# Using technical artificial intelligence Modeling to forecast the management of the water quality index.

**Dr. Dalia Younis**
**Orcid ID -0000-0001-5104-2460**
**Arab Academy of Science Technology - And Maritime Transport (AASTMT) - Egypt**
**Dyounis1@aast.edu , Dalia_Younis1@Yahoo.Com**

## Abstract:

An essential tool for risk-based management of water resource systems is the assessment and forecasting of water quality. Numerous water quality indicators, charts, and standards have been produced for these objectives, and they have been applied based on water uses. Since the advent of computing technology, numerical models have been used frequently to simulate processes affecting water quality. But because these numerical models are not sufficiently user-friendly, there is a significant knowledge gap between model developers and practitioners. Since Artificial Intelligence (AI) has advanced over the past ten years, it is now viable to incorporate the technologies into numerical modeling systems to fill in the gaps. Among the numerous AI-based algorithms available, For predicting water quality, artificial neural networks are more often used. These models, however, need a sizable dataset for both training and validation. The management of water resources for conservation has increased the necessity for forecasting techniques today. In this study, the parameters for the water quality index were determined using an artificial neural network model (ANN). In the calibration of an ANN model, we can obtain a set of coefficients for a linear model. In 2020, seven Sohag and kena water quality metrics were selected at four different locations. The results show that, in comparison to the Multiple Regression Model, the Water Quality Index (WQI) predicted with ANN model produces better output (correlation coefficient).

**Keyword:** water quality index, wáter modelling by AI, water quality in AI, water model prediction, neural network water quality

## 1. Introduction:

Even though access to safe water is a basic human right and a requirement for health and development, hundreds of millions of people in developing nations still do not have it. Inadequate access to clean water supplies and poor sanitation and hygiene practices result in 3.4 million people, mostly children, dying each year from water-related diseases. Despite ongoing government efforts, more than a billion people still do not have access to better water sources (Daliakopoulos et al. , the international community, and civil society. The problem with water quality is much bigger in scope. It is becoming more and more clear that many of the current developed sources in place in developing countries do not provide water of sufficient quality for domestic use. A well-known example of this is the widespread poisoning of tubewells in Asia by naturally occurring arsenic. Despite how terrible this and other instances of chemical contamination are, microbiological contamination, particularly from feces, is the main cause for concern. Groundwater is typically of significantly higher microbiological quality than surface water, despite the fact that a growing number of sources and systems used by people for drinking and cooking water are not adequately protected from faecal contamination (Daliakopoulos et al. , 2020). This is due to a variety of factors, including urbanization, population pressure, and improper water infrastructure design, implementation, and maintenance. Families cannot be guaranteed access to safe water despite properly protected supplies and well-managed systems (Doan et al. , 2020). Many people still need to physically transport and store water in their homes because the majority of people on the planet lack access to dependable household water connections. Studies show that even water gathered from trustworthy sources is prone to faecal contamination while being transported and stored. Safe sources are important, but the quality of water that people consume can only be ensured with better hygiene, better water storage and handling, improved sanitation, and in some cases, domestic water treatment (Doan et al. , 2020). Because it is less affected by evaporation and pollutants, groundwater serves as the primary source of freshwater in dry and semi-arid regions of the world (Dogan et al. , 2019). Sadly, groundwater withdrawal, pollution, and resource depletion have increased as the human population has grown and as agricultural and industrial activities have occurred (Dogan et al. , 2019). Due to increased groundwater demand brought on by climate change and altered precipitation patterns, aquifer depletion poses a serious threat to human existence in these areas (Dogan et al. , 2019). therefore, access to groundwater is.

## Literature review

In essence, assessing water quality is a categorization issue. Research on unsupervised approaches is extremely active as a result of the inconsistent present standards for measuring water quality (Khalil et al., 2017). Principal component analysis (PCA) and cluster analysis (CA), particularly hierarchical cluster analysis (HCA), are the two most popular unsupervised classification techniques (Rupal et al., 2015). These techniques have been utilized extensively in water quality management, however because of the growth and complexity of data in the aquatic environment, employing these techniques to evaluate water quality is difficult in terms of handling the data (Ouarda et al., 2017). Today's leading technologies for huge data analysis include fuzzy C-means clustering, fuzzy logic, evolutionary algorithms, and K-means clustering (Rene et al., 2016). The Euclidean distances with equal weights method is frequently used in K-means clustering. Recent studies have concentrated on applying different weights to the Euclidean distance (Perrone et al., 2016).

## **Indicators and Numbers:**

The topic of artificial intelligence is the one that computer science and information technology specialists, as well as other scientists and decision-makers, are debating the most. The ultimate goal of artificial intelligence is to create a machine that can think like a human (Dwivedi, Pathak, 2019). These machines are capable of making the best decisions in the most difficult situations by combining cutting-edge analytical techniques with the skills of analysts, engineers, politicians, and other scientists based on the vast amount of information resources at their disposal. The hidden layer plays a significant role in the network power; counting the number of hidden layers is the second step. Therefore, a single figure must be discovered to represent the appropriateness of groundwater for agricultural or human usage.

