# Assessment of River Suitability for Aquaculture Using Random Forest Algorithm

P.S.Madhumitaa
*Department of Computer Science & Engineering*
*National Institute of Technology*
Puducherry, India
madhumitaa2005@gmail.com

M.Srimahalakshmi
*Department of Computer Science & Engineering*
*National Institute of Technology*
Puducherry, India
srimahalakshmi62004@gmail.com

V.Varshini
*Department of Computer Science & Engineering*
*National Institute of Technology*
Puducherry, India
varshu2025@gmail.com

Dr. Venkatesan M
*Assistant Professor*
*HOD – Department of Computer Science & Engineering*
*National Institute of Technology*
Puducherry, India

Dr. P Prabhavathy
*Department of Computer Science & Engineering*
*Vellore Institute of Technology*
Vellore, India
pprabavathy@vit.ac.in

***Abstract*—In this paper, we examine a dataset of water samples carefully collected from rivers across India's vast landscape. Our unwavering commitment revolves around the strategic effective management of water resources which is crucial in order to quickly identify the changes in water quality. In this expanding industry of aquaculture where water quality affects the vitality of fishes, by taking a proactive approach, it is possible to mitigate environmental risks in a timely manner, which in turn reduces the potential for economic losses. In order to comprehensively assess water quality, we scrutinize a range of pivotal parameters such as the water's pH, biochemical oxygen demand (BOD), dissolved oxygen (DO), electrical conductivity (EC), nitrate nitrogen (NO3-N), and ammonium nitrogen (NH3-N). These parameters shed light on the water's acidity or alkalinity, ability to support aquatic life and nutrient content. Various machine learning algorithms were considered for this research, including the Long Short Term Memory (LSTM), Support Vector Machine (SVM), Least Squares Support Vector Machine (LSSVM), and Gradient Boosting Machine (GBM). However, the Random Forest Classifier was identified as the most accurate algorithm for water quality detection. In addition to these algorithms, other machine learning techniques were also explored to enhance the water quality detection capabilities of this study, in accurately detecting and predicting variations and anomalies in these critical water quality parameters. The findings from our study clearly demonstrate the immense potential of these algorithms in revolutionizing water quality monitoring.**

***Keywords—Water quality index, Water quality prediction, river water quality degradation, Random forest.***

## 1. INTRODUCTION

Aquaculture is a crucial and expanding industry globally, providing a vital source of protein for human consumption. However, conventional fishing methods have their limitations, and to meet the increasing demand for fish products, there is a need for new techniques. Artificial Intelligence (AI) and Machine Learning (ML) are the latest technologies that can transform the Pisciculture industry [1].

Farmers can benefit greatly from precise and real-time predictions of the water quality parameters, as it allows for necessary adjustments to be made in advance. This can help improve the breeding environment, leading to increased fish production efficiency. By predicting and adjusting water quality, drug use can be reduced for green and precision agriculture [2].

In today's world, water pollution caused by human activities is a significant issue affecting the quality of water. In the field of Aquaculture, cultivating in freshwater under controlled conditions and maintaining high water quality is crucial for effective farming of aquatic organisms fish [3]. Water quality indicates the attributes of a water source which will determine its effectiveness for a particular use [4].

Fish are completely connected to water and depend on this element for their most important functions, including respiration, nutrition, reproduction, and growth. In scenarios where water quality changes negatively, it casts a shadow of vulnerability on these aquatic populations, compromising their immune defenses. The amount of DO in the water is particularly important among the important parameters that significantly affect the health of fish. Deviations from the desired oxygen concentration can seriously inhibit fish growth and general well-being.

In addition, water with an elevated pH can induce elevated ammonia concentrations, which can damage the delicate gills and livers of the fish. Another dangerous consequence of declining water quality is gas oversaturation, a condition that can lead to gas bubble disease, characterized by the formation of bubbles in various parts of the fish's anatomy, including the eyes, skin and gills. The consequences of water pollution due to the deterioration of water quality are diverse and include many problems that impair the growth and health of aquatic species [5].

