International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
www.computersciencejournals.com/ijcai
IJCAI 2024; 5(2): 07-16
Received: 10-05-2024
Accepted: 13-06-2024

Esraa Raheem Alzaidi
College of Since, University of Al-Qadisiyah, Al-Qadisiyah, Iraq

Corresponding Author:
Esraa Raheem Alzaidi
College of Since, University of Al-Qadisiyah, Al-Qadisiyah, Iraq

Water quality classification using machine learning algorithms

Esraa Raheem Alzaidi

DOI: https://doi.org/10.33545/27076571.2024.v5.i2a.91

Abstract
Sustainable development in any field must include achieving social sustainability, environmental sustainability, and economic sustainability. In the field of water, this can be achieved by classifying water quality. When we classify water quality and determine pollution rates, solutions can be found and thus control pollution rates and the nature of pollution. Thus, we have we have achieved environmental sustainability. Also, by classifying water quality, human health can be preserved, and thus social sustainability is achieved. By classifying water quality and using it in irrigation sources and using it in various manufacturing fields, economic sustainability can be achieved. In this study, which aims to clarify the role of machine learning techniques in classifying water quality and how to apply it, the study also aims to clarify the importance of both machine learning techniques and classifying water quality, and through a methodology that is a mixture of methodologies of description, analysis, comparison, and quantitative methodology, a model was created using Neural network technology is one of the machine learning techniques in classifying water quality and predicting any future changes that may affect its quality. The results indicated that the neural network technique, the proposed model, and machine learning techniques are an important and very effective tool in classifying water quality, as the accuracy of the proposed model reached More than 90%, which is consistent with many previous and current studies.

Keywords: Sustainable, sustainability, classify, water quality, machine learning, analysis, neural network technology

1. Introduction

In light of the terrible development that humanity is witnessing, there must be a great struggle to achieve sustainability in all fields, in order to keep pace with this development that is occurring at a terrible rate of acceleration. Sustainability can be defined as preserving the rights of current generations without compromising the rights of future generations, and sustainability is achieved by three parallel axes are economic sustainability, environmental sustainability, and social sustainability. Water is the source of life on the surface of this planet. Therefore, if we want to achieve sustainability for the source of life on the planet, several measures must be followed to achieve sustainability in this sector, including achieving environmental sustainability by preserving water from pollution. Achieving economic sustainability, as water is involved in many industries, and is also the basis of agricultural operations, which is the primary source of food on the surface of this planet [1]. Achieving social sustainability by preserving human health and protecting them from diseases resulting from water pollution. Therefore, it is considered a research topic or topic. The study "Using machine learning techniques to classify water quality" is one of the very important topics that preoccupies many and many workers in the field of water. They conduct a lot of research that depends on technology and its tremendous development in classifying water quality and predicting the causes that can cause its pollution. It also enables through machine learning techniques, predicting the requirements of expected water quantities and people's consumption in light of population increase [2].

Through machine learning techniques, water quality can be classified, as machine learning techniques are techniques that rely on artificial intelligence. Through these techniques, specific training programs and algorithms can be created, which are fed with information about a specific topic and trained to infer, predict, and analyze data. Machine learning can also be defined as a branch of artificial intelligence that focuses on developing algorithms and computer models through learning, meaning that acquiring knowledge from Data that is explicitly programmed to carry out specific tasks, which results in extracting specific patterns and analyzing them to predict future results or take specific strategies according to the inferred data [3]. Machine learning can be divided into many types, including wave learning, undirected learning, semi-directed learning, and reinforcement learning [4]. The importance of machine learning lies in that it analyzes data with great efficiency, and analyzes it with the same speed, efficiency and flexibility, which helps in making decisions based on data analysis [5]. The larger the volume of this data and the greater its diversity, this leads to the accuracy of the analysis and thus the accuracy of the decisions taken. Machine learning techniques can be used to predict future events that have previously been fed into programs. Machine learning techniques can also be used in personalized recommendations for users [6]. It is also used in intelligent automation processes, where through this automation it is possible to identify routine tasks that may be repeated continuously, which provides A lot of human effort and time. Machine learning techniques can be used to improve health services and all sciences, including computer science, as machine learning helps in cybersecurity operations by exploring loopholes or unusual patterns that may indicate the presence of certain attacks, which enhances security and prevention of threats. In short, it can be said that machine learning techniques have become the cornerstone of many important applications, such as applications for analyzing and improving the quality of sounds and images, predicting financial stock prices, medical and engineering fields, computer science, and search engines [7].

This study aims to clarify machine learning techniques, especially neural network techniques, in classifying water quality and predicting ways to apply these techniques and the possibility of finding new methods through which water can be preserved from pollution. The study also aims to clarify the importance of machine learning techniques, and the importance of classifying water quality to achieve sustainability in Water field. All of this is done through a methodology that is a mixture of different methodologies, such as methodologies of description, analysis, comparison, scientific methodology, and quantitative methodology, to formulate a model that classifies water quality, as well as evaluating the performance of this model through accuracy and robustness tests. The study gains its importance from the importance of the topic itself, as the study can It is considered an important scientific reference for researchers and students interested in the field. The importance of the study is also due to the fact that it explains the importance of water quality classification, the importance of machine learning techniques, and their role in achieving sustainability in the field of water.

2. Basic concepts and theoretical analyses

In this section, we will explain and analyze some basic concepts and theoretical analyzes as mentioned in some previous studies

2.1 Techniques of the application of learning techniques

Algorithms and techniques to enable computer systems to extract data according to the following steps:

1. Data collection: It is necessary to provide a large set of data related to the problem. The larger this data is, the more accurate the results will be and the more flexible the learning model will be. This data should also be diverse. 2-Data processing: After collecting the data, the data must be organized, cleaned, and filtered to make it easy to use, remove any missing values and any noise, and convert the data into an appropriate mathematical formula [8].