## **Techniques and model requirements:**

Groundwater Quality Index (GWQI) calculates a single number that represents the total groundwater quality at a specific time and location after analyzing multiple water quality indicators. In order to make complex data useable and understandable for the general public, WQI transforms it into a straightforward indication of groundwater quality. The suitability of groundwater for different purposes has been categorized using WQI in a number of studies on groundwater quality assessment. The weighted arithmetic index approach and the limits for drinking water of the Egypt Standard Specification were utilized in the current study to calculate WQI (Hagan et al., 2019).

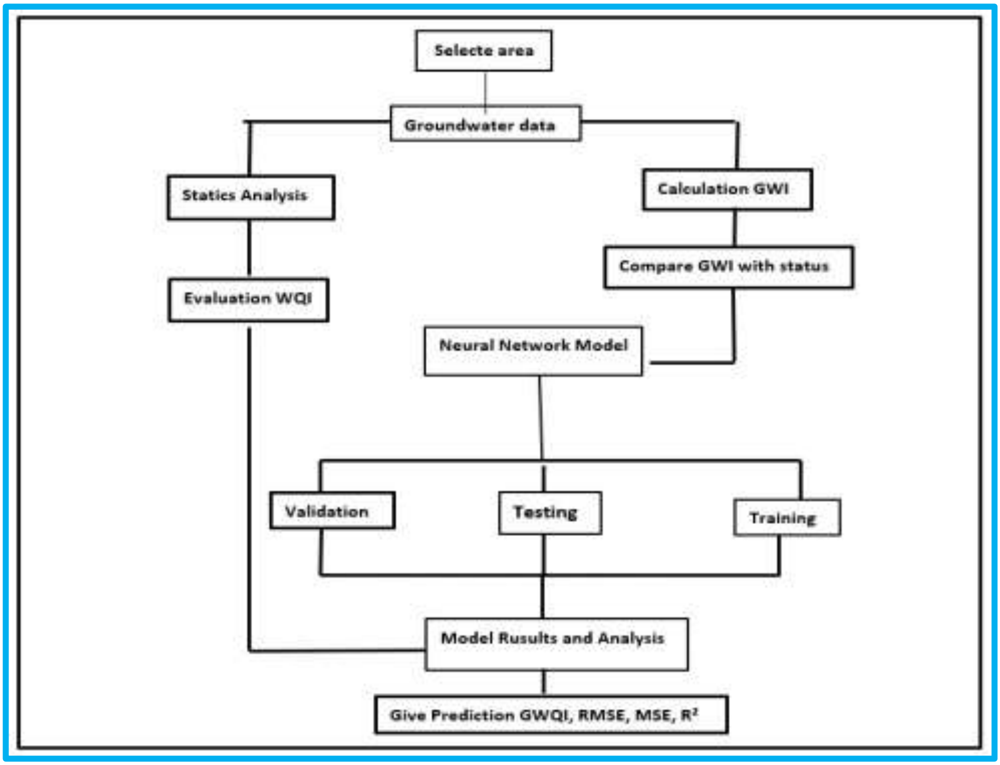



**Figure (1) represented the ANN steps**

The executive and structural steps of the suggested model to find (GWQI) are shown in Figure No. 1. Using a network of artificial neurons, an output data might be predicted from input data in a manner similar to how the human brain functions, where neurons receive input signals and make output signals (Hsieh et al., 2018). Additionally, by training and testing the model using historical data, the mathematical link between the independent variables (chemical parameters of water quality) and the dependent variable (WQI) was investigated for the purpose of modeling water quality. Additionally, the network was used to store the learned information from the input data in order to use it for future WQI predictions (Hsieh et al., 2018). When the inaccuracy of the sample data between the actual and anticipated outputs is minimal, the artificial neurons of the network are typically expanded to other layers while maintaining a link with one another via a weight (Kaurish et al., 2018).

## 4.1. Model Validation

The sum of square error function can be used to calculate the difference between observed and expected values. The prediction efficiency of ANNM was found to be high when the error function was low since the network worked to reduce it during training (Khalil, Ouarda TBMJ, Hilaire, 2017). As an alternative, Khan et al. looked at the R2 coefficient of determination, another statistical measure. R2 values range from 0.0 to 1.0, further demonstrating that the greater the R2 value, The more closely the ANNM matches the input data, the better (Nasiri F, Maqsood I, Huang G, Fuller N, 2017).
The following equations show how to find (GWQI): -

$$\boldsymbol{Par}^{-1} = (\boldsymbol{Ph}^{-1} + \boldsymbol{CL}^{-1} + \boldsymbol{DO}^{-1} + \boldsymbol{Ca}^{-1} + \boldsymbol{EC}^{-1} + \boldsymbol{TDS}^{-1} + \boldsymbol{SO4}^{-1} + \boldsymbol{N}^{-1})^{-1} \quad \textbf{(2)}$$

$$\boldsymbol{Wi} = \frac{\boldsymbol{Par}^{-1}}{\boldsymbol{Si}}$$
(3)

$$\boldsymbol{WQI} = \frac{\mathbf{100}* \sum\{(\boldsymbol{Si}.(\boldsymbol{Si}-\boldsymbol{Ii}).\boldsymbol{Par})^{-1}*(\boldsymbol{Oi}-\boldsymbol{Ii})\}}{\sum(\boldsymbol{Si}*\boldsymbol{Par})^{-1}}$$
(4)

## Empirical results

For determining the quality of groundwater, the chemical parameters such as pH, Cl, SO4, and TDS, were analyzed.

