UNIVERSITI TEKNOLOGI MARA
Cawangan Kedah
Kampus Sungai Petani

2021 iCMS
INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e-PROCEEDINGS

of The 5th International Conference on Computing, Mathematics and Statistics (iCMS2021)

4-5 August 2021

Driving Research Towards Excellence

e-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

Editor-in-Chief: Norin Rahayu Shamsuddin

Editorial team:
Dr. Afida Ahamad
Dr. Norliana Mohd Najib
Dr. Nor Athirah Mohd Zin
Dr. Siti Nur Alwani Salleh
Kartini Kasim
Dr. Ida Normaya Mohd Nasir
Kamarul Ariffin Mansor

e-ISBN: 978-967-2948-12-4
DOI

Library of Congress Control Number:

Publication by
Department of Mathematical Sciences
Faculty of Computer & Mathematical Sciences
UiTM Kedah

TABLE OF CONTENT

PART 1: MATHEMATICS

PART 2: STATISTICS

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

PART 4: OTHERS

PREDICTION OF BIOCHEMICAL OXYGEN DEMAND IN MEXICAN SURFACE WATERS USING MACHINE LEARNING

Maximiliano Guzmán-Fernández[1], Misael Zambrano-de la Torre[2], Claudia Sifuentes-Gallardo[3], Oscar Cruz-Dominguez[4], Carlos Bautista-Capetillo[5], Juan Badillo-de Loera[6], Efrén González Ramírez[7], Héctor Durán-Muñoz[8]

[1,2,3,6,7,8] Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, México,
[4] Universidad Politécnica de Zacatecas, [5] Doctorado en Ciencias de la Ingeniería, Universidad Autónoma de Zacatecas, México.
([1] maxguzman1@hotmail.com, [2] misaelzambrano1997@gmail.com, [3] clauger17@gmail.com, [4] racso_zurc@hotmail.com, [5] baucap@uaz.edu.mx, [6] l_badillo@uaz.edu.mx, [7] gonzalezefren@uaz.edu.mx, [8] hectorduranm@hotmail.com)

The monitoring of surface water quality is insufficient in Mexico due to the limited water monitoring stations. The main monitoring parameter to evaluate surface water quality is the biochemical oxygen demand. This parameter estimates the biodegradable organic matter present in the water. Concentrations above 30 mg/l indicates a high level of contamination by domestic and industrial waste. Therefore, the aim of this work to provide a reference to the conventional process of determining biochemical oxygen demand using machine learning. The database used was collected by the National Water Commission (CONAGUA). Pearson's correlation and Forward Selection techniques were applied to identify the parameters with the most important contribution to prediction of biochemical oxygen demand. Two groups were formed and used as input to four machine learning algorithms. Random forest algorithm obtained the best performance. Group 1 and 2 of parameters obtained a 0.76 and 0.75 coefficient of determination respectively. This allows choosing an adequate group of parameters that can be determined with the chemical analysis instruments available in the study area.

***Keywords*:** Machine Learning, Biochemical Oxygen Demand, Mexican Surface Waters.

1. Introduction

The preservation, treatment, access, and efficient use of water is fundamental for humanity, and several countries consider it as a national security resource. Due to the importance of this resource, the concept of water security has been discussed and defined by several institutions, such as the United Nations Water Group (UN-Water), the Economic Commission for Latin America and the Caribbean (CEPAL), among others. The use of water from rivers, wells and lagoons is of great importance, as they are the main sources of water supply in the municipalities and states for various uses. However, activities such as mining, livestock, agriculture and industrial demand generate water exploitation and contamination (Raynal, 2020). Water quality is an important factor to consider, whether for ecosystem needs or for contamination levels that directly impact food, hygiene, health, and economy. To ensure the safe use of water, continuous monitoring of water quality parameters in the water supply sources and discharge areas should be carried out.