2. Data processing: After collecting the data, the data must be organized, cleaned, and filtered to make it easy to use, remove any missing values and any noise, and convert the data into an appropriate mathematical form. The process of converting the data into a mathematical model takes place according to a specific methodology, which is to accurately understand the data and determine the required outputs and inputs. necessary, then determining the characteristics of the data, whether they are digital data, text data, or images, then determining their dimensions and converting them into vectors or matrices, after identifying all missing and distorted values. Then the data is converted to a unified measure in a specific mathematical formula; this formula is encoded if the data was categorical, represented numerically and the most important features in the data were selected that would be used for prediction and inference [9].

3. Model selection: An appropriate machine learning algorithm is selected to implement the model. The selection method depends on a group of factors, including:

Choosing the appropriate machine learning algorithm depends on several factors, which are as follows:

- **a) Data characteristics:** Data varies according to its size and type, so this must be taken into account, as there are algorithms that suit small data, such as algorithms such as decision trees, and algorithms that suit large and diverse data, such as (such as CNN). Algorithms that suit the nature of the relationship between variables, so data Random relationships between variables are suitable for algorithms such as (Random Forests), and data with a linear relationship between variables are suitable for algorithms such as (Logistic Regression).
- **b)** The type of application and problem is one of the most important factors, as classification applications are suitable for classification algorithms such as SVM or k-Nearest Neighbors. Regressive applications are suited to algorithms such as regression trees, and clustering applications are suited to clustering algorithms such as k-means or DBSCAN [10].

4. Training the model: The model is trained using a set of data that has been extracted and filtered to train the model in learning, where the model is provided with input and output data and an instruction on how to connect them to each other. Training processes include choosing a learning algorithm, setting training parameters such as learning rate, number of iterations and elicitation via specific algorithms to learn patterns from data

5. Testing the model: Where the model is tested and its performance is evaluated using a set of actual data that was not used during training. These can be called test sets to measure the accuracy and ability of the model to generalize. These tests are as follows:

- **a)** Performance measurement tests use metrics such as accuracy, recall, precision, and F1-Score.

b) **Analysis tests:** These are tests through which the performance of the model on the test set is analyzed and areas that can be improved are identified.

6. Improving the model: The model is improved through a set of procedures, including:

a) **Improving data quality:** Data cleaning: dealing with missing data or outliers, removing noise from the data, and using techniques such as normalization or standardization to improve the quality of anomalous data [11].

b) **Feature selection:** The most important features are selected using methods such as analysis of variance (ANOVA), feature extraction, and dimensionality reduction using techniques such as principal components analysis (PCA) or linear discriminant analysis (LDA).

c) **Hyperparameter Tuning:** Parameters are tuned by Grid Search, where a set of values for the parameters are tested to determine the optimal values, Random Search, where a random set of values for the parameters are tested to improve the model, and Bayesian Optimization, where statistical methods are used to optimize parameters more efficiently [12].

d) **Using advanced techniques:** For example, deep neural network techniques, and reinforcement learning techniques: using reinforcement learning techniques to deal with problems that require making sequential decisions.

e) **Model regularization:** It is a method used to organize the model using L1 Regularization (Lasso), for simplification, and L2 Regularization (Ridge) to prevent overtraining.

f) **Missing data handling:** Missing values are replaced using techniques such as averaging or model.

g) **Error analysis and model modification:** The errors committed by the model are studied and identified to understand their causes.

h) **Monitoring and maintaining the model:** Monitoring the model's performance and updating the model periodically with new data are among the most important factors on which the model's performance depends.

i) **Retraining:** The model must be retrained when the data or basic patterns in the data change.

2.2 Neural network technology

Neural network technology is one of the machine learning techniques. Neural networks are used to imitate the processes carried out by the human brain in processing data and making decisions [13]. This technology relies on an interconnected structure of nodes called layers, where these nodes are organized into multiple layers including the input layer, hidden layers, and output layer.

2.2.1 The neural networks basic components

The neural network consists of two basic components as shown in figure (1), as the following:

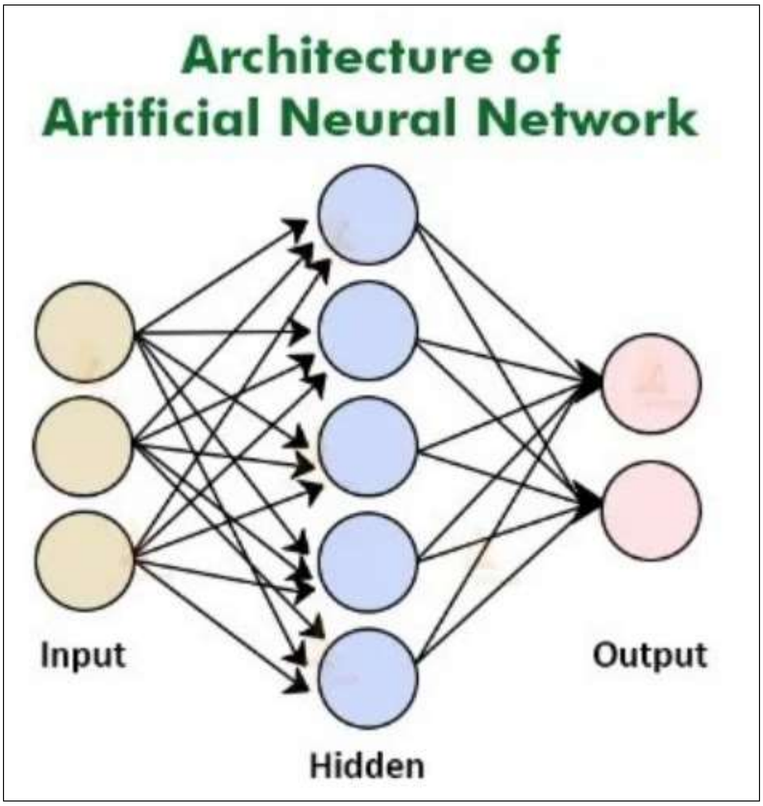

Fig 1: Shows neural networks

a) **Layers:** where the layers represent the input layers, which are the raw data and features that will be processed. The hidden layers, which are special layers for processing data by applying a set of mathematical operations. The output layers, which are the layers responsible for predicting and analysing the results [14].