Given the significant advances in machine learning techniques and artificial intelligence in solving complex classification and decision-making problems, our task is to exploit the potential of these intelligent algorithms. Their ability to learn from real data sets and accurately detect deviations from desired settings is a testament to their effectiveness. Motivated by these, our research focuses on using machine learning methods, especially the Random Forest Classifier to predict water quality by looking at several parameters that govern it. With this approach, we aim not only to increase our understanding of water quality dynamics, but also to protect the integrity of precious water resources in a proactive, sustainable and thoughtful way.

## 2. LITERATURE REVIEW

In 2020, the study was conducted at a marine aquaculture base in Xincun Town, China, the researchers collected data to predict water quality using several approaches. These approaches included the time series method, Markov method, grey system theory method and a few. Among them, a model called Bi-S-SRU (bi-directional simple recurrent unit) was constructed using pre-processed data and its correlation information. This model demonstrated good robustness, fault tolerance and was able to fit complex nonlinear relations well. Despite these strengths, the study also identified some drawbacks of these methods. These include low generalization, low computational efficiency and unstable prediction accuracy. Despite these limitations, the results showed that the Bi-S-SRU-based prediction method had only slightly higher time complexity than the traditional methods. Thus, the Bi-S-SRU model showed promise in predicting water quality with good robustness and high fault tolerance, it also highlighted areas for improvement such as enhancing generalization ability, increasing computational efficiency, and stabilizing prediction accuracy [6].

This study focuses on the application of Recirculating Aquaculture Systems (RAS) in maintaining stable water quality parameters to ensure suitable conditions for fish. The study emphasizes the role of RAS in improving water purification technology, equipment, and promoting sustainability. However, it also acknowledges that RAS can be energy and cost-intensive, leading to suggestions for economic improvements. It uses the new Internet of Things (IoT) based monitoring systems that can monitor basic water quality parameters. These systems provide sufficient data for regular monitoring purposes. The study presents a hybrid neural network approach for predicting and controlling water quality in RAS, offering a promising solution to some of the challenges faced in aquaculture [7].

This study conducted an investigation on water quality in the Macuco and Queixada rivers. They developed a water quality index based on the measurements of turbidity, total phosphorous, and dissolved oxygen. This was applied to 17 monitoring points over a two-year period to assess the impact of water quality on aquaculture. The index provides a simple and cost-effective method to assess the impact of aquaculture on water bodies. However, one of the limitations of this study is that the water quality index is calculated based on only three environmental parameters, which may not capture other environmental parameters that affect water quality. The researchers suggest that further research is necessary to refine this index and possibly include more parameters. Despite this limitation, the study provides insights into the use of water quality index as an indicator of the effects of aquaculture on aquatic bodies [8].

In this study, researchers explored various machine-learning techniques used for water-quality management aides. While Artificial Neural Networks (ANNs) are commonly used, the study also examined other methods like Adaptive Neuro Fuzzy Inference System (ANFIS), Recurrent Neural Networks (RNN), Extreme Machine Learning (EML), Regression Trees (RT), Support Vector Machines (SVM), Holt-Winters (HW), and hybrid models. The paper defines Artificial Intelligence (AI) as algorithms performing tasks requiring human intelligence and Machine Learning (ML) as a subset of AI where intelligent computer systems modify their behavior based on new data during training. The study emphasizes the importance of data management, public and legal opinions, reproducibility, transparency in research for advancing intelligent applications. It also discusses the concept of transfer learning, which mimics human learning by using past knowledge to improveperformance on new tasks. Despite the complexity of these techniques, the study underscores their potential in enhancing water treatment and monitoring processes[10].

In this research paper, the researchers conducted a study on water quality prediction in fish farming using data collected from a pond in Colombia. They used machine learning techniques such as random forest, multivariate linear regression, and artificial neural networks to predict water quality parameters. The researchers used mobile-based information systems to run their machine learning model, which allowed fish farmers to receive real-time predictions and make immediate decisions about their farming practices. However, this study had a limitation in that the water-quality variables were measured manually, and any errors in these measurements could impact the model's performance. Despite this limitation, the study provides valuable insights into the use of ML for water quality prediction in fish farming and suggests that future research will focus on creating a better prediction model, especially where data is collected manually [11].

## 3. METHODOLOGY

### *[3.1]Water Quality Prediction Model*

The dataset has captured critical water quality parameters, such as pH (ph), total coliform (tc), conductivity (co), dissolved oxygen (do), biological oxygen demand (bod), and nitrate content (na). These parameters have been meticulously collected and subsequently normalized, resulting in standardized values. Normalization is essential to ensure that each parameter contributes equally to the overall Water Quality Index (WQI) computation, irrespective of their original measurement units.