In Mexico, the agency in charge of managing, regulating, controlling and protecting the country's national waters is the National Water Commission (CONAGUA). CONAGUA performs the monitoring of the main water bodies in the country, both surface and groundwater. Biochemical oxygen demand is one of the main parameters when evaluating surface water quality at monitoring sites in Mexico. This parameter indicates the biodegradable organic material present in the sample of surface water bodies after 5 days. Based on the contamination level established by CONAGUA, if the biochemical oxygen demand is above 30 mg/l, the water is considered contaminated. This parameter is normally obtained by taking samples in the study area and then transferring them to a laboratory for subsequent analysis of the sample. Sample analysis is through biochemical and

manometric methods, involving specialized instruments and reagents. This conventional process of collecting samples from the study area and analyzing them in the laboratory requires considerable time and labor. As a consequence, it is not possible to have real-time monitoring of water quality. In addition, the diagnosis of contamination is reduced to identifying and analyzing more frequently surface water monitoring sites that are located near certified laboratory infrastructure.

To assist the study process performed by the specialists, it is possible to perform statistical analyses and generate predictive models using artificial intelligence, based on measurement data previously obtained by the specialists. The implementation of artificial intelligence through machine learning and data mining using algorithms for the prediction of water quality parameters in different monitoring zones has been reported in the literature. They are characterized by different stages such as preprocessing, normalization and evaluation of the supervised learning algorithms with goodness-of-fit statistics. For the analysis in rivers, water quality was classified by temperature, pH, turbidity and total dissolved solids. Data are collected in a river and water quality is classified using K-Nearest Neighbors, support vector machine, Bayesian classifier and decision trees. The performance of the algorithms is evaluated by sensitivity, specificity, accuracy and precision (Rosero et al., 2020). This indicates the possibility of involving easily determinable parameters in the study area to diagnose contamination. Similarly, the implementation of the support vector machine algorithm for water quality prediction obtained a correlation coefficient of 0.97 and 0.058 mean square error. The data was collected from the Malaysian Department of Environment. The parameters used as input to the algorithms were ph, dissolved oxygen, biochemical oxygen demand, chemical oxygen demand and ammonia-nitrogen (Abobakr Yahya et al., 2019). These results show that machine learning algorithms can be implemented for the prediction of particular parameters according to local conditions. Different machine learning algorithms such as polynomial regression, model tree and gene expression programming have been implemented for the prediction of biochemical oxygen demand. Ca2+, Na+, Mg2+, NO-2, NO-3, PO3-4, electrical conductivity, pH and turbidity were the input parameters. The gene expression programming algorithm performed acceptably with 5.388 root mean square error and 0.86 correlation coefficient (Najafzadeh et al., 2019). However, most of the parameter groups used as input to algorithms for prediction of biochemical oxygen demand are chosen depending on laboratory capabilities and no matter the determination time. Therefore, choosing parameter groups that are obtained quicker than determining biochemical oxygen demand would save analysis time and allow more study areas to be diagnosed.

The aim of this work is to predict the biochemical oxygen demand in surface waters of Mexico using machine learning algorithms. This study is presented using measurements of water quality parameters from the 2764 surface water monitoring sites in Mexico, acquired by CONAGUA from 2012 to 2019. Pearson's correlation and Forward Selection techniques were applied to select two groups of parameters as input to the multiple linear regression, ridge regression, random forest and elastic net algorithms. The groups of parameters were: (1) if it is possible to transfer the sample to a laboratory, a group of parameters that are obtained quicker than determining biochemical oxygen demand. (2) a group of parameters that can be measured in the study area. The database was split by cross-validation for training and testing of the algorithms. The performance of the algorithms was evaluated by goodness-of-fit statistics. Random forest obtained the best prediction. Similar results were obtained when using both groups of parameters as input. Therefore, this work provides a group of parameters that can be measured in the study area and a group of parameters that can be quickly determined in a laboratory.

2. Materials and Methods

The methodology used in this work, to obtain the prediction of biochemical oxygen demand in surface waters of Mexico, consists of three stages. The first stage was data preprocessing. The second stage consisted of data analysis for the prediction of biochemical oxygen demand.

In the last stage, the machine learning algorithms are validated. The methodology is shown in Figure 1 and was implemented with R studio software version 4.0.2. The following sections describe each stage.