b) **Nodes:** that can be weights that represent the strength of the connection between the nodes and add biases to the output in order to help the model learn better. The process of training a neural network includes operations Feed-forward, where data is passed through the network from the input layer to the output layer, passing through the other layers, calculating the final output, then calculating the error coefficient and comparing the resulting final output by comparing the final output with the actual output comparison. One of the most common types of neural networks is neural companies with one or more industries. Among the hidden layers there are recursive RNNs which are used to process serial data [15].

2.2.2 Types of neural networks

There are many neural networks, including, as shown in figure (2):

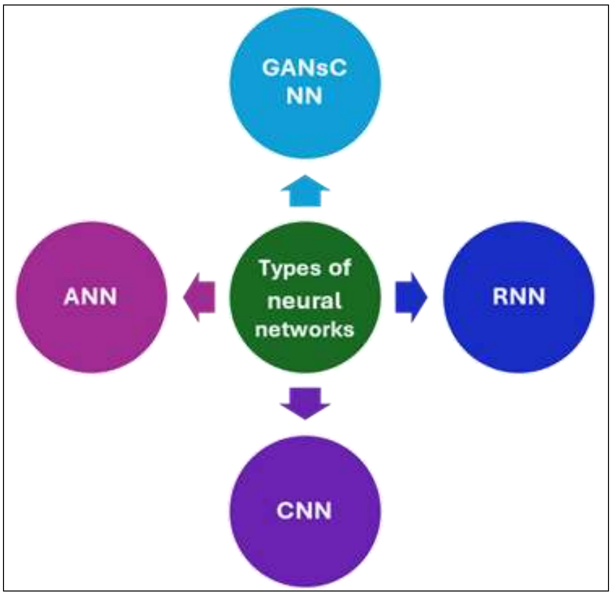

Fig 2: Shows the types of neural networks

a) **Convolutional Neural Networks - CNN:** They are mainly used in image processing and computer vision.
b) **Recurrent Neural Networks - RNN:** They are used to process sequential data such as texts, which are subject to a temporal sequence and depend on an internal memory in which the data is kept.
c) **33. Generative Adversarial Networks - GANs:** They are used in applications that require the generation of new data, and they consist of two models, the Generator and the Discriminator, where each model generates part of the original data set [16].
d) **Advanced Convolutional Neural Networks (Capsule Networks):** These are networks that improve the performance of traditional convolutional neural networks in recognizing patterns and shapes.

Neural network technology is also used to provide personalized recommendations to users, and contributes to intelligent automation processes, through which repetitive routine tasks can be identified, saving a lot of human effort and time. Neural network technologies can be used to improve health services and all sciences, including computer science, as they help in cybersecurity operations by exploring vulnerabilities or unusual patterns that may indicate the presence of certain attacks, which enhances security and threat prevention. In short, it can be said that neural network techniques have become the cornerstone of many important applications, such as analyzing and improving the quality of sounds and images, predicting financial stock prices, medical and engineering fields, computer science, and search engines [17].

3. Methods and methodology

This section refers to the software and hardware devices and tools that were used in the study.

3.1 Sample specifications: The study sample in this study, a sample of water was used from one of the water stations, and it was analysed in one of the laboratories affiliated with the approved Ministry of Health. The data was obtained and the machine learning algorithm (neural network) was fed with this data to test the model that was used to ensure its accuracy. The data was as follows.

Table 1: Shows the standard specifications for water quality

Metal type	Permissible percentage mg/L
lead (pb)	0.01
Mercury (Hg)	0.001
Cadmium (Cd)	0.003:0.001
Chromium (Cr)	0.05:0.1
Nickel (Ni)	0.05:0.02
Arsenic(As)	0.02
irom (Fe)	0.3:05
Magnesium(mg)	0.1:05
Zinc (Zn)	3.0:5.0
Manganese(Mn)	0.1:0.5
Copper(Cu)	1.0:2.0

Table 2: Actual value for the percentages of heavy metals in a water sample from sewage

Metal type	Metal values in a wastewater sample
Lead (PH)	0.05
Mercury (Hg)	0.03
Cadmium (Cd)	0.02
Zinc (Zn)	0.00
Copper (Cu)	0.2
Iron (Fe)	0.9

Table 3: Actual valuse for Bacteria Range in a water sample from sewage

Expected Bacteria Range	Actual Bacteria Range
Over 1000000(cells/mL)	123000000(cells/mL)Municipal tap

3.2 Materials and tools

In conducting the study, a set of tools and data were used, as follows:

- Databases that contain data on algorithms, methods of selecting and applying them, and sample data.
- Computers
- **Data processing tools:** These are tools for processing, analyzing and cleaning data.
- **Neural network technology:** It is one of the machine learning techniques
- **Evaluation and analysis tools:** These are tools for determining accuracy and sensitivity.

3.3 Machine learning model using neural network

3.3.1 Neural network design

Structure of a neural network as the following:

A. Input Layer

Number of nodes: Equal to the number of features in the data set. It is 10 features in the data. They are as follows:

1. pH.
2. (Biochemical Oxygen Demand - BOD).
3. Total dissolved solids (TDS).
4. Dissolved Oxygen – DO.
5. Total Nitrogen – TN.
6. The amount of phosphorus in the water, which is a nutrient that can cause algae growth if it is in high concentrations.
7. Turbidity: A measure of the transparency of water and the amount of suspended matter that affects light transmittance.
8. The ability of water to conduct electricity, which depends on the amount of dissolved ions.
9. Chlorides concentration.
10. Total Organic Carbon (TOC).

