver high-performance processing capabilities to the edge, where sensors and IOT devices are situated, which is its most significant benefit. The results of the suggested method are contrasted with those of the most recent research, and it is observed that the proposed technique provides acceptable performance while requiring fewer input variables.

Table 1. Sample data

DO	EC	pH	Temp.	Nitrate	BOD	COD
1.96	365	8.7	31	0.84	1.16	12.86
9.06	727	8.3	31.3	0.56	4.64	25.2
7.43	182	7.84	28.5	0.01	3.345	17.99
6.76	198	7.6	29.6	1.28	2.123	14.152
6.34	464	7.86	30.1	2.55	3.21	21.57
6.84	203	8.09	29.4	1.32	2.21	13.7
7.26	192	7.9	29.5	1.42	1.68	12.45
7.48	203	8.2	29.6	1.57	3.06	19.87
6.69	315	7.84	30	2.18	2.23	13.35
0.6	910	8.11	29	2.39	20.469	58.469l

Table 2. Evaluation metrics of proposed method

Multi-Parameter Prediction Model	MAE	MSE	RMSE	NSE	Training Time (in sec)	Prediction Time (in sec)
BOD and COD Prediction with Extra Trees	0.52	2.69	1.64	BOD: 0.994 COD: 0.990	0.574	0.053
Nitrate Prediction with PCA for 5 Components	0.1	0.08	0.28	0.9215	0.618	0.093

Table 3. Comparison table

Ref. No.	Predicted Parameter	No. of Inputs	Algorithm Used	Performance Metric
[4]	COD	8	ANN	RMSE=0.0108
[5]	COD	6	ANN	MSE=0.0449
[23]	BOD	6	IBK ML	RMSE=0.199
[24]	BOD	9	MNN	RMSE=0.0079
[9]	Nitrate	18	ANN	RMSE=3.91
[17]	Nitrate	5	ANN	RMSE=15.95
[21]	Nitrate	4	RF	MSE=0.081
-	BOD & COD	4	Proposed method using Extra Trees	RMSE=1.64
-	Nitrate	6	Proposed method using Extra Trees with PCA	RMSE=0.28