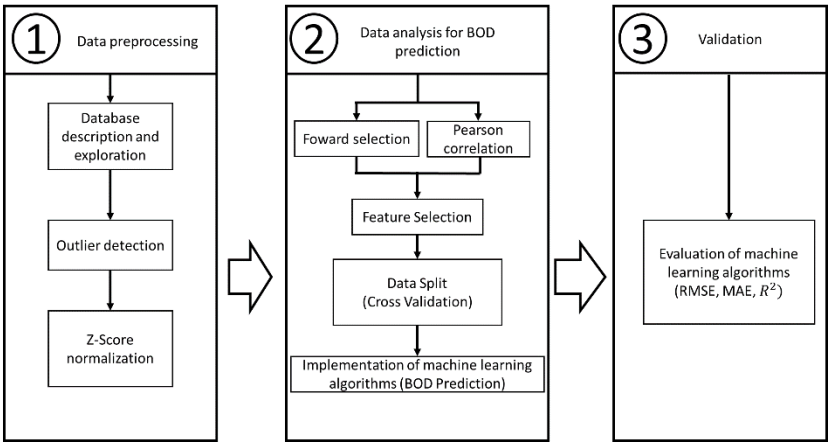

Figure 1: Stages of the methodology: (1) Data preprocessing, (2) Data analysis for prediction of biochemical oxygen demand and (3) Validation

2.1 Data preprocessing

In this first stage, the database collected by CONAGUA was used, which contains indicators of monitoring sites and water parameters from 2012 to 2019. The database is composed of 177 chemical and biological parameters of surface water in Mexico, with a total of 110827 samples. Errors of capture were eliminated and 31 chemical and biological parameters were obtained with a total of 58824 samples. The parameters used and their basic statistics are presented in Table 1.

Table 1: Basic statistics of database parameters.

Parameter (Units)	Min	Mean	Max	Parameter (Units)	Min	Mean	Max
Fecal Coliform (NMP/100mL)	1	55772	24196000	Total Suspended Solids (mg/L)	0.1	105	20812
Escherichia Coli (NMP/100mL)	1	46459	24196000	Turbidity (UNT)	0.01	75	21500
Biochemical Oxygen Demand (mg/L)	0.1	23.2	7667	Arsenic (mg/L)	0.0001	0.006	1
Chemical Oxygen Demand (mg/L)	0.9	77.7	14489	Cadmium (mg/L)	0.00002	0.0002	0.1
Phosphorus (mg/L)	0.001	1.3	95.2	Chromium (mg/L)	0.0002	0.01	76.5
Organic Nitrogen (mg/L)	0	2.5	827.8	Mercury (mg/L)	0.00001	0.0003	0.5
True Color (U Pt/Co)	2.5	55.2	8000	Nickel (mg/L)	0	0.005	7.3
UV Absorbance (U Abs/cm)	0.002	0.17	17	Lead (mg/L)	0.001	0.003	1.8
Total Dissolved Solids (mg/L)	2.4	354.5	159520	Hardness (mg/L)	3.8	295.2	37965
Electrical Conductivity(uS/cm)	3.8	1056	199400	Temperature (°C)	-6	27.6	51
PH (UpH)	2.9	7.8	11.8	Water Temperature (°C)	4	24.9	62
% Dissolved Oxygen (% Saturation)	0.6	73.2	1113.3	Total Organic Carbon (mg/L)	0.06	12.8	2490
Dissolved Oxygen (mg/L)	0.05	5.7	762	Nitrogen (mg/L)	0.008	7.4	1244.1
Ammoniacal Nitrogen (mg/L)	0.003	3.7	497	Kjeldahl Nitrogen (mg/L)	0.003	6.34	1239.8
Nitrogen Dioxide (mg/L)	0.0005	0.1	21.84	Orto-Phosphate (mg/L)	0.0005	0.87	144.4
Nitrate Nitrogen (mg/L)	0.0004	1	336.2				