A. Hidden Layers

Through the hidden layers, the appropriate activation functions for each layer can be determined. Define the backpropagation algorithm as a learning algorithm for the model.

Through the hidden layers, the appropriate activation functions for each layer can be determined [17]. The backpropagation algorithm can be determined as a learning algorithm for the model. The layers are as follows:

The first hidden layer

- **Number of nodes:** 64.
- **Activation function:** ReLU.

The second hidden layer

- **Number of nodes:** 32.
- **Activation function:** ReL.

Below are examples of activation functions

B. Output Layer

It is a layer of 3 nodes and is as follows:

1. High.
2. Medium.
3. Low.

C. Mathematical modelling

Input Features

$x=[x1,x2,\ldots,x10]x=[x1,x2,\ldots,x10]$

x1: pH.
x2: Biochemical Oxygen Demand (BOD).
x3: Total Dissolved Solids (TDS).
: Dissolved Oxygen (DO).
x5: Total Nitrogen (TN).
x6: Phosphorus.
x7: Turbidity.
x8: Electrical Conductivity.
x9: Chlorides Concentration.
x10: Total Organic Carbon (TOC).

Outputs: yy (three water quality outputs)

y1: Low quality.
y2: Medium quality.
y3: High quality.

The first hidden layer

- **Number of units in the first hidden layer:** n1.
- **Weights for the first hidden layer:** $W1\in Rn1\times 10$.
- **Biases of the first hidden layer:** $b1\in Rn1$.
- **Output of the first hidden layer:** $h1=\sigma(W1x+b1)$.

The second hidden layer

- **Number of units in the second hidden layer:** n2.
- **Weights for the second hidden layer:** $W2\in Rn2\times n1$.
- **Biases for the second hidden layer:** $b2\in Rn2$.
- **The output of the second hidden layer:** $h2=\sigma(W2h1+b2)$.

Final layer (output)

- **Weights for the final layer:** $W3\in R3\times n2$.
- **Biases for the final layer:** $b3\in R3$.
- **Final layer output (output):** $y=SoftMax\ (W3h2+b3)$.

Final mathematical formulas

First hidden layer

$$h1 = \sigma(W1x+b1)$$

Second hidden layer

$$h2 = \sigma(W2h1+b2)\ )$$

Final layer (output)

$$y = SoftMax(W3h2+b3)$$

Details of the functions used

- **σ:** activation function (e.g. ReLU, tanh, sigmoid).
- **SoftMax:** A function that converts values into covalent probabilities

In this way, we have a neural network with water quality classification outputs based on ten features and inputs using two hidden layers.

3.3.2 Training a neural network model

At this stage, the training data set is fed to the neural network, then the error rate is calculated, which is the difference between the actual values and the resulting values divided by the actual values, and the training process is repeated until the model reaches an acceptable performance level [18]. To complete the training process, the following steps can be followed:

1. It is assumed that we have a neural network consisting of L layers, with n_l node in layer l.

The output of layer l is denoted by the vector a_l. The weight of the connections between layer $(l-1)$ and layer l is denoted by the matrix W_l. The output of layer l can be calculated using the following equation:

$$a\_l = f(W\_l * a\_\{l-1\} + b\_l) \qquad eq(1)$$

Where:
f is the activation function.
b_l is the biased class l.

2. Calculating the prediction error

The forecast error is denoted by the variable E, and can be calculated using the following equation:

E = L(y, a_{out}) eq(2)

Where:
y is the desired vector (desired output).
a_{out} is the output of the last layer (layer L).
L is a loss function.

3. Calculating the gradients of the prediction error with respect to the weight of the links

The prediction error gradients are calculated with respect to the weight of each link in the neural network. Calculate the prediction error gradients with respect to the link weight w_lj (from node j in layer (l - 1) to node l in layer l) using the following equation:
The gradients of prediction error are denoted
For the link weight w_ij (of node i in layer (l - 1)
to node j in layer l) by the variable delta_w_ij.

delta_w_lj = (dE / da_l) * (da_l / dw_lj) eq(3)

Where:
dE/da_l is the derivative of the prediction error with respect to the outputs of layer l.
da_l / dw_lj is the derivative of the output of layer l with respect to the link weight w_lj.
4- Updating link weights
The weights of the connections in the neural network are updated, Using the following equation:

w_ij <- w_ij - alpha * delta_w_ij eq(4)

Where:
alpha is the learning rate.
Momentary regression learning rule:

w_ij <- w_ij - alpha * delta_w_ij * a_{l-1j} eq(5)

3.3.3 Neural network model evaluation

The model is tested and its performance is evaluated using a set of actual data that was not used during training.
These can be called test sets to measure the accuracy of the model. The data was divided into 80% data for the model and 20% data for testing its ability to generalize. These tests are as follows:
A. Benchmarking tests use metrics such as Precision, Recall, Precision, and F1-Score. These metrics can be calculated as the following relationships illustrate:

Accuracy = (TP + TN) / (TP + TN + FP + FN) = 100% eq (6)

Recall = TP / (TP + FN) = 100% eq (7)

F1SCORE = 2 *(Precision *Recall)/(Precision + Recall) = 100% eq (8)

F1 Score = 2 * (Precision * Recall) / (Precision + Recall) eq (9)

B. Analysis tests: These are tests through which the model's performance on the test set is analyzed and areas that can be improved are identified.

3.3.4 Model improvement

The model will be improved using hyperparameter tuning, where the parameters are tuned by Grid Search, where a set of parameter values are tested to determine optimal values, Random Search, where a random set of parameter values is tested to improve the model, and Bayesian optimization where statistical methods are used to improve the parameters More efficiently.