Furthermore, the normalized parameters are assigned individual weights, reflecting their relative importance in influencing overall water quality. This step is crucial as it recognizes that not all parameters have the same significance in determining water quality. By applying specific weights to each parameter, the importance of tailoring the WQI calculation to capture the unique characteristics of water bodies in India is highlighted.

The central focus lies in the aggregation process, where the weighted parameters are summated to compute the Water Quality Index (WQI). This aggregated value, wqi serves as a comprehensive measure of water quality.

The WQI allows for the classification of rivers in India into categories such as "poor," "fair," and "good." This classification is a vital step in assessing the suitability of these water bodies for aquaculture, which has significant implications for environmental management and the sustainable development of the aquaculture industry in India.

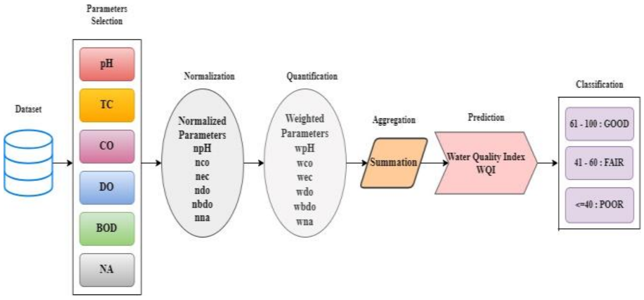


Fig. 1. Architecture

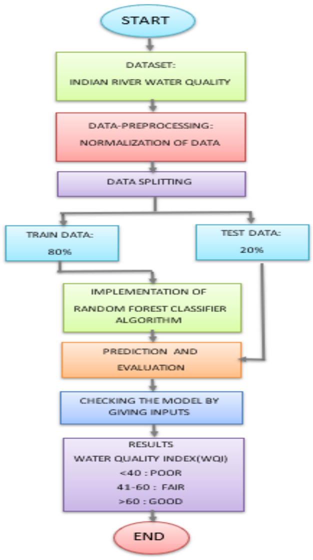


Fig. 2. Work Flow of the Proposed System

### *[3.2]Working Principle of Proposed System*

The Proposed system begins with the collection of Dataset from various rivers in India. This Dataset contains State, Locations and measurements of the various parameters to predict and analyze the Water quality.

Then, it undergoes the Preprocessing Stage. This Stage involves Normalization and Data Splitting. Normalization is a technique that organizes the data efficiently, the main goal of normalization is to transform the values of the numeric columns in the dataset to a common scale without distorting differences in the ranges of the values or losing information. After Normalization, the data is split into training and testing sets, where 80% of the data is used for training the model and the remaining 20% is used for testing its accuracy.

After the Preprocessing Stage, the next step involves the implementation of a Random Forest Algorithm. This Machine Learning Algorithm is known for its efficiency and accuracy. The Random Forest classifier takes the input derived from the training dataset. It creates a collection of decision trees and combines their output to make decisions and predict the water Quality.

The Final step involves checking the results. The model predictions are compared with the actual Water quality index (WQI) Ranges. If the WQI prediction value is less than 40, then the water quality is considered as "poor". If the WQI prediction value is between 41 to 60, then the water quality is considered as "Fair" and if the WQI prediction value is greater than 60, then the water quality is considered as "Good".

### *[3.3] Random Forest Classifier*

The Random Forest Classifier Technique:

The Random Forest Classifier is a powerful machine-learning algorithm used for classification and regression tasks. It belongs to the ensemble learning family of algorithms, which means it combines the predictions from multiple machine learning models to make more accurate and robust predictions.

---

**Algorithm 1** Random Forest algorithm

---

**INPUT:**

i. Input dataset (features and labels): $(x, y)i$=1 to $N$
ii. Number of trees in the forest: $M$
iii. Splitting criterion for tree construction
iv. Maximum tree depth or stopping criteria
v. Number of features to consider at each split
vi. Other hyperparameters (e.g, minimum samples per leaf)

**OUTPUT:** Ensemble of trees

Initialize an empty forest $F$.