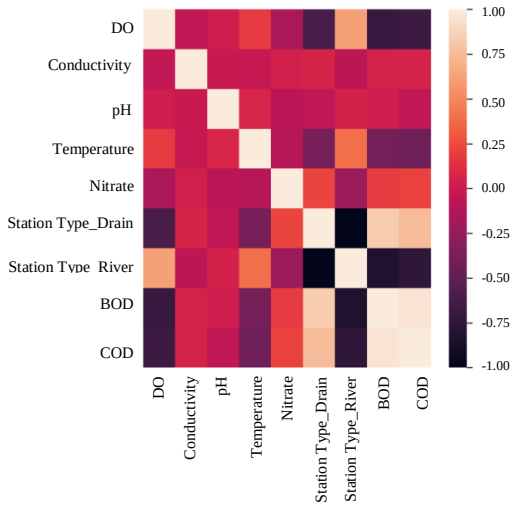

Fig. 1 Correlation heatmap

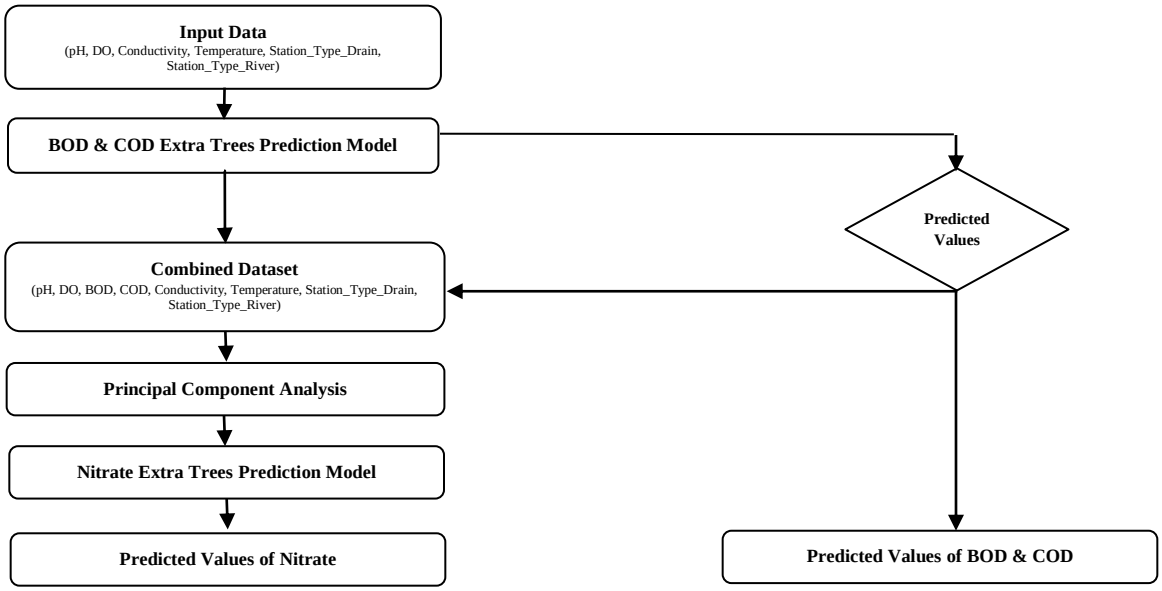

Fig. 2 Modular network design flow

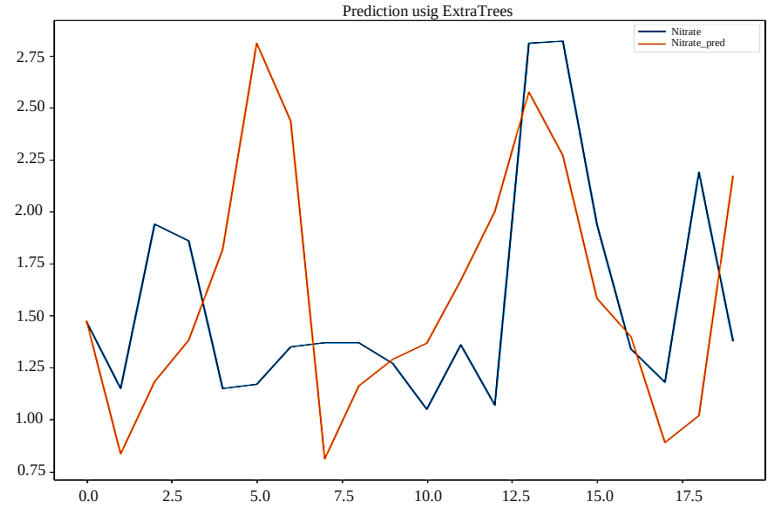

Fig. 3 Nitrate prediction without PCA

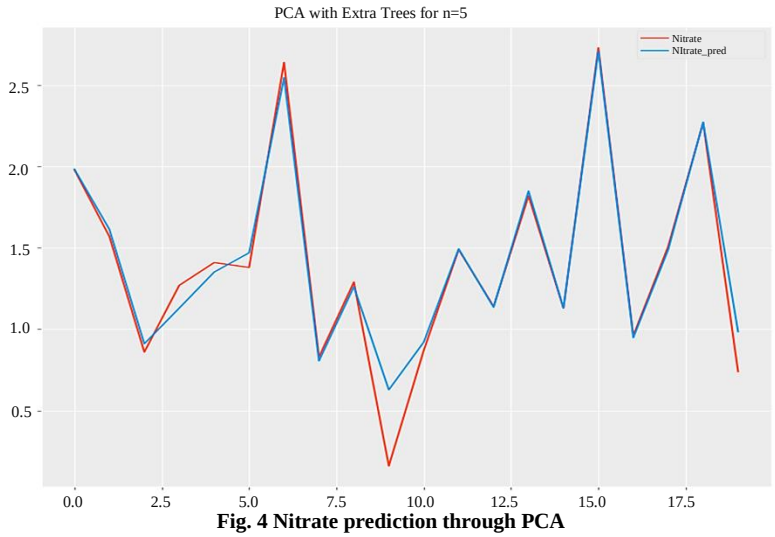

Fig. 4 Nitrate prediction through PCA

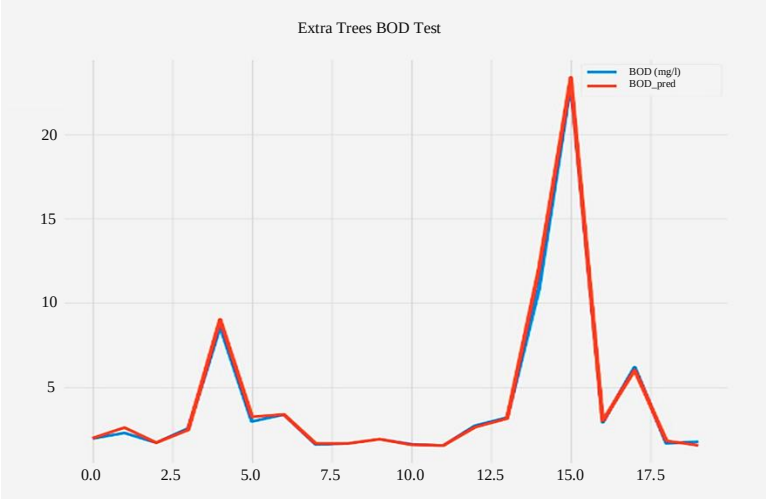

Fig. 5 BOD prediction using extra trees algorithm

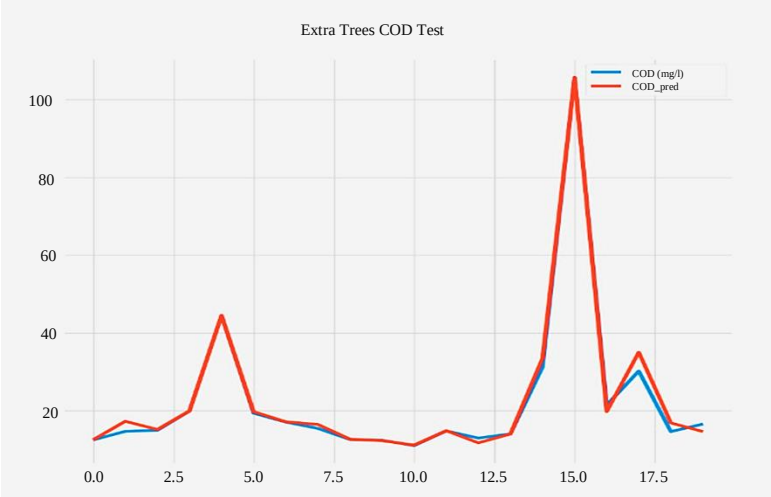

Fig. 6 COD prediction using extra trees algorithm

5. Conclusion

This work attempts to develop a machine learning-based multiparameter-sensing soft sensor with minimal input parameters. The Extra Trees algorithm, a powerful ensemble method, is used to enhance the performance of the developed module by reducing overfitting and increasing generalization. The results also demonstrate the effectiveness of PCA in refining the performance of the predictive unit. Modular networks reduce the data needed for modelling and analysis by using predicted values instead of actual data points, improving the model's reliability.

The presented metrics and the scatter plots provided valuable insights into the model's predictions' accuracy, reliability, and robustness. The proposed model is cost-effective and can be a potential low-cost alternative to traditional methods and costly online sensors for monitoring pollutants in water. It opens up new opportunities while incorporating the suggested technique into macro-systems employed in substantial water bodies. Future research can concentrate on optimizing the selection of principal components to minimize dimensionality reduction and investigate the effects of distinct feature selection techniques on the effectiveness of the predictive model.

Furthermore, the scope of this study can be broadened by integrating additional input parameters into the machine learning model to predict other quantities affecting water quality, such as ammonia and chloride. Also, the focus can be on developing a low-cost hardware model for big water bodies that combines the developed soft sensor for monitoring water quality in real time.

Funding Statement

Thankful to Shri Ramdeobaba College of Engineering and Management for providing the necessary funds to support the validation of results by testing samples in the laboratory.

References

[1] Mengyuan Zhu et al., "A Review of the Application of Machine Learning in Water Quality Evaluation," *Eco-Environment Health*, vol. 1, no. 2, pp. 107–16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Andres Felipe Zambrano et al., "Machine Learning for Manually-Measured Water Quality Prediction in Fish Farming," *PLOS ONE*, pp. 1-17, 2021.[CrossRef] [Google Scholar] [Publisher Link]