Box plot and basic statistics were chosen to detect outliers. Most of the parameters varied their maximum values due to measurement errors or collection anomalies. Each parameter was analyzed separately per year and outliers were adjusted to a threshold value. This value was considered as a maximum due to the rest of the measurement values. E.g. in 2012, the Chemical Oxygen Demand had outliers of 14489 mg/L, so the limit value of 250 mg/L was determined and all values exceeding the limit value were assigned to 250 mg/L (Ahmed et al., 2019). To finalize stage one, the parameters were normalized in order to stablish the parameter values on a common scale. The z-score is a method for normalization and standardization that represents the number of standard deviations and allows one to know how far away one is from the mean for each point or raw parameter. Equation (1) shows the z-score normalization expression applied to each parameter, where x represents the parameter value, μ is the mean of the parameter and σ is the standard deviation:

$$z - score = (x - \mu)/\sigma \tag{1}$$

2.2 Data analysis for prediction of biochemical oxygen demand

To begin stage 2, a correlation analysis was performed using Pearson's method. In order to find the dependent and independent variables that have a linear behavior. This method allows us to extract the parameters that have the highest relationship. Equation (2) presents the Pearson correlation between the values of two vector Xi and Yi, where $\bar{x}$ is the mean of the vector x_i $\bar{y}$ is the mean of the vector y_i and n the number of total values in the sample.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2} \tag{2}$$

Continuing with stage 2, the Forward Selection technique was applied to support the selection of parameter groups. This technique first evaluates the individual contribution of each parameter to the prediction of biochemical oxygen demand. Then, the parameters with the highest individual contribution are sorted in descending order and grouped together. Creating database sets by adding one parameter at a time. This set of parameters is used as input to the algorithm and the coefficient of determination is evaluated when predicting biochemical oxygen demand. This process is performed with 70% of measurements for algorithm training and 30% for algorithm testing (Melesse et al., 2020). The algorithm used to apply Forward Selection was multiple linear regression. The goodness-of-fit statistic used to evaluate the individual and joint contribution of the parameters was the coefficient of determination R^2.

After applying the forward selection technique, data split was carried out. The purpose of this division is to validate the algorithms with a balance of the measurements. Cross-validation was the technique used, as this technique divides the data into k subparts and iterates on all subparts of the entire database, having for training k-1 subparts and 1 subpart for testing. In this work was used k=3 since the database consists of 58824 measurements and allows us to use a large balanced number of data for training and testing. 41176 was the number of measurements used for training and 17648 was the number of measurements used for testing.

To finalize stage two, four machine learning algorithms were implemented to predict the biochemical oxygen demand in surface water. The first algorithm used was multiple linear regression, with this algorithm it was possible to obtain an equation of the output variable as a function of the input variables. The second algorithm used was Random Forest, which is based on a decision tree and generates several base models giving good efficiency, it can be used for regression and classification. Also, in this work the Ridge Regression algorithm was used, which uses the same principles as a linear regression, and adds some bias to avoid the effect of having high variances. It also minimizes the sum of the squared residuals. Finally, the Elastic net algorithm, which combines the efficiency of ridge regression, was used in this work. It minimizes the cost function by combining the penalty methods of both algorithms.

2.3 Validation

In stage 3, the algorithms were evaluated in training and testing. The evaluation was through the goodness- of-fit statistics of the root mean square error, mean absolute error and the coefficient of determination. The coefficient of determination was used. It determines the variation that exists between the predictions, the true values and the mean of the values. Equation (3) presents the expression of the coefficient of determination, where y_i are the actual values, y_i' are the predictions, $\bar{y}$ represents the mean of the values and n the number of total values in the sample:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

It was also necessary to use the root mean square error, by scaling the values to the range of the mean square error values. Equation (4) shows the expression, where y_i are the actual values, y_i' are the predictions and n the number of total values in the sample:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2} \tag{4}$$

In addition, the mean absolute error, which represents the sum of the absolute value of the error, was taken and then divided by the total number of values in the sample. Equation (5) shows the expression, where y_i are the actual values, y_i' are the predictions and n the number of total values in the sample:

$$MAE = \frac{1}{n}\|y_i - y_i'\| \tag{5}$$

3. Results and Discussion

3.1 Outlier detection

The box-plot analysis and basic statistics showed that most of the parameters had outliers, with maximum values significantly off the mean, so these values are replaced by the measurement limits for each parameter. Table 2 presents the parameters used and the basic statistics after removing the outliers. By changing the values for each parameter, the data used for training and testing the algorithms were free of bias. Only by modifying the values that seemed to be out of the limits.