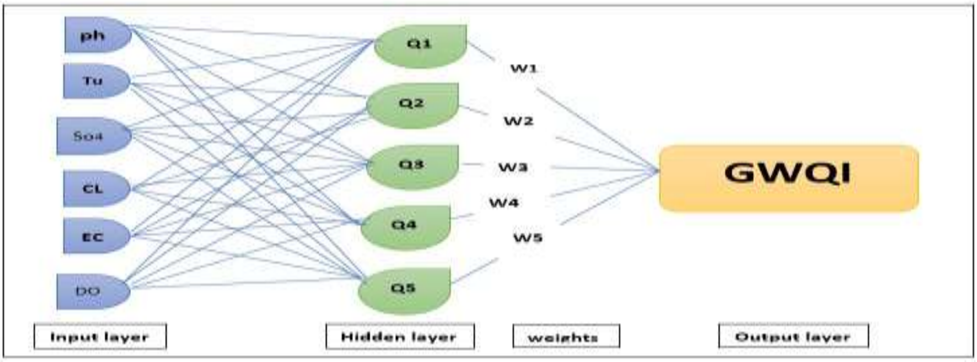



**Figure (2) Structure of NN**

Figure (2) displays how many parameters were chosen based on the relative importance of the total number of parameters, which was reduced to six parameters, including:

## 5.1. PH Parameters: -

The pH value ranges from 6.4 to 11.3, with the highest pH values of 11.3, 10.5, and 9.6 (see Figure). As one of the most important factors of water quality, pH indicates the presence of alkaline or acidic substances in groundwater. As a result, it appears that the pH of the area under inquiry is within the ideal ranges recommended for drinking water.

## 5.2. Chloride (Cl):

In general, Cl is widely distributed in all kinds of rocks. Furthermore, high Cl concentrations in groundwater tests imply that the water has a high organic content, rendering it unfit for human consumption and livestock irrigation (Ouarda TBMJ, Shu, 2017).

**5.3. Sulfate ($SO_4$) :**

Sulfate content in natural water is around 50 mg/l. According to the WHO, 2019 drinking water with a high sulfate content may cause corrosion in the distribution network's pipe systems, along with an unpleasant flavor. 400 mg/l of sulfate is the recommended amount for human intake. The current investigation shows that gypsum content in the soil causes sulfate concentrations in groundwater to reach a maximum value of 12000 mg/L.

**5.4. Total Dissolved Solids (TDS):**

High TDS concentrations might cause gastrointestinal discomfort. Long-term consumption of water with a high TDS can also cause kidney stones and a number of heart ailments. One of the main causes of such high TDS values may be anthropogenic sources, including the disposal of solid waste, home sewage, and agricultural operations. Averaging 8215 mg/L and 4733 mg/L, high TDS levels were found in the majority of the groundwater samples.

Figure No. (3) shows the analyzed values in the laboratories of the National Research Center in Dokki in Figures (3) (B - D - F - H). The parameter values were found, as well as its equation While Figures (3) (E - G - A - C) shows the parameter values in addition to the prediction values of up to thirty wells. Table No. (1) shows the correlation coefficients between the parameters, as well as the relationships between them (direct or inverse) and the extent of the correlation between them (strong or weak). It is noted from the table that all the relationships between the parameters are direct relationships and there is no inverse relationship and this is evidence that the samples obtained They are close to each other and this is evidence of soil homogeneity despite their different locations and despite the presence of five invalid wells, while the column shown in blue indicates that the relationships between the sulfate group and all parameters except for the first four parameters is a weak direct relationship, and this indicates that the values of the rest of the coefficients are small can be neglected. The research of the impact of altering the activation function of the output layer on the performance of the model is then followed by the achievement of the optimal partition dataset. The sigmoid functions were found and the hyperbolic tangent was checked by the automatic architecture. The sigmoid activation function, among others, was discovered to have the smallest error for both training and testing.

The objective was accomplished by separating the input data into training, testing, and holdout samples, and then using the model training approach to generate the model structure (connection weights and hidden neuron counts). Additionally, when creating the model, the data set for the holdout sample was not taken into account (IBM® SPSS® Statistics 19 User Guide). The sum of squares error for training and testing data is used in this study to identify the best partition dataset. The quantity of hidden neurons has an impact on the model's capacity to predict the output data. The model underfits because it is unable to learn properly with a small number of hidden neurons. On the other side, if there are too many hidden neurons, the model loses its ability to generalize, which causes it to be overfit.

## 6. Model Prediction

The scaled conjugate gradient was used as the necessary optimization procedure for calculating the network weights when the ANNM was constructed using the conventional method for rescaling input data. Additionally, by comparing the anticipated and observed WQI values using the ANNM (Perrone MP, Cooper LN, 2016). The importance of the independent variables is found in the measuring of the model's variance during the forecasting of the output values, which is made possible by the independent variables. In this situation, pH has a greater influence on the WQI prediction method used by the model than chloride, total dissolved solids, or sulfate.