[3] Zifei Wang, and Yi Man, "Artificial Intelligence Algorithm Application in Wastewater Treatment Plants: Case Study for COD Load Prediction," *Applications of Artificial Intelligence in Process Systems Engineering*, pp. 143-164, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Sani Isa Abbaa, and Gozenelkiran, "Effluent Prediction of Chemical Oxygen Demand from the Wastewater Treatment Plant using Artificial Neural Network Application," *Procedia Computer Science*, vol. 120, pp. 156-163, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Hemant Bhutada et al., "COD Prediction in Water using Edge Artificial Intelligence," *10th IEEE International Conference on Emerging Trends in Engineering & Technology Signal and Information Processing*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Arunima Sambhuta Pattanayak et al., "Development of Chemical Oxygen on Demand (COD) Soft Sensor Using Edge Intelligence," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14892-14902, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Arunima Sambhuta Pattanayak et al., "Development of Chemical Oxygen on Demand (COD) Soft Sensor using Edge Intelligence," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14892-14902, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Yesilnacar, M.Irfan et al., "Neural Network Prediction of Nitrate in Groundwater of Harran Plain, Turkey," *Environmental Geology*, vol. 56, no. 1, pp. 19–25, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[9] Shine Bedi et al., "Comparative Evaluation of Machine Learning Models for Groundwater Quality Assessment," *Environmental Monitoring and Assessment*, vol. 192, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Okpare Onoharigho Anthony et al., "Development of a Wireless System to Check the Water Level in A Storage Tank," *International Journal of Recent Engineering Science*, vol. 10, no. 2, pp. 40-45, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Masoud Haghbin et al., "Application of Soft Computing Models for Simulating Nitrate Contamination in Groundwater: Comprehensive Review, Assessment, and Future Opportunities," *Archives of Computational Methods in Engineering*, vol. 28, pp. 3569–3591, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Deepak Kumar et al., "Source Characterization and Human Health Risk Assessment of Nitrate in Groundwater of Middle Gangetic Plain, India," *Arabian Journal of Geosciences*, vol. 12, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Remya R S, and Ebin Antony, "Water Quality Analysis of Different Water Sources in Kerala, India," S*SRG International Journal of Agriculture & Environmental Science*, vol. 9, no. 4, pp. 6-11, 2022. [CrossRef] [Publisher Link]