Table 2: Basic statistics of the parameters after assigning a limit value.

Parameter (Units)	Min	Mea	Max	Parameter (Units)	Min	Mean	Max
Biochemical Oxygen Demand (mg/L)	0.1	14.7	120	Total Suspended Solids (mg/L)	0.1	66.8	400
Chemical Oxygen Demand (mg/L)	0.9	55.2	250	Phosphorus (mg/L)	0.001	1.2	20
Dissolved Oxygen (mg/L)	0.05	5.7	40	Temperature (°C)	-6	27.6	51
True Color (U Pt/Co)	2.5	45.1	200	Turbidity (UNT)	0.01	49.2	500
UV Absorbance (U Abs/cm)	0.002	0.17	2	Water Temperature (°C)	4	24.9	62
Ammoniacal Nitrogen (mg/L)	0.003	3.7	200	Kjeldahl Nitrogen (mg/L	0.003	6.3	400
Electrical Conductivity(uS/cm)	3.8	900	5000	Total Dissolved Solids (mg/L)	2.4	455.4	1000
Total Organic Carbon (mg/L)	0.06	12.5	1000				

3.2 Features Selection

Based on Pearson's correlation and Forward Selection, parameters were selected for group 1 and group 2. Figure 2 shows the heat map representing the Pearson correlation between the parameters.

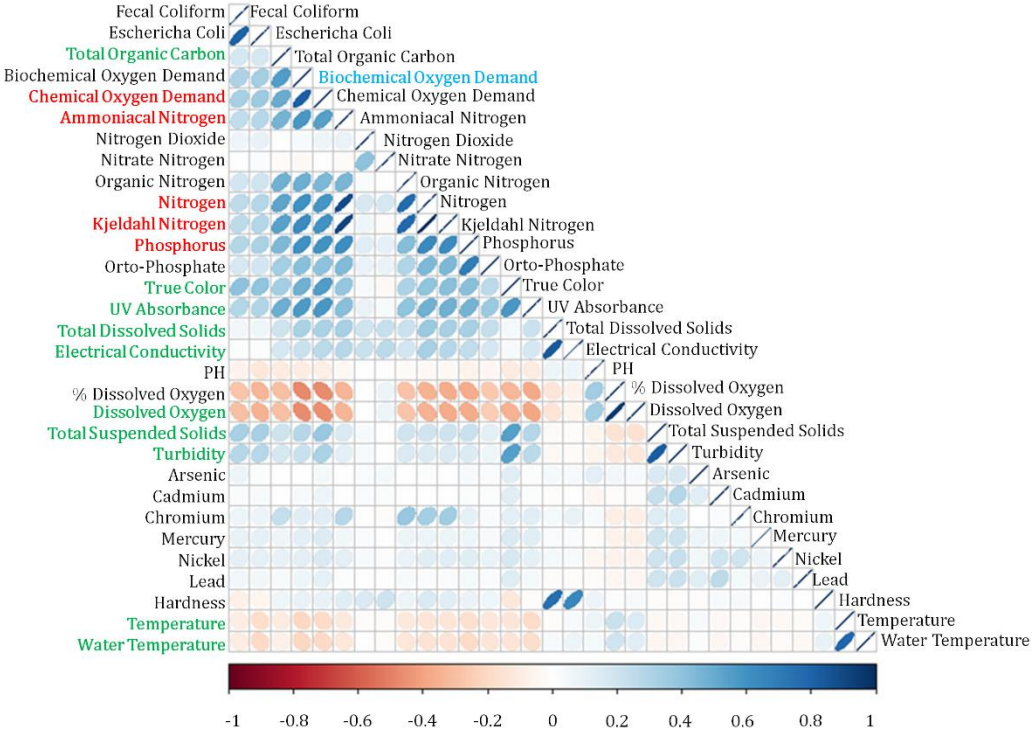

Figure 2: Heat map for Pearson's correlation for all parameters in the processed database