**for** $t$ = 1 to $M$ **do**

Create a bootstrapped dataset by randomly sampling $N$ samples with replacement from the original dataset $(x, y)$.

Create a decision tree $Tt$:

At each node of the tree, randomly select a subset of features from the available features.

Choose the best feature and split point based on a criterion (e.g., Gini impurity or entropy) and construct the node.

Continue splitting nodes until a stopping criterion is met (e.g., maximum depth or minimum samples per leaf).

Add the tree $Tt$ to the forest $F$.

**end for**

---

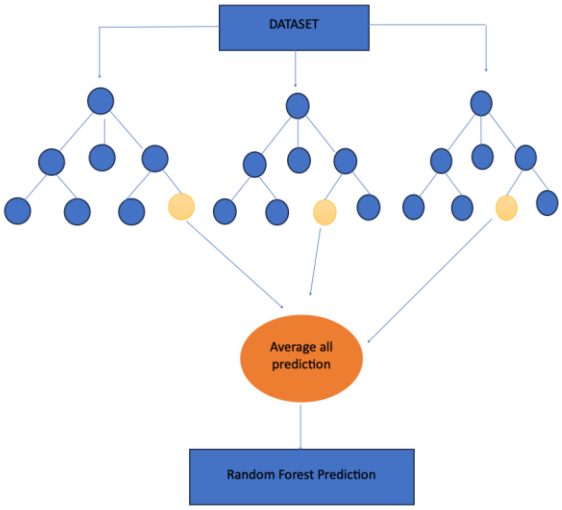


Fig. 3. Working Procedure of Random Forest

The random forest algorithm details used in the simulation are as shown in Table I:

TABLE I. RANDOM FOREST ALGORITHM DETAILS

| NAME OF PARAMETER | VALUE |
|---|---|
| Number of trees in the forest | 10 |
| min_samples_split | 2 |
| min_samples_leaf | 1 |
| random_state | 0 |

Table II shows the model parameters used in this paper for the calculation and prediction of water quality index.

TABLE II. MODEL PARAMETERS

| PARAMETER NAME | TYPE | UNIT |
|---|---|---|
| STATION CODE | object | |
| LOCATION | object | |
| STATE | object | |
| Temp | float64 | |
| D.O. | float64 | mg/l |
| pH | float64 | |
| CONDUCTIVITY | Object | mhons/cm |
| B.O.D. | float64 | mg/l |
| NITRATENAN N+ NITRITENANN | float64 | mg/l |
| FECAL COLIFORM | Object | MPN/100ml |
| TOTAL COLIFORM | float64 | (MPN/100ml) Mean |
| Year | int64 | |

The Temp parameter represents the temperature of the water body. Information such as station code, location and state is given to identify the sampling location. Dissolved Oxygen (D.O.) is a critical measure of the free oxygen in water, impacting aquatic life and water quality. Deviations from optimal D.O. levels can harm aquatic ecosystems. pH signifies the acidity or alkalinity of water, influencing the solubility and biological impact of contaminants like nutrients and heavy metals. Conductivity assesses the water's electrical conductivity, reflecting the presence of dissolved ions. Biochemical Oxygen Demand (BOD) quantifies the oxygen consumed by aerobic microorganisms during organic matter decomposition in water. High BOD values signal the presence of organic matter, which can deplete oxygen levels, affecting aquatic life. Nitrate is a compound produced by combining oxygen or ozone with nitrogen. While nitrogen is essential for all life, excessive nitrate levels can be detrimental to aquatic organisms. Coliform Bacteria are found in animals, including humans. Although most coliforms are harmless, certain strains like E.coli can indicate faecal contamination and the potential presence of harmful pathogens in water.