Table No. 2 on the criteria used to assess the suitability of well water for use in agriculture and drinking

Table No. (3) shows the method of calculating (GWQI) for any well, by which a well was calculated in the city of Quesna, Menoufia, and it was confirmed that the well was not suitable for drinking or agriculture, as the groundwater quality index was (82.4) .

as well as Table (4) data and the calculation of the groundwater quality index for fifteen A well in Tahta, which is in the governorates of Shohag and Qena, respectively. Figure (5) shows the relative importance of the parameters, which shows that the effect of pH is one of the largest in the selected sample, and this is specific to the selected sample only.

## Conclusion

Some hybrid methods have been These hybrid methods were devised and used by the researcher to combine the strengths of each approach and address the limitations of distinct AI models. Many researchers have used these hybrid approaches to evaluate and forecast water quality. A number of water quality indicators were created with various water quality standards to assess water quality in 16 wells in Egypt using the statistical analysis system, Statistical Package for Social Sciences (SPSS). The results revealed that using artificial intelligence, some of the necessary parameters can be removed without reducing the accuracy of the water quality indicators. And to predict the groundwater quality index in several places in Egypt's wells, a brief summary of the most common hybrid algorithms for artificial intelligence models with statistical analysis that are widely used in predicting water quality has given results very close to each other, but the accuracy of the algorithms for artificial intelligence gave higher accuracy results. Determining the potential of water that may be used safely and continuously for different purposes without lowering its quality is essential, as is alerting people in charge of water management and water users of this. Identifying the necessary groundwater uses and their requirements; establishing sound planning for their development, exploitation, and management; and creating long- and short-term plans for their exploitation. Periodic monitoring and review of the aquifer's behavior to keep track of any changes that may take place in terms of quantity and quality. In order to determine the WQI, which is necessary to assess the suitability of groundwater for drinking purposes as well as its suitability and, consequently, the validity of the artesian well, we conducted this study. The purpose of this study was to analyze the quality of groundwater in Cairo, Egypt. We have observed that the majority of the chemical parameters surpassed the permitted standard limits in Egypt when the results were evaluated.

It thus renders them toxic if utilized in Egyptian agriculture and unfit for human consumption. Furthermore, the quality of the groundwater in the research area has been impacted by the presence of sewage, the dumping of solid and liquid waste, and all other human activities like different enterprises. ANNM was created with the intent of improving future groundwater quality predictions. The lowest values of error functions for training and testing as well as a high R2 value were attained using this model. As a result, the model has a high forecast accuracy. The most important parameters, pH and chloride, had an impact on the model.