Chemical Oxygen Demand shown a high correlation ($|r| > 0.7$). This parameter involves Biochemical Oxygen Demand, by measuring the complete oxidation of the sample, both organic, biodegradable and non-biodegradable materia (r=0.81). Total Organic Carbon, Ammoniacal Nitrogen, Nitrogen, Kjeldahl Nitrogen, Phosphorus, UV Absorbance, Fecal Coliform, Escherichia Coli, Organic Nitrogen, Ortho-Phosphate, True Color, Total Dissolved Solids, and Dissolved Oxygen, shown a moderate correlation ($0.3 < |r| < 0.7$). This can be related to the method and technique of parameter determination. Nitrogen Dioxide, Nitrate Nitrogen, Electrical Conductivity, PH, Total Suspended Solids, Turbidity, Arsenic, Cadmium, Chromium, Mercury, Nickel, Lead, Hardness, Temperature and Water Temperature shown a weak correlation ($0 < |r| < 0.3$). Electrical Conductivity provides general information on the concentration of salts and ions, so it shows a weak correlation with the Biochemical Oxygen Demand (r=0.21) and high correlation with Total Dissolved Solids (r=0.83).

The results of applying Forward Selection were as follows. The coefficient of determination individually obtained for Chemical Oxygen Demand, Ammoniacal Nitrogen, Kjeldahl Nitrogen and Phosphorus was 0.66, 0.328, 0.39 and 0.36, respectively. The coefficient of determination individually obtained for Total Organic Carbon, True Color, UV Absorption, Total Dissolved Solids, Electrical Conductivity, Total Suspended Solids, Turbidity, Dissolved Oxygen, Water Temperature and Temperature were 0.31, 0.22, 0.31, 0.1, 0.04, 0.07, 0.04, 0.21, 0.04 and 0.04, respectively. After several tests using different combinations of parameters with coefficients of determination greater than 0.3 and less than 0.3 as input to the multiple linear regression algorithm, two sets were formed. This process also allowed us to determine the performance of using a multiple linear regression algorithm with all the parameters available in the processed database and get a reference of the maximum performance of that algorithm. Table 3 shows the increase in the coefficient of determination when grouping the parameters into sets.

Table 3: Coefficient of determination when combining the parameters by forward selection into two sets.

Sets of parameters used as input to multiple linear regression algorithm	Coefficient of Determination [0-1]	
	Training	Testing
Chemical Oxygen Demand, Ammoniacal Nitrogen	0.69	0.69
(Set 1) Chemical Oxygen Demand, Ammoniacal Nitrogen, Kjeldahl Nitrogen, Phosphorus.	0.70	0.70
Total Organic Carbon, True Color, UV Absorbance, Total Dissolved Solids, Electrical Conductivity.	0.48	0.46
(Set 2) Total Organic Carbon, True Color, UV Absorbance, Total Dissolved Solids, Electrical Conductivity, Total Suspended Solids, Turbidity, Dissolved Oxygen, Temperature, Water Temperature.	0.53	0.51

The parameters selected for group 1 were: Chemical Oxygen Demand, Ammoniacal Nitrogen, Kjeldahl Nitrogen and Phosphorus. The parameters selected for group 2 were: Total Organic Carbon, True Color, UV Absorption, Total Dissolved Solids, Electrical Conductivity, Total Suspended Solids, Turbidity, Dissolved Oxygen, Water Temperature and Temperature. According to the methodology used in this work, Forward Selection confirmed the performance and relationship of the Pearson correlation, showing the same parameters that complied with the established groups.

3.3 Validation of machine learning algorithms

After implementing the multiple linear regression, ridge regression, random forest and elastic net algorithms, it was found that the best performing algorithm was random forest. The performance of the algorithms when using group 1 and 2 as input are shown in Table 4 and Table 5.

Table 4: Results in the testing stage of the algorithms using group 1 parameters as input.

Algorithm	Goodness of fit		
	Root Mean Square Error	Coefficient of Determination	Mean Absolute Error
Multiple Linear Regression	0.53	0.7	0.30
Ridge Regression	0.53	0.7	0.30
Random Forest	**0.48**	**0.76**	**0.23**
Elastic Net	0.53	0.7	0.30

Table 5: Results in the testing stage of the algorithms using group 2 parameters as input.