4. Results and discussion

In this part, we will explain and analyze the results that were extracted from the practical part of the study and analyze and discuss these results

4.1 Results

Real data for a water sample was used, and machine learning algorithms, especially training algorithms, were fed this data in percentage terms. Evaluation tests were conducted, namely accuracy tests, recall tests, and F1 score tests. A comparison was also made between the expert classifications and the model classifications, the correlation coefficient between both classifications was studied, and the relationship between accuracy and recall F1 score was studied. The results were recorded as follows.

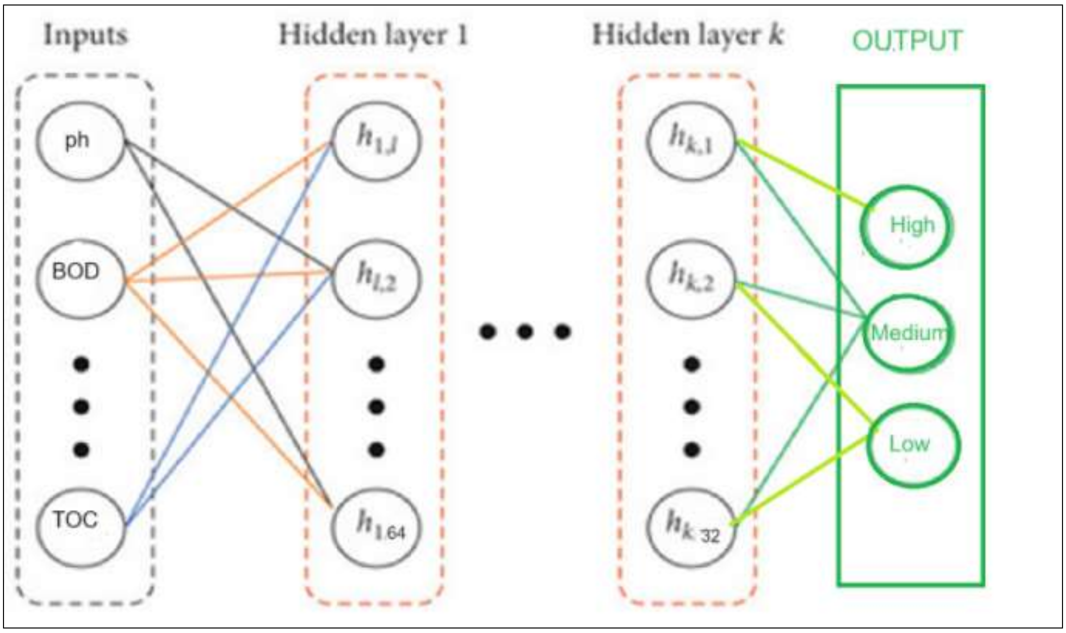

Fig 3: Shows the ANN neural network diagram

According to fig (3), the neural network used in the study, which consists of 10 input layers and three output layers that express the water classification, which is high classification, medium classification, and low classification, in addition to the data processing layers.

Table 4: Percentage accuracy result on the test set

Result of percentage accuracy in the test set		
Number of Records	**Number of Correctly Classified Records**	**Accuracy Percentage (%)**
60	55	91.60%

According to Table No. 4 shows the number of complete registration times, the number of successful registration times, and the accuracy rate, which reached 91.6%. The total number of registration times was 60, and the successful ones were 55. Two attempts at medium output were recorded as low attempts, and three attempts were recorded as low, while these are medium attempts.

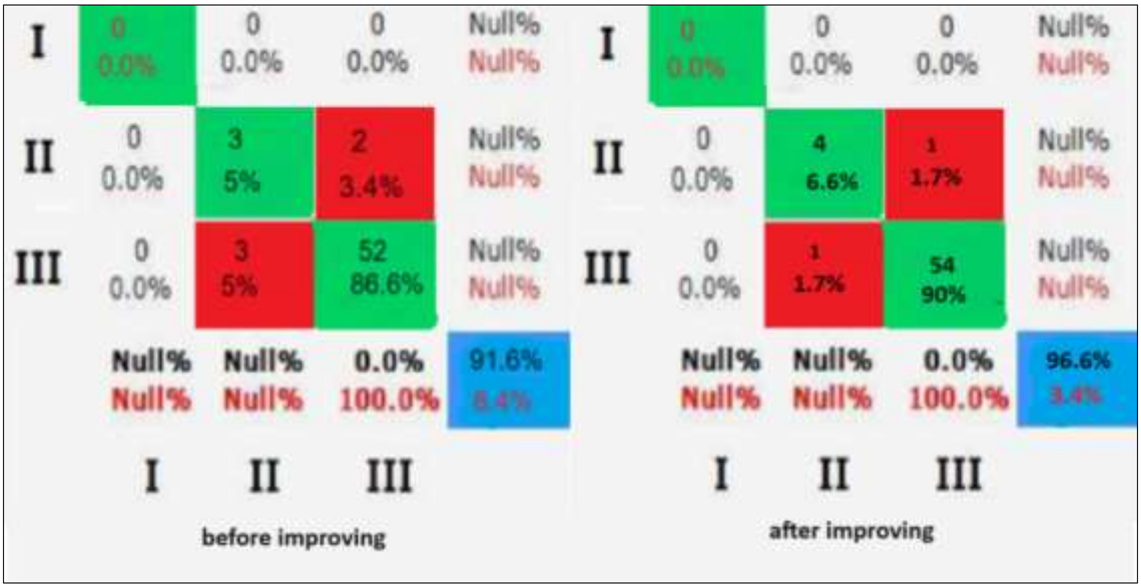

Fig 4: Shows the confusion matrix before and after improving

According to Figure (4), the confusion matrix of the model shows the accuracy of the model before and after the optimization processes, where adjusting the parameters was used as a means to improve the neural network algorithm as a machine learning technique, using Grid Search, where a set of parameter values are tested to determine the optimal values, and random search, Where a random set of parameter values is tested to improve the model, and Bayesian optimization where statistical methods are used to improve parameters more efficiently.