[14] Mengyuan Zhu et al., "A Review of the Application of Machine Learning in Water Quality Evaluation", *Eco-Environment & Health*, vol. 1, no. 2, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Umair Ahmed et al., "Efficient Water Quality Prediction using Supervised Machine Learning," *Water*, vol. 11, no. 11, p. 2210, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Sakshi Khullar, and Nanhey Singh, "Machine Learning Techniques in River Water Quality Modelling: A Research Travelogue," *Water Supply*, vol. 21, no. 1, pp. 1–13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Eustace M. Dogo et al., "A Survey of Machine Learning Methods Applied to Anomaly Detection on Drinking-Water Quality Data," *Urban Water Journal*, vol. 16, no. 3, pp. 235-248, 2019. [CrossRef] [Google Scholar] [Publisher Link

[18] Aayush Bhattarai et al., "Prediction of Nitrate and Phosphorus Concentrations using Machine Learning Algorithms in Watersheds with Different Landuse," *Water*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Hussam Eldin Elzain et al., "Comparative Study of Machine Learning Models for Evaluating Groundwater Vulnerability to Nitrate Contamination," *Ecotoxicology, and Environmental Safety*, vol. 229, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Christina Stylianoudaki, Ioannis Trichakis, and George P. Karatzas, "Modeling Groundwater Nitrate Contamination using Artificial Neural Networks," *Water*, vol. 14, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Meenakshi Rawat et al., "Modeling of Groundwater Nitrate Contamination Due to Agricultural Activities-A Systematic Review," *Water*, vol. 14, no. 24, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Azadeh Atabati et al., "Modeling Groundwater Nitrate Concentrations using Spatial and Non-Spatial Regression Models in A Semi-Arid Environment," *Water Science and Engineering*, vol. 15, no. 3, pp. 218-227, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Issoufou Ouedraogo, Pierre Defourny, and Marnik Vanclooster, "Application of Random Forest Regression and Comparison of Its Performance to Multiple Linear Regression in Modelling Groundwater Nitrate Concentration at the African Continent Scale," *Hydrogeology Journal*, pp. 1081–1098, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[24] Victor Rodriguez-Galianon et al., "Predictive Modeling of Groundwater Nitrate Pollution using Random Forest and Multisource Variables Related to Intrinsic and Specific Vulnerability: A Case Study in An Agricultural Setting (Southern Spain)," *Science of the Total Environment*, vol. 476–477, pp. 189-206, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[25] Kingdom Kponanyie Dune et al., "Production, Activation and Characterisation of PKS-Biochar from Elaeis Guineensis Biomass Activated with HCL for Optimum Produced Water Treatment," *International Journal of Recent Engineering Science*, vol. 9, no. 1, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Bhawani Shankar Patnaik et al., "Machine Learning-Based Soft Sensor Model for BOD Estimation using Intelligence at Edge," *Complex & Intelligent Systems*, vol. 7, pp. 961–976, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] Wenjing Li, and Junkai Zhang, "Prediction of BOD Concentration Wastewater Treatment Process using A Modular Neural Network in Combination with the Weather Condition," *Applied Sciences*, vol. 10, no. 21, pp. 1-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Opeyemi Oyelesi, and Akingbade Kayode Francis, "Face Recognition for Access Control using PCA Algorithm," *SSRG International Journal of VLSI & Signal Processing*, vol. 4, no. 2, pp. 22-27, 2017. [CrossRef [Publisher Link]

[29] Anju, and Anu Rani, "Face Recognition using PCA Gobor Filter SVM Techniques," *International Journal of Computer & Organization Trends*, vol. 6, no. 3, pp. 20-23, 2016. [CrossRef] [Publisher Link]

[30] Jake Lever, Martin Krzywinski, and Naomi Altman, "Principal Component Analysis," *Nature Methods*, vol. 14, pp. 641–642, 2017. [CrossRef] [Publisher Link]