Algorithm	Goodness of fit		
	Root Mean Square Error	Coefficient of Determination	Mean Absolute Error
Multiple Linear Regression	0.67	0.52	0.42
Ridge Regression	0.67	0.52	0.42
Random Forest	**0.48**	**0.75**	**0.24**
Elastic Net	0.67	0.52	0.42

The random forest algorithm obtained optimal results when using the two groups of parameters as input. At the test step, 0.48 RMSE, 0.76 $R2$ and 0.23 MAE were obtained when using group 1. These water quality parameters are determined in laboratories based on Mexican Standards. Selecting these parameters for group 1 can accelerate the determination of biochemical oxygen demand. These parameters do not require a long analysis time in the laboratory.

Similarly, the random forest algorithm obtained 0.48 RMSE, 0.75 $R2$ and 0.24 MAE using group 2. Selecting these parameters for group 2 allows the number of monitoring sites to be greatly expanded. As these parameters can be determined with instruments or sensors in the study area, the diagnosis of water pollution by predicting biochemical oxygen demand is facilitated. This reduces sample transport and analysis time. It offers the possibility of analyzing the required surface water regardless of its location or proximity to chemical laboratories. Additionally, using the groups of parameters identified by this work, different algorithm training techniques could be applied to increase performance such as ensemble learning and genetic algorithms.

4. Conclusion

Water quality is essential for the human life development. Through the present work it was possible to identify the best algorithm that can predict the biochemical oxygen demand in surface waters of Mexico. Also, the parameters that have the most influence. Random forest showed flexibility when implemented in the prediction of biochemical oxygen demand by obtaining 0.48 RMSE, 0.76 R^2 and 0.23 MAE using the parameters Chemical Oxygen Demand, Ammoniacal Nitrogen, Kjeldahl Nitrogen and Phosphorus. In addition, 0.48 RMSE, 0.75 R^2 and 0.24 MAE were obtained using the parameters Total Organic Carbon, True Color, UV Absorption, Total Dissolved Solids, Electrical Conductivity, Total Suspended Solids, Turbidity, Dissolved Oxygen, Water Temperature and Temperature. This indicates that based on the local conditions and the study area, the biochemical oxygen demand can be obtained in a similar way and diagnose water contamination in Mexico in a relatively short time. As a future work, it is proposed to design and develop a real-time electronic monitoring device to measure the parameters of group 2 obtained in this work.

Acknowledgment

We would like to thank the Mexican National Council for Science and Technology (CONACYT) for their support in all activities.

References

M. E. Raynal Gutierrez. (2020). Water use and consumption: industrial and domestic in water resources of Mexico. Vol. 6, J. A. Raynal-Villasenor, ed. GEWERBESTRASSE 11, 6330 Cham, Switzerland: Springer Nature, 2020, pp. 103–116.

P. D. Rosero-Montalvo, V. F. López-Batista, J. A. Riascos, and D. H. Peluffo-Ordóñez. (2020). Intelligent WSN system for water quality analysis using machine learning algorithms: a case study (Tahuando river from Ecuador). *Remote Sens*, 12(12).

A. S. Abobakr Yahya., A. N. Ahmed, F. Binti Othman, R. K. Ibrahim, H. A. Afan, A. El-Shafie, C. Fai, M.S. Hossain, M. Ehteram and A. Elshafie. (2019). Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios. *Water*, 11(6).

M. Najafzadeh, A. Ghaemi, and S. Emamgholizadeh. (2019). Prediction of water quality parameters using evolutionary computing-based formulations. *Int. J. Environ. Sci. Technol*, 16(10).

U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, and J. García-Nieto. (2019). Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*, 11(11).

A. M. Melesse, K. Khosravi, J.P. Tiefenbacher, S. Heddam, S. Kim, A. Mosavi, B. T. Pham. (2020). River Water Salinity Prediction Using Hybrid Machine Learning Models. *Water*, 12(10).n

2021 ICMS

INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e ISBN 978-967-2948-12-4

9 7 8 9 6 7 2 9 4 8 1 2 4