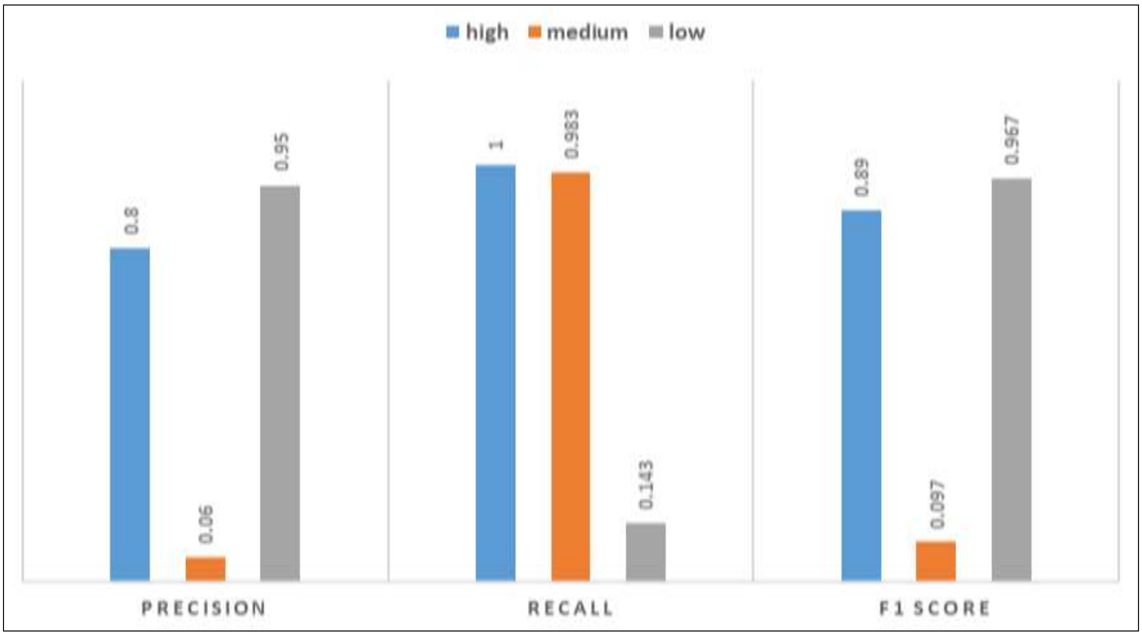

Fig 5: Shows the accuracy, recall, and f1 score values

Appearance. (5) Shows the accuracy, recall, and f1-score values of the model, before performing the above-mentioned optimization operations.

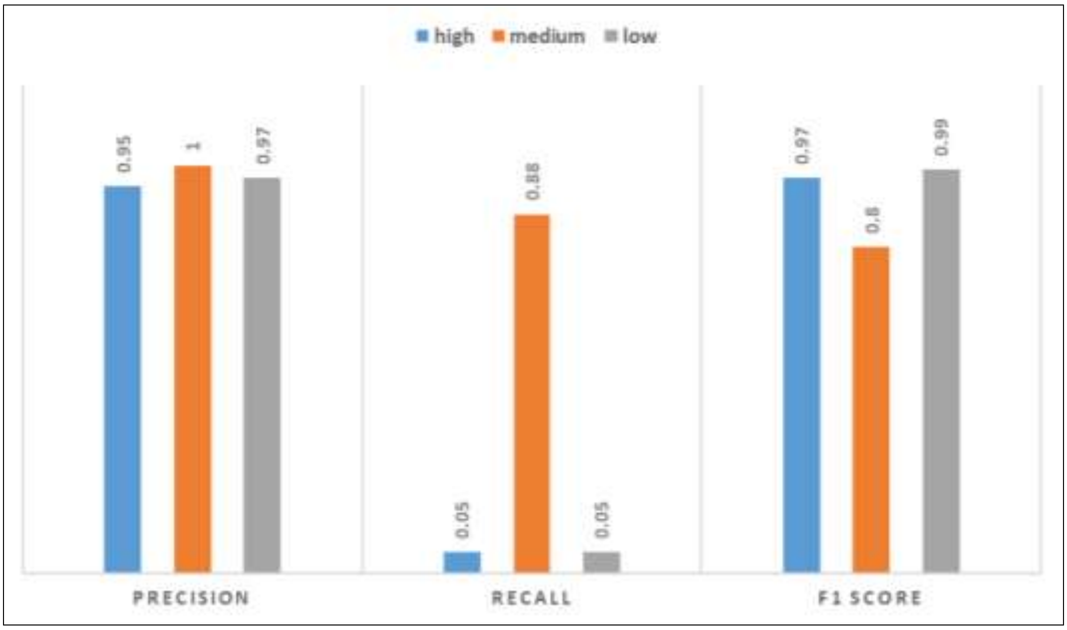

Fig 6: Shows the accuracy, recall, and f1 score values after improving

Appearance. (6) shows the accuracy, recall, and f1-score values of the model, before performing the above-mentioned optimization operations.