## 4. COMPARATIVE ANALYSIS

TABLE III. COMPARISON OF ML MODELS USING R2 VALUE

| ML MODEL | MAE | MSE | RMSE | R2 |
|---|---|---|---|---|
| RFC | 1.0063408521303325 | 6.195191609022559 | 2.4890141841746423 | 0.9662300204628745 |
| LSTM | 1.2023489567825623 | 7.556815642789550 | 2.7254951354687125 | 0.9446213548795617 |
| SVM | 1.1002157856413648 | 6.813254978432135 | 2.6454623666367849 | 0.9623449873564125 |
| LSSVM | 1.3456136674213578 | 8.213568413236578 | 2.9123584659874626 | 0.9315844596668561 |
| GBM | 1.0582589546879846 | 6.012478943154517 | 2.4875945230058458 | 0.9613589500984640 |

Table 3 presents the findings of our experiment, where we evaluated the performance of different classifiers. The results indicate that the Random Forest Classifier (RFC) performed exceptionally well in our analysis. In general, lower Mean Absolute Error (MAE) values are better, and the RFC has the lowest MAE (1.0063), indicating that it makes the smallest absolute errors on average. Similarly, lower Mean Squared Error (MSE) values are better, and although the Gradient Boosting Machine (GBM) has the lowest MSE (6.0124), the RFC (6.1952) is also relatively close to this value, indicating smaller squared errors compared to other models. Additionally, lower Root Mean Squared Error (RMSE) values are better, and the RFC once again has the lowest RMSE (2.4890), implying smaller errors on average. The R-squared (R2) value measures the proportion of the variance in the dependent variable that can be predicted from the independent variables. In general, higher R2 values are desirable, and the RFC has the highest R2 (0.9662), indicating that it explains a large portion of the variance in the data.

Our findings suggest that the Random Forest Classifier is the most accurate and reliable model for this particular experiment. Moreover, all the classifiers are highly efficient, providing fast prediction times, making them ideal for real-time applications that require immediate results. In conclusion, the results suggest that the Random Forest Classifier is a superior model that can produce accurate and efficient results across multiple applications.

## 5. WATER QUALITY INDEX ESTIMATION

Water quality is determined by a variable WQI which is constructed based on the normalized parameters nph, nbdo, nec, ndo, nbdo, nna and weighted parameters wph, wdo, wbdo, wec, wna and wco. These variables are used to perform mathematical calculations, as shown in Table IV

TABLE IV. NORMALIZATION OF VARIOUS PARAMETERS IN WATER

| PARAMETERS | PARAMETER RANGE | NORMALIZED VALUES | WEIGHTED VALUES |
|---|---|---|---|
| pH - npH - wpH | 7 to 8.5 | 100 | npH * 0.165 |
| | 8.5 to 8.6 or 6.8 to 6.9 | 80 | |
| | 8.6 to 8.8 or 6.7 to 6.8 | 60 | |
| | 8.8 to 9 or 6.5 to 6.7 | 40 | |
| | Else | 0 | |
| TC - nco - wco | 0 to 5 | 100 | nco ∗ 0.281 |
| | 5 to 50 | 80 | |
| | 50 to 500 | 60 | |
| | 500 to 1000 | 40 | |
| | Else | 0 | |
| CO - nec - wec | 0 to 75 | 100 | $n$ec ∗ 0.009 |
| | 75 to 150 | 80 | |
| | 150 to 225 | 60 | |
| | 225 to300 | 40 | |
| | Else | 0 | |
| DO - ndo - wdo | 6+ | 100 | $n$do ∗ 0.281 |
| | 5.1 to 6 | 80 | |
| | 4.1 to 5 | 60 | |
| | 3 to 4 | 40 | |
| | Else | 0 | |
| BOD - nbdo - wbdo | 0 to 3 | 100 | $n$bdo ∗ 0.234 |
| | 3 to 6 | 80 | |
| | 6 to 80 | 60 | |
| | 80 to 125 | 40 | |
| | Else | 0 | |
| NA - nna - wna | 0 to 20 | 100 | $n$na ∗ 0.028 |
| | 20 to 50 | 80 | |
| | 50 to 100 | 60 | |
| | 100 to 200 | 40 | |
| | Else | 0 | |

Then, wqi = wp$h$ + wco + wec + wdo + wbdo + wna WQI measured as shown here is used in the learning algorithm to predict the water quality based on the constituting parameters and classify the rivers based on their WQI. [5]