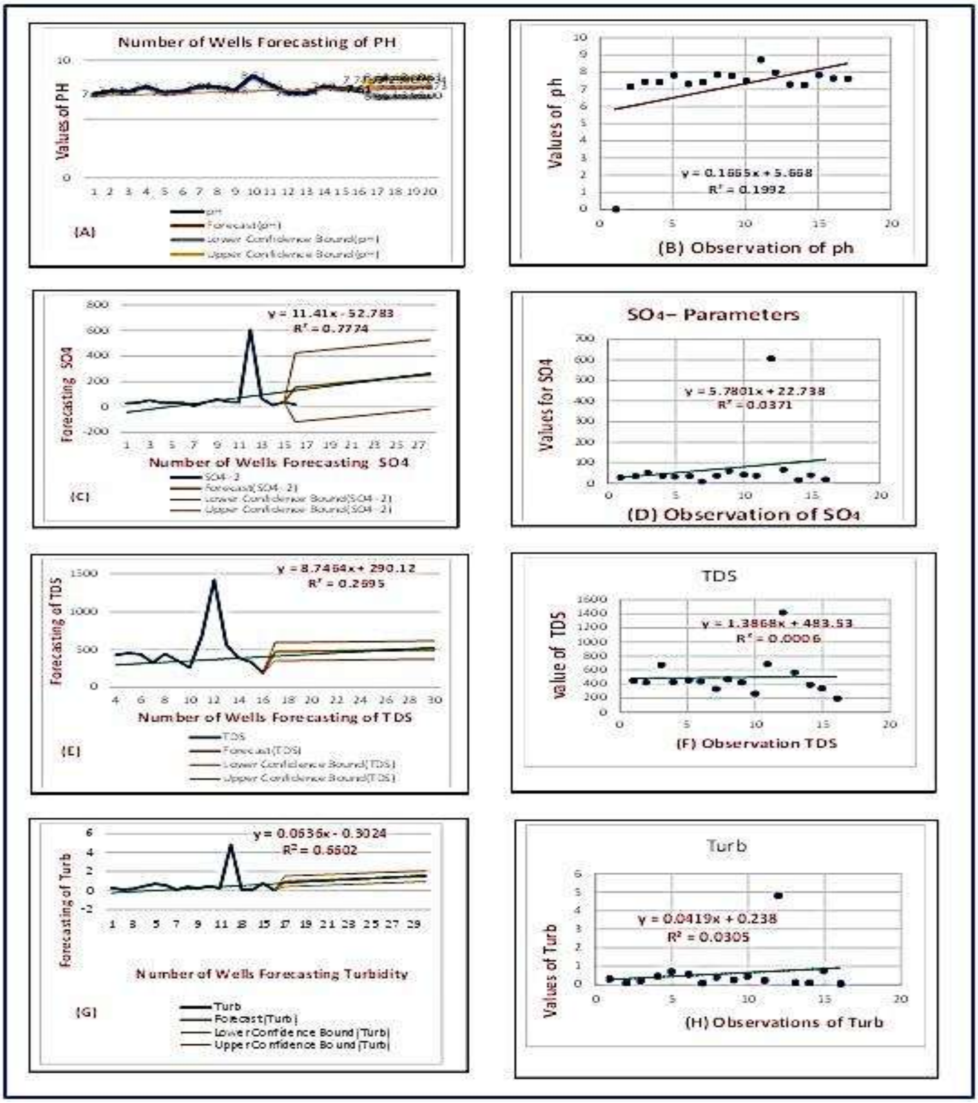



**Figure (3) Observations and Forecasting of Parameters from (A- H)**

## Table (1) Correlation between all Parameters

| | ph | TDS | TH | Turb | SO4-2 | Cl- | NO3- | F- | Na+ | Cu | Zn | Pb | Fe | Mn | Cr | Ecoli |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pH | 1 | | | | | | | | | | | | | | | |
| TDS | 0.997528 | 1 | | | | | | | | | | | | | | |
| TH | 0.974456 | 0.983667 | 1 | | | | | | | | | | | | | |
| Turb | 0.970257 | 0.979898 | 0.992056 | 1 | | | | | | | | | | | | |
| $SO_4^{-2}$ | 0.696613 | 0.695695 | 0.648574 | 0.643348 | 1 | | | | | | | | | | | |
| $Cl^-$ | 0.983985 | 0.990348 | 0.995534 | 0.997359 | 0.656035 | 1 | | | | | | | | | | |
| $NO_3$- | 0.980312 | 0.983488 | 0.985687 | 0.99459 | 0.653771 | 0.995472 | 1 | | | | | | | | | |
| F- | 0.919643 | 0.937239 | 0.98378 | 0.97562 | 0.608319 | 0.969781 | 0.958388 | 1 | | | | | | | | |
| $Na^+$ | 0.964724 | 0.974035 | 0.996747 | 0.987369 | 0.636312 | 0.990595 | 0.97759 | 0.985061 | 1 | | | | | | | |
| Cu | 0.703775 | 0.737612 | 0.82817 | 0.843212 | 0.46451 | 0.810408 | 0.809391 | 0.905338 | 0.833748 | 1 | | | | | | |
| Zn | 0.880504 | 0.903472 | 0.962041 | 0.957078 | 0.583527 | 0.945787 | 0.931418 | 0.992286 | 0.965125 | 0.917739 | 1 | | | | | |
| Pb | 0.912118 | 0.921576 | 0.942224 | 0.947683 | 0.590961 | 0.944848 | 0.93328 | 0.932674 | 0.957363 | 0.839339 | 0.903139 | 1 | | | | |
| Fe | 0.989008 | 0.991364 | 0.990413 | 0.976014 | 0.649577 | 0.988498 | 0.976314 | 0.953098 | 0.987967 | 0.751574 | 0.919664 | 0.936391 | 1 | | | |
| Mn | 0.988186 | 0.990938 | 0.993214 | 0.992896 | 0.660091 | 0.998284 | 0.996116 | 0.962849 | 0.987685 | 0.789132 | 0.935887 | 0.934646 | 0.990238 | 1 | | |
| Cr | 0.989202 | 0.990564 | 0.991707 | 0.986705 | 0.663442 | 0.994974 | 0.989375 | 0.958536 | 0.990377 | 0.777998 | 0.926459 | 0.950324 | 0.99536 | 0.996819 | 1 | |
| E.coli | 0.96477 | 0.974672 | 0.995941 | 0.996845 | 0.640906 | 0.995282 | 0.990117 | 0.986659 | 0.994715 | 0.857736 | 0.967701 | 0.953929 | 0.979437 | 0.991667 | 0.988653 | 1 |

**Table (2) WQI and status of water quality**

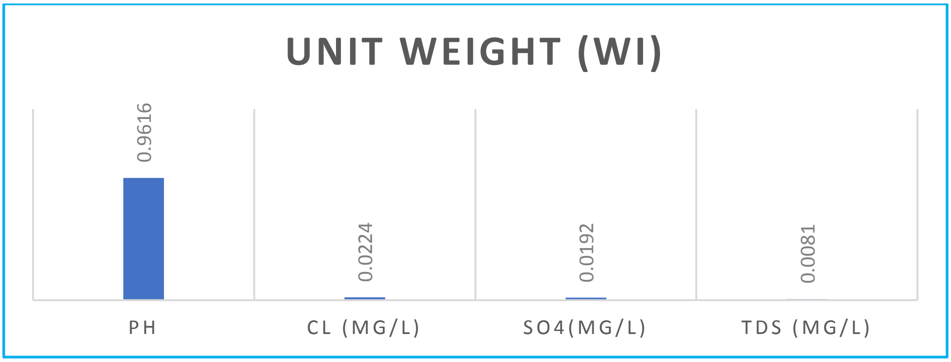
| WQI value | Water Quality Status |
|---|---|
| 0 - 25 | Excellent |
| 26 - 50 | Good |
| 51 - 75 | Poor |
| 76 - 100 | Very Poor |
| > 100 | Unfit for human drinking purpose |