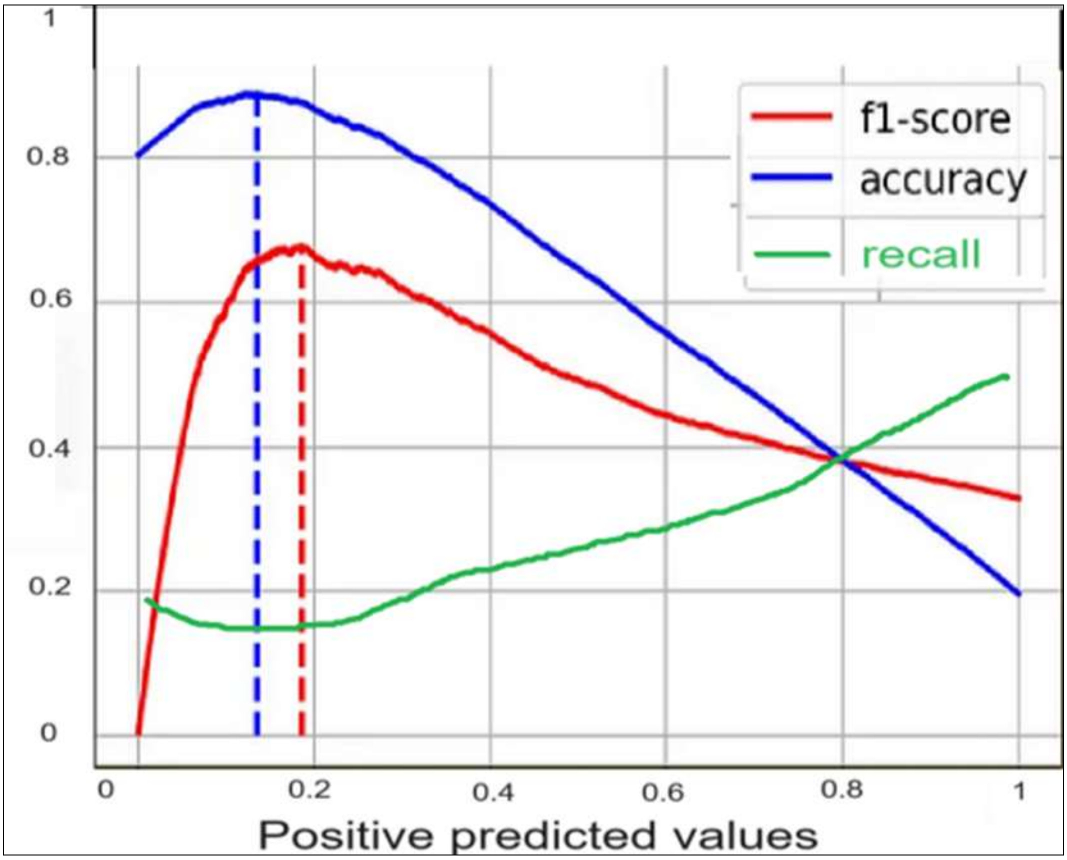

Fig 7: Shows the relation between the accuracy, recall, and f1 score

According to Figure 7, it is clear that the relationship between accuracy and recovery is an inverse relationship, and that the relationship is a direct relationship between F-Score and accuracy.

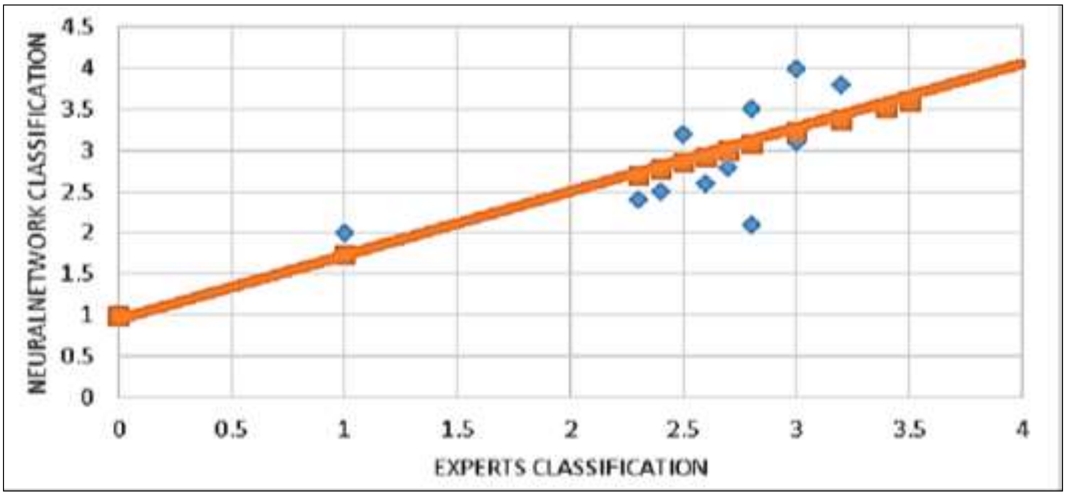

Fig 8: Shows the relation between neural network classification experts classification

Figure No. 8 shows the classification of the relationship between the experts' classification and the neural network model's classification, Figure 8 shows the classification relationship between expert classification and neural network model classification. As is clear from the diagram, neural network classifications tend to be consistent with expert classifications.

4.2 Discussion

Through Figure (4), the confusion matrix of the model shows the extent of the success of the model and achieving an accuracy rate of 91.6%. The accuracy of the model also increased after optimization operations, to 96.6%.

Figure (5) shows the values of accuracy, recall, and f1 score. The values show the success of the model, as it achieved an accuracy rate of 91%, and the recall values are, f1-score Very acceptable, but the model needs improvement. Figure (6) shows the values of recall, f1-score, and accuracy after improvement using parameter resetting methods. The values have improved significantly, as the accuracy of the model reached 96.4%.

Figure (7) shows the relationship between recall, which can be defined as the ratio of positive samples that were correctly classified to all positive samples already present in the data. While Precision can be defined as the ratio of samples that were correctly classified as positive to all samples that were classified as positive [18]. It turns out that the inverse relationship is clear, as when we try to increase recall, we usually expand the range in which we consider predictions to be positive, which leads to an increase in Number of true positives detected. This may also lead to an increase in the number of false positives, and thus a decrease in accuracy. As for the relationship between accuracy and F1 score, it is a direct relationship [19].

Figure No. (8), shows that, it is clear that the relationship between the classification of experts and the classification of neural networks takes a positive direction, meaning that the classifications of neural networks tend to be consistent with the classifications of experts. However, there are some discrepancies and some points show a significant deviation, which indicates that the neural network may Some samples were rated differently than experts [20].