TABLE V. DATASET

| STATION CODE | LOCATIONS | STATE | Temp | D.O. (mg/l) | PH | CONDUCTIVITY (μmhos/cm) | B.O.D (mg/l) | NITRATENAN N+ NITRITENANN (MG/L) | FECAL COLIFORM (MPN/100ml) | TOTAL COLIFORM (MPN/100ml)Mean | year |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1393 | DAMANGANGA AT D/S OF MADHUBAN, DAMAN | DAMAN & DIU | 30.6 | 6.7 | 7.5 | 203 | NAN | 0.1 | 11 | 27 | 2014 |
| 1399 | ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOIN… | GOA | 29.8 | 5.7 | 7.2 | 189 | 2 | 0.2 | 4953 | 8391 | 2014 |
| 1475 | ZUARI AT PANCHAWADI | GOA | 29.5 | 6.3 | 6.9 | 179 | 1.7 | 0.1 | 3243 | 5330 | 2014 |
| 3181 | RIVER ZUARI AT BORIM BRIDGE | GOA | 29.7 | 5.8 | 6.9 | 64 | 3.8 | 0.5 | 5382 | 8443 | 2014 |
| 3182 | RIVER ZUARI AT MARCAIM JETTY | GOA | 29.5 | 5.8 | 7.3 | 83 | 1.9 | 0.4 | 3428 | 5500 | 2014 |

## 6. EXPERIMENTAL SETUP AND TRAINING

The Random Forest library was installed and utilized to train a Random Forest Model. To effectively operate the Random Forest model, specific parameters and computations are essential, as outlined in the methodology. These parameters are employed to generate predictor variables, and if response variables are not explicitly defined, the model makes educated guesses. Subsequently, the dataset is selected for analysis. We have taken a clean and well pre-processed dataset. This dataset includes relevant water quality features and the WQI values. Random Forest, which is an ensemble of decision trees, also requires specifying the number of trees (n-trees) as a parameter. Dataset is split into a training set and a testing set, 80% for training and 20% for testing. Tools like scikit-learn in Python to achieve this split. Once these configurations are set, the training process commences. Following the training and testing phases, feature selection/engineering is done to identify the most important features. Then the model generates results and predictions. The model's performance is evaluated on the testing set using appropriate evaluation metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score.

## 7. RESULT AND DISCUSSION

This Research uses the powerful Random Forest model, a sophisticated ensemble learning algorithm that plays a central role in assessing river suitability for aquaculture based on the Water Quality Index (WQI) Ranges. The Random Forest algorithm lies in its remarkable ability to construct a group of decision trees during each training cycle. From the decision trees , the Random Forest model make prediction and classify the river suitability for aquaculture based on diverse water quality parameters defined by the Water quality Index Ranges.

TABLE VI. DATA ANALYSIS

| | station | location | state | Temp | do | Ph | co | bod | na | tc | ... | nbdo | nec | nna | wph | wdo | wec | wna | wco | wqi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1393 | DAMAGAN AT D/S OF MADHUBAN, DAMAN | DAMAN & DIU | 30.600000 | 6.7 | 7.5 | 203.0 | 6.940049 | 0.100000 | 27.0 | … | 60 | 60 | 100 | 16.5 | 28.10 | 0.54 | 2.8 | 22.48 | 84.46 |
| 1 | 1399 | ZUARI AT D/S OF PT,WHERE KUMBARJRIA CANAL JOL…. | GOA | 29.800000 | 5.7 | 7.2 | 189.0 | 2.000000 | 0.200000 | 8391.0 | … | 100 | 60 | 100 | 16.5 | 22.48 | 0.54 | 2.8 | 11.24 | 76.96 |
| 2 | 1475 | ZUARI AT | GOA | 29.500000 | 6.3 | 6.9 | 179.0 | 1.700000 | 0.100000 | 5330.0 | … | 100 | 60 | 100 | 13.2 | 22.48 | 0.54 | 2.8 | 11.24 | 79.28 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PANCHAWADI | | | | | | | | | | | | | | | | | | |
| 3 | 3181 | RIVER ZUARI AT BORIM BRIDGE | GOA | 29.700000 | 5.8 | 6.9 | 64.0 | 3.800000 | 0.500000 | 8443.0 | … | 80 | 100 | 100 | 13.2 | 22.48 | 0.90 | 2.8 | 11.24 | 69.34 |
| 4 | 3182 | RIVER ZUARI AT MARCAIM JETTY | GOA | 29.500000 | 5.8 | 7.3 | 83.0 | 1.900000 | 0.400000 | 5500.0 | … | 100 | 80 | 100 | 16.5 | 22.48 | 0.72 | 2.8 | 11.24 | 77.14 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| 1986 | 1330 | TAMBIRAPARANI AT ARUMUGANERI, TAMILNADU | NAN | 26.209814 | 7.9 | 738.0 | 7.2 | 2.700000 | 0.518000 | 202.0 | … | 100 | 100 | 100 | 0.0 | 28.10 | 0.90 | 2.8 | 16.86 | 72.06 |
| 1987 | 1450 | PALAR AT VANIYAMBADI WATER SUPPLY HEAD WORK,T… | NAN | 29.000000 | 7.5 | 585.0 | 6.3 | 2.600000 | 0.155000 | 315.0 | … | 100 | 100 | 100 | 0.0 | 28.10 | 0.90 | 2.8 | 16.86 | 72.06 |
| 1988 | 1403 | GUMI AT U/S SOUTH TRIPURA, TRIPURA | NAN | 28.000000 | 7.6 | 98.0 | 6.2 | 1.200000 | 1.623079 | 570.0 | … | 100 | 100 | 100 | 0.0 | 28.10 | 0.90 | 2.8 | 11.24 | 66.44 |
| 1989 | 1404 | GUMI AT U/S SOUTH TRIPURA, TRIPURA | NAN | 28.000000 | 7.7 | 91.0 | 6.5 | 1.300000 | 1.623079 | 562.0 | … | 100 | 100 | 100 | 0.0 | 28.10 | 0.90 | 2.8 | 11.24 | 66.44 |
| 1990 | 1726 | CHANDRAPUR, AGARTALA D/S OF HAORA RIVER… | NAN | 29.000000 | 7.6 | 110.0 | 5.7 | 1.100000 | 1.623079 | 546.0 | … | 100 | 100 | 100 | 0.0 | 28.10 | 0.90 | 2.8 | 11.24 | 66.44 |