## Table (3) Calculation the GWQI with eight parameters

| Parameters | Observe values( Oi) | Standard values(Si) | Initial values(Ii) | $\frac{1}{Si}$ | K | Wi | Qi | Wi .Qi |
|---|---|---|---|---|---|---|---|---|
| Ph | 7.7 | 8.5 | 7 | 0.1176 | 2.74 | 0.322 | 46.67 | 14.93 |
| CL | 750 | 350 | 755 | 0.0028 | 2.74 | 0.008 | 1.350 | 0.011 |
| SO4 | 8.0 | 200 | 0.42 | 0.005 | 2.74 | 0.014 | 3.800 | 0.053 |
| TDS | 47.0 | 500 | 3.28 | 0.002 | 2.74 | 0.005 | 8.800 | 0.044 |
| EC | 98 | 250 | 1.14 | 0.004 | 2.74 | 0.011 | 38.520 | 0.423 |
| DO | 3.6 | 5 | 11.66 | 0.2000 | 2.74 | 0.548 | 121.00 | 66.308 |
| N2 | 1.9 | 50 | 0.2 | 0.0200 | 2.74 | 0.055 | 3.410 | 0.1876 |
| Ca | 10.5 | 75 | 0.64 | 0.0130 | 2.74 | 0.040 | 13.250 | 0.530 |
| Σ | | | | 0.3644 | | 1.001 | | 82.487 |

## Table (4) Calculation the GWQI with fifteen wells

| Parameters | Well 1 | Well 2 | Well 3 | Well 4 | Well 5 | Well 6 | Well 7 | Well 8 | Well 9 | Well 10 | Well 11 | Well 12 | Well 13 | Well 14 | Well 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pH | 7.58 | 7.34 | 7.4 | 7.8 | 7.45 | 6.7 | 7.65 | 7.79 | 6.73 | 8.65 | 7.52 | 7.12 | 7.52 | 7.55 | 7.65 |
| TDS | 460 | 433 | 671 | 429 | 498 | 455 | 344 | 470 | 453 | 265 | 671 | 749 | 576 | 431 | 379 |
| TH | 455 | 371 | 470 | 263 | 433 | 363 | 222 | 245 | 342 | 321 | 296 | 798 | 568 | 299 | 283 |
| Turb | 0.34 | 0.2 | 0.2 | 0.45 | 0.87 | 0.55 | 0.12 | 0.46 | 0.31 | 0.45 | 0.28 | 0.86 | 0.37 | 0.35 | 0.79 |
| $SO_4^{-2}$ | 31 | 39 | 53 | 36 | 43 | 35 | 39 | 36 | 43 | 49 | 55 | 60 | 67 | 68 | 58 |
| $Cl^-$ | 55 | 73 | 160 | 68 | 90 | 74 | 54 | 143 | 102 | 55 | 159 | 180 | 143 | 112 | 87 |
| $NO_3$- | 45 | 63 | 45 | 53 | 49 | 38 | 22 | 38 | 54 | 46 | 53 | 47 | 48 | 43 | 39 |
| F- | 0.22 | 0.36 | 0.48 | 0.45 | 0.37 | 0.36 | 0.36 | 0.64 | 0.64 | 0.54 | 0.44 | 0.65 | 0.69 | 0.48 | 0.86 |
| $Na^+$ | 28 | 45 | 99 | 59 | 46 | 54 | 76 | 91 | 57 | 67 | 78 | 89 | 79 | 54 | 59 |
| Cu | 0.019 | 0.015 | 0.018 | 0.02 | 0.019 | 0.02 | 0.017 | 0.016 | 0.02 | 0.019 | 0.02 | 0.018 | 0.02 | 0.02 | 0.018 |
| Zn | 0.156 | 0.013 | 0.015 | 0.03 | 0.015 | 0.04 | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 | 0.06 | 0.016 | 0.016 | 0.016 |
| Pb | 0.004 | 0.007 | 0.005 | 0.006 | 0.006 | 0.05 | 0.008 | 0.006 | 0.008 | 0.006 | 0.009 | 0.007 | 0.008 | 0.01 | 0.008 |
| Fe | 0.04 | 0.043 | 0.03 | 0.04 | 0.03 | 0.03 | 0.04 | 0.06 | 0.07 | 0.1 | 0.04 | 0.11 | 0.04 | 0.04 | 0.04 |
| Mn | 0.016 | 0.017 | 0.016 | 0.017 | 0.011 | 0.021 | 0.019 | 0.015 | 0.018 | 0.019 | 0.018 | 0.017 | 0.015 | 0.017 | 0.016 |
| Cr | 0.05 | 0.042 | 0.049 | 0.05 | 0.14 | 0.06 | 0.06 | 0.048 | 0.042 | 0.05 | 0.049 | 0.047 | 0.04 | 0.045 | 0.041 |
| *E*.coli | 48 | 45 | 34 | 53 | 54 | 51 | 54 | 43 | 36 | 54 | 59 | 39 | 48 | 52 | 56 |
| GWQI | 44 | 48 | 56 | 48 | 51 | 43 | 37 | 49 | 54 | 42 | 57 | 92 | 48 | 41 | 42 |
| water type | Good | Good | Poor | Good | Poor | Good | Good | Good | Poor | Good | Poor | V. Poor | Good | Good | Good |