5. Conclusion

Through the study, some conclusions were reached, the most important of which is that machine learning techniques are very effective tools in classifying water quality in particular, and in many applications in general, as they save a lot of time, effort and money, and they also give very accurate results and are flexible tools. Very easy, and the greater the size of the data and the greater its diversity, the more accurate and better the results will be [21]. It was also concluded that the factors that affect the choice of algorithms or machine learning techniques must be taken into account in terms of the size of the data, in terms of the diversity of the data, and in terms of the main problem that is to be solved using machine learning techniques, as there are techniques that are useful for applications that have a lot of and diverse data, such as There are applications for neural network algorithms for data of small size and low diversity. Also, the nature of the problem in terms of regression and in terms of grouping or classification is an important factor in determining the appropriate techniques for the applications. It is also possible to combine more than one algorithm into one model to benefit from the advantages of each algorithm, but the type [22], importance and nature of the application must be taken into account. It is also necessary to take into account the steps of designing a model using machine learning techniques and implementing them with all precision, especially with regard to the issue of monitoring the model, as monitoring the model, modifying its data, and improving it are among the most important factors for the success of the model and ensuring the continuity of this model. It can also be concluded that water quality classification is one of the most important factors that can achieve Sustainability in its full form in the water sector, whether environmental sustainability, social sustainability, or economic sustainability [23].

6. Recommendations

There are some important proposals, such as enhancing the role of water classification in achieving sustainability in the field of water, enhancing the role of machine learning techniques in classifying water quality in particular, and enhancing the role of machine learning techniques in all fields in general. There is also the need to make more efforts to spread and define the importance of using machine learning techniques. More efforts should be made in developing these techniques and discovering new algorithms that can improve the accuracy of the results, improve the efficiency of training the model, and the efficiency of learning the model to link outputs and inputs.

7. References

1. Nasir N, Kansal A, Alshaltone O, Barneih F, Sameer M, Shanableh A, *et al*. Water quality classification using machine learning algorithms. Journal of Water Process Engineering. 2022;48:102920.
2. Abuzir SY, Abuzir YS. Machine learning for water quality classification. Water Quality Research Journal. 2022;57(3):152-164.
3. Azhar SC, Aris AZ, Yusoff MK, Ramli MF, Juahir H. Classification of river water quality using multivariate analysis. Procedia Environmental Sciences. 2015;30:79-84.
4. Barinova S. On the classification of water quality from an ecological point of view. International Journal of Environmental Sciences & Natural Resources. 2017;2(2):1-8.
5. Mahesh B. Machine learning algorithms-a review. International Journal of Science and Research (IJSR). 2020;9(1):381-386.
6. Angra S, Ahuja S. Machine learning and its applications: A review. In: 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC); c2017 Mar. p. 57-60.
7. Malhotra R. A systematic review of machine learning techniques for software fault prediction. Applied Soft Computing. 2015;27:504-518.
8. Dhall D, Kaur R, Juneja M. Machine learning: A review of the algorithms and its applications. Proceedings of ICRIC 2019: Recent innovations in computing; c2020. p. 47-63.
9. Gautam R, Sharma M. Prevalence and diagnosis of neurological disorders using different deep learning techniques: A meta-analysis. Journal of Medical Systems. 2020;44(2):49.
10. Mishra S, Mishra D, Santra GH. Applications of machine learning techniques in agricultural crop production: A review paper. Indian Journal of Science and Technology.
11. Kadhim AI. Survey on supervised machine learning techniques for automatic text classification. Artificial Intelligence Review. 2019;52(1):273-292.
12. Poornappriya TS, Gopinath R. Application of machine learning techniques for improving learning disabilities. International Journal of Electrical Engineering and Technology (IJEET). 2020;11(10):392-402.
13. Abdolrasol MG, Hussain SS, Ustun TS, Sarker MR, Hannan MA, Mohamed R, *et al*. Artificial neural networks-based optimization techniques: A review. Electronics. 2021;10(21):2689.
14. Belkacem AN, Jamil N, Khalid S, Alnajjar F. On closed-loop brain stimulation systems for improving the quality of life of patients with neurological disorders. Frontiers in Human Neuroscience. 2023;17:1085173.
15. Nagabushanam P, Thomas George S, Radha S. EEG signal classification using LSTM and improved neural network algorithms. Soft Computing. 2020;24(13):9981-10003.
16. Uddin MG, Nash S, Rahman A, Olbert AI. Performance analysis of the water quality index model for predicting water state using machine learning techniques. Process Safety and Environmental Protection. 2023;169:808-828.
17. Ahmed U, Mumtaz R, Anwar H, Shah AA, Irfan R, García-Nieto J, *et al*. Efficient water quality prediction using supervised machine learning. Water. 2019;11(11):2210.
18. Kaur A, Khurana M, Kaur P, Kaur M. Classification and analysis of water quality using machine learning algorithms. In: Proceedings of International Conference on Communication, Circuits, and Systems: IC3S 2020. Springer Singapore; c2021. p. 389-398.
19. Ibrahim I, Abdulazeez A. The role of machine learning algorithms for diagnosing diseases. Journal of Applied Science and Technology Trends. 2021;2(01):10-19.
20. Mijwil M, Salem IE, Ismaeel MM. The significance of machine learning and deep learning techniques in cybersecurity: A comprehensive review. Iraqi Journal for Computer Science and Mathematics. 2023;4(1):87-101.
21. Sarker IH. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science. 2021;2(6):420.
22. Torney CJ, Lloyd-Jones DJ, Chevallier M, Moyer DC, Maliti HT, Mwita M, *et al*. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. Methods in Ecology and Evolution. 2019;10(6):779-787.
23. Dezfooli D, Hosseini-Moghari SM, Ebrahimi K, Araghinejad S. Classification of water quality status based on minimum quality parameters: Application of machine learning techniques. Modeling Earth Systems and Environment. 2018;4:311-324.