Rivers that fall within the WQI range of 61 to 100 are accorded the highly coveted designation of "Good." This classification also recognizes the river is "Very suitable without restrictions" for a broad spectrum of aquaculture endeavours. On the other hand, for rivers that reside within the WQI range of 41 to 60, the classification gracefully transitions to "Fair", also it classifies the river as Relatively suitable conditions with certain restrictions" for aquaculture activities. Meanwhile, in the final echelon of this comprehensive classification scheme, rivers with WQI values ranging between 0 and 40 find themselves recognised as "Poor", it also recognizes that the river is "Unsuitable water conditions for aquaculture activity". In summation, the rivers can be classified based on this category which is used to predict the suitable river for Aquaculture.

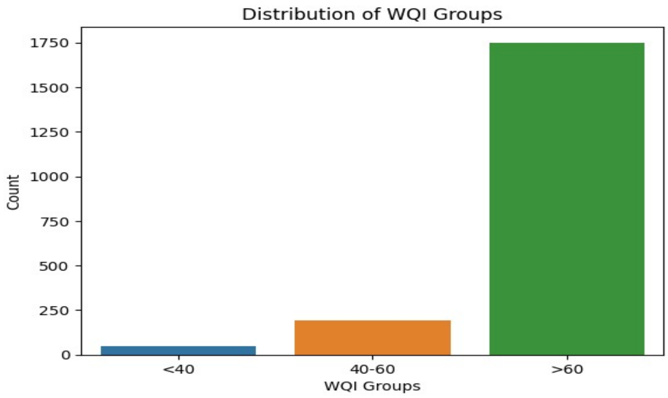


Fig. 4. Classification of Rivers

.The graph above depicts the Water Quality Index (WQI) ranges used in our aquaculture suitability assessment model. The WQI serves as a crucial indicator of water quality, which is a fundamental factor in determining the suitability of rivers for aquaculture activities. As shown in the graph, we have categorized the WQI into three distinct ranges as shown in table VI.