## Figure (4) the percentage important of parameters for well

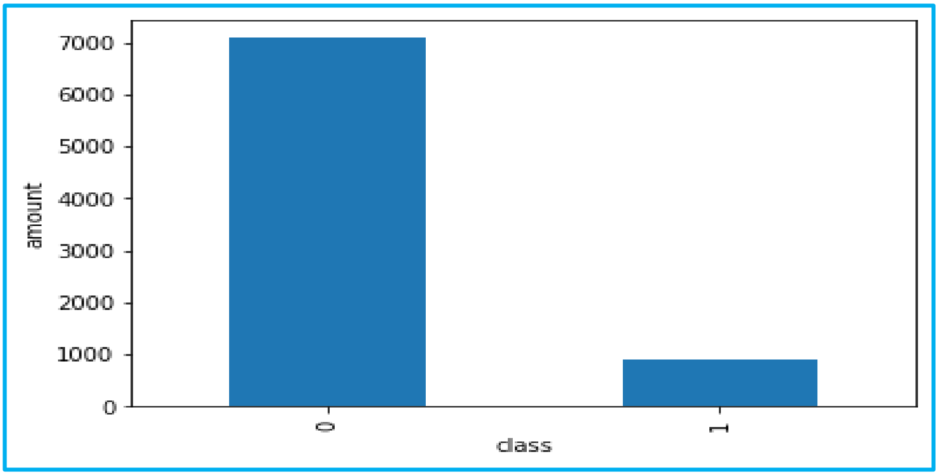



**Figure 1: The class label distribution over the dataset.**

## Table 1: The results of machine learning models.

| Model | Accuracy | Cross-Validation Accuracy | Cross-Validation Balanced accuracy |
|---|---|---|---|
| Dummy classifier | 0.871613 | 0.888945 | 0.500000 |
| Logistic regression | 0.897874 | 0.769500 | 0.598439 |
| Random forest | 0.950813 | 0.788326 | 0.747457 |
| Balanced Random Forest Classifier | 0.932055 | 0.688322 | 0.809368 |
| Logistic regression with proportional class weights | 0.804502 | 0.605859 | 0.686354 |
| Random forest with proportional class weights | 0.947895 | 0.854197 | 0.780058 |
| Logistic regression with Under-sampling SMOTE | 0.794498 | 0.602845 | 0.685647 |
| Random forest with Under-sampling SMOTE | 0.924135 | 0.681049 | 0.797367 |

**Table 2: Classification report for Random forest with balanced class weights**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.95 | 0.99 | 0.97 | 2091 |
| 1 | 0.93 | 0.64 | 0.76 | 308 |
| accuracy | | | 0.95 | 2399 |
| macro avg | 0.94 | 0.82 | 0.86 | 2399 |
| weighted avg | 0.95 | 0.95 | 0.94 | 2399 |

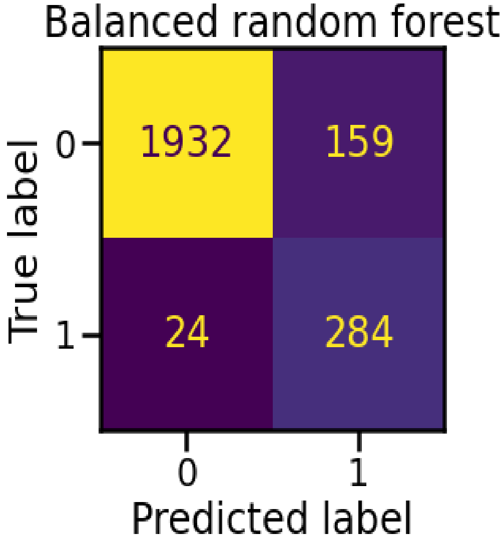



**Figure 2: The confusion matrix for Balanced random forest model.**

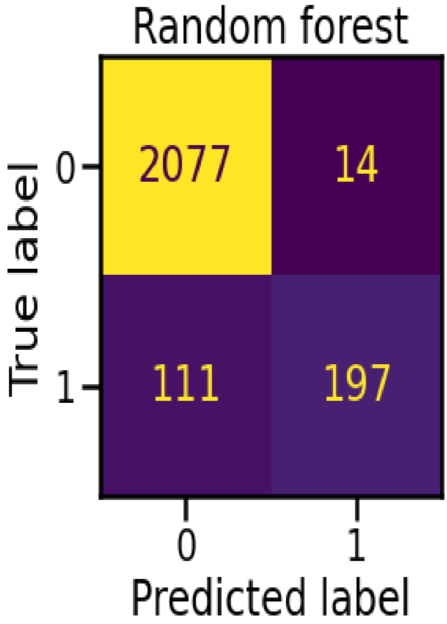



**Figure 3: The confusion matrix for random forest model.**

## Table 3: The the cross validation's findings for Balanced RF Classifier

| Iteration | recall score | accuracy | Balanced accuracy |
|---|---|---|---|
| 1 | 0.976 | 0.976 | 0.976 |
| 2 | 0.981 | 0.972 | 0.972 |
| 3 | 0.981 | 0.976 | 0.976 |
| 4 | 0.987 | 0.977 | 0.977 |
| 5 | 0.988 | 0.9829 | 0.9831 |