TABLE VII. WQI INDEX

| **Range score** | **Status** | **Description** |
|---|---|---|
| 61-100 | Good | Very suitable without restrictions |
| 41-60 | Fair | Relatively suitable with restrictions |
| 0-40 | Poor | Unsuitable water condition for aquaculture activity |

## 8. CONCLUSION

The main goal of this research is to identify Rivers that would be most suitable for aquaculture. For this research, a large amount of water quality data was collected from various river sources in India. This large data collection helps us to understand the variations in water quality parameters across different regions of India. In this Research, we compared various Machine Learning Models. In this comparison,the Random Forest Classifier (RFC) emerged as the most effective machine learning algorithm for this task due to its robustness, ability to handle complex datasets, and outstanding predictive performance. The Random Forest Classifier (RFC) emerges as the top-performing model, showcasing superior accuracy with lower MAE, MSE, RMSE, and a higher R2 value compared to LSTM, SVM, LSSVM, and GBM. Random Forest algorithm consistently demonstrated its remarkable performance, delivering predictions of water quality that can only be described as outstanding. The model's ability to accurately predict the WQI Range is the power of the Random Forest machine learning algorithm. The predicted WQI value by the model is compared with the Actual WQI Range based on the comparison the suitability of these rivers for aquaculture was predicted and classified. This classification provides valuable insights for decision makers and stakeholders in the aquaculture industry .It can guide them in selecting the suitable location for aquaculture.

In future work, we propose developing a web-based application that allows users to input their own water quality data and receive a prediction of river suitability for aquaculture. This would make the research more accessible to a wider audience and could be used to support decision-making in the aquaculture sector.

## REFERENCES


[1] Dr. Rajesh Kumar Panda & Prof. Dipak Baral. (2023). Adoption of AI/ML in Aquaculture: a study on Pisciculture. https://doi.org/10.17762/sfs.v10i1.473

[2] Sen, Sohom & Maiti, Samaresh & Manna, Sumanta & Roy, Bibaswan & GHOSH, ANKIT.(2023) Smart Prediction of Water Quality System for Aquaculture using Machine Learning Algorithms. 10.36227/techrxiv.22300435.

[3] Li, Tingting & Lu, Jian & Wu, Jun & Zhang, Zhenhua & Chen, Liwei. (2022). Predicting Aquaculture Water Quality Using Machine Learning Approaches. Water. 14. 2836. 10.3390/w14182836.

[4] Nayan, Al-Akhir & Kibria, Muhammad & Rahman, Md & Saha, Joyeta. (2020). River Water Quality Analysis and Prediction Using GBM. 219-224..

[5] Nayan, A.-A. ., Saha, J. ., Mozumder, A. N. ., Mahmud, K. R. ., Al Azad, A. K. ., & Kibria, M. G. . (2021). A Machine Learning Approach for Early Detection of Fish Diseases by Analyzing Water Quality. Trends in Sciences, 18(21), 351.

[6] Liu, Juntao & Yu, Chuang & Zhuhua, Hu & Zhao, Yaochi & Bai, Yong & Xie, Mingshan & Luo, Jian. (2020). Accurate Prediction Scheme of Water Quality in Smart Mariculture with Deep Bi-S-SRU Learning Network. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2971253.

[7] Junchao Yang, Lulu Jia, Zhiwei Guo, Yu Shen, Xianwei Li, Zhenping Mou, Keping Yu, Jerry Chun-Wei Lin, Prediction and control of water quality in Recirculating Aquaculture System based on hybrid neural network, Engineering Applications of Artificial Intelligence, Volume 121, 2023, 106002, ISSN 0952-1976

[8] Simões, Fabiano & Moreira, Altair & Bisinoti, Márcia & Gimenez, Sonia & santos yabe, Maria. (2008). Water Quality Index as a Simple Indicator of Aquaculture Effects on Aquatic Bodies. Ecological Indicators. 8. 476-484. 10.1016/j.ecolind.2007.05.002.

[9] Chen, C.-H.; Wu, Y.-C.; Zhang, J.-X.; Chen, Y.-H. IoT-Based Fish Farm Water Quality Monitoring System. Sensors 2022, 22, 6700.

[10] Lowe M, Qin R, Mao X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. *Water*. 2022; 14(9) : 1384. https://doi.org/10.3390/w14091384

[11] Zambrano AF, Giraldo LF, Quimbayo J, Medina B, Castillo (2021) Machine learning for manually-measured water quality prediction in fish farming. PLoS ONE 16(8): e0256380. https://doi. org/10.1371/journal.pone.0256380

[12] Indian River Water Quality Dataset available at: https://www.kaggle.com/datasets/anbarivan/indian- water-quality-data/