## Table 4: The results of the grid search for Balanced Random Forest Classifier

| | Number of Estimators | Max Features | Balanced Accuracy | Mean Recall Score |
|---|---|---|---|---|
| 0 | 10 | 2 | 0.960613 | 0.942787 |
| 1 | 100 | 4 | 0.978073 | 0.982352 |
| 2 | 200 | 6 | 0.983083 | 0.982747 |
| 3 | 1000 | 10 | 0.981615 | 0.983754 |
| 4 | 2000 | 20 | 0.977094 | 0.979137 |

## References:

1 -Daliakopoulos IN, Coulibaly P, Tsanis IK (2020) Groundwater level forecasting using artificial neural networks. J Hydrol 309:229–240
2 - Doan CD, Liong SY (2020) Generalization for multilayer neural network Bayesian regularization or early stopping. The 2nd APHW Conference, Singapore
3 - Dogan E, Sengorur B, Koklu R (2019) Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. J Environ Manage 90:1229–1235
4 - Dwivedi SL, Pathak V (2019) A preliminary assignment of wáter quality index to Mandakini river, Chitrakoot. Ind. J Environ Prot 27:1036–1038
5 - Gazzaz NM, Yusoff MK, Aris AZ, Juahir H, Ramli MF (2019) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar Pollut Bull 64:2409–2420
6 - Hagan MT, Demuth HB, Beale MH (2019) Neural network design. PWS Publishing Company, Boston
7 - Hanrahan G (2019) Computational neural networks driving complex analytical problem solving. Anal Chem 82:4307–4313
8 - Hanrahan G (2018) Artificial neural network in biological and environmental analysis. Taylor and Francis Group, USA
9 - Horton RK (2018) An index number system for rating water quality. J Water Pollut Control Fed 37:300–305
10 - Hsieh WW, Tang B (2018) Applying neural network models to
prediction and data analysis in meteorology and oceanography. Bull Am Meteorol Soc 79:1855–1870
11 - Kaurish FW, Younos T (2018) Developing a standardized wáter quality index for evaluating surface water quality. J Am Water Resour Assoc 43:533–545
12 - Khalil B, Ouarda TBMJ, St-Hilaire A (2017) Estimation of wáter quality characteristics at un-gauged sites using artificial neural networks and canonical correlation analysis. J Hydrol 405:277–287
13 - Nasiri F, Maqsood I, Huang G, Fuller N (2017) Water quality index: a fuzzy river-pollution decision support expert system. J Water Resour Plann Manage 133(2):95–105
14 - Nouri J, Mahvi AH, Babaei AA, Jahed GR, Ahmadpour E (2017) Investigation of heavy metals in groundwater. Pak J Biol Sci 9:377–384

15 - Ouarda TBMJ, Shu C (2017) Regional low-flow frequency análisis using single and ensemble artificial neural networks. Water Resour Res 45:W11428
16 - Patel S, Desai KK (2016) Studies on water quality index of some villages of Surat district. Pollut Res 25:281–285
17 - Perrone MP, Cooper LN (2016) When networks disagree: ensemble methods for hybrid neural networks. Technical report, Institute for Brain and Neural System& Brown University Providence, Rhode Island
18 - Ray C, Klindworth KK (2016) Neural networks for agrichemical vulnerability assessment of rural private wells. J Hydrol Eng 5:162–171
19 - Rene ER, Saidutta MB (2016) Prediction of water quality indices by regression analysis and artificial neural networks. Int J Environ Res 2(2):183–188
20 - Rezania AR, Naseri AA, Albaji M (2015) Assessment of soil properties for irrigation methods in North Andimeshk Plain, Iran. J Food Agric Environ 7:728–733
21 - Rupal M, Tanushree B, Sukalyan C (2015) Quality characterization of
groundwater using water quality index in Surat city, Gujarat, India. Int Res J Environ Sci 1:14–23
22 - Sakizadeh M, Malian A, Ahmadpour E (2015) Groundwater quality
modeling with a small data set. Ground Water. doi:10.1111/ gwat.12317
23 - Shu C, Ouarda TBMJ (2015) Flood frequency analysis at un-gauged sites using artificial neural networks in canonical correlation analysis physiographic space. Water Resour Res 43, W07438
24 - Singh KP, Basant A, Malik A, Jain G (2014) Artificial neural network modeling of the river water quality-A case study. Ecol Model 220:888–895
25 - Tiwari TN, Mishra MA (2104) A preliminary assignment of wáter quality index of major Indian rivers. Indian J Environ Prot 5:276–279 .