2023 7th International Conference on Design Innovation for 3 Cs Compute Communicate Control (ICDI3C) | 979-8-3503-9555-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDI3C61568.2023.00040

Detecting Lake Water Quality Using Machine Learning and Image Processing

Dr. Utsab Banerjee.
Assistant Professor
Department of Electronics and Communication Engineering
MVJ College of Engineering
Bengaluru, India
utsabbanerjee@mvjce.edu.in

Dr. Vinod P V
RRSC-SOUTH
NRSC,ISRO,
Bangalore, India.
vinod.isro@gmail.com

Karuna Sharma
Department of Electronics and Communication Engineering
MVJ College of Engineering
Bengaluru, India
karunasharma21032002@gmail.com

Jasna Thomas
Department of Electronics and Communication Engineering
MVJ College of Engineering
Bengaluru, India
jasnathomas0709@gmail.com

Prithviraj Sawant
Department of Electronics and Communication Engineering
MVJ College of Engineering
Bengaluru, India
prithvi456t@gmail.com

Vineet Kumar Pandey
Department of Electronics and Communication Engineering
MVJ College of Engineering
Bengaluru, India
vineetvineet059@gmail.com

***Abstract* — Accurate water quality detection is the basis of water environment management and is of great significance for water environment protection. There is no doubt that the accuracy of water quality will be improved if the multivariate correlation and time sequence data of water quality are fully used. With respect to the lakes in Bengaluru, it is being alarmingly polluted due to drains opening into the water bodies. The lakes exhibit severe pollution with high values of alkalinity, hardness, phosphates and a low level of dissolved oxygen and other non-degradable substances. This paper proposes a framework for future enhancement for water quality detection using Machine learning [6][8] and Image Processing [7][13][14] which can significantly help in assessing water sources without the usage of traditional methodologies different sensors which are manual, time consuming, costly and lack real-time feedback. The methods for water quality detection using machine learning includes Multiple Linear Regression, Random Forest Regression, Support Vector Regression, Gradient Boosting Regression, Decision Tree Regression for detecting water quality index and uses Random Forest Classification, Support Vector Machine, Accuracy, Gradient Boosting Classification, Decision Tree Classification for classification purpose. These methods give a data-driven model utilized in surrogate modelling for mapping physical processes and obtaining the second degree of realism, primarily to reduce physical simulation time and expense. These algorithms can be employed in a variety of scientific domains since they have good generalization ability, high accuracy, are less prone to over-fitting, and simultaneously minimize estimate mistakes.**

Keywords: Water Quality Detection; Image Processing; Machine Learning; Real-Time sensing [5];

I. INTRODUCTION

Water quality is an important aspect of the health and well-being of human society and aquatic ecosystems. Poor water quality of lakes can lead to the degradation of aquatic habitats, the decline of aquatic species, and the spread of waterborne diseases which will affect human society. Consuming contaminated drinking water can lead to various health problems, such as gastrointestinal infections, typhoid fever, cholera, and dysentery. Exposure to toxins in water can cause skin irritation, respiratory problems, reproductive problems, cardiovascular problems, liver and kidney damage, and cancer
Water pollution can disturb the food chain, affect the respiratory system of fishes, and damage the crop production.
Water pollution can also spread diseases through pathogens, microplastics, and chemicals that affect the plants, animals, and seafood that humans consume and Water pollution can reduce the availability and quality of freshwater resources, leading to water stress and scarcity.
Monitoring water quality [1][3][4] of lakes is essential for a variety of reasons. Lakes are a vital resource for human society, agriculture, and industries. Ensuring that the lake water is safe and suitable for human use is critical for maintaining public health and important for sustaining economic development. Lakes are ecosystems that support diverse aquatic life, from fish and algae to microorganisms. Any change in water quality can have significant impacts on the survival of these aquatic organisms, which can, in turn, affect the overall ecological health of the lake. Lakes are sensitive to human activities and environmental changes such as Climate change, pollution, land-use changes, and other human impacts can all affect the water quality of lakes. Regular monitoring of water quality is therefore crucial for identifying and mitigating these impacts, and for maintaining the long-term health and sustainability of lake ecosystems. Monitoring water quality of lakes is essential and important for supporting aquatic life, protecting human health, ensuring recreational opportunities, and preserving the ecological health of lake ecosystems. A smart technology application is presented in this paper that aims to monitor and detect the quality of water using smart technologies. Using an on-site measurement technique with image processing that will provide real-time [5] estimation of water quality, as well as the detection of lake water quality using machine learning [6][8][5], as well as a website that is fully configurable for monitoring the data on a real-time basis.

Table 1. Classification based on WQI

WQI (Water Quality Index) Value	Water Quality Classification
Less than 50	Excellent
50-100	Good Water
100-200	Poor Water
200-300	Very Poor Water
>300	Unsuitable for drinking

979-8-3503-9555-6/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDI3C61568.2023.00040

Below shown are some statistics and indicators about water pollution in the world and in India:

Table 2. Statistics on water quality in India

Indicator	World	India
Population without safely managed sanitation services (%)	55	59.4
Population living in water-stressed countries (%)	31	100
Discharges of COVID-19-associated plastic waste to the ocean (million tons)	1.56	0.3
Surface water unfit for consumption (%)	N/A	70
Health costs relating to water pollution (billion USD per year)	N/A	6.7-8.7
Population affected by waterborne diseases (millions)	N/A	37.7
Children dying of diarrhoea (millions)	N/A	1.5
Water quality index rank (out of 122 countries)	N/A	120
Level of dissolved oxygen (mg/L)	N/A	4-8
Level of total organic carbon (mg/L)	N/A	2-10
Level of total phosphorus (mg/L)	N/A	0.01-0.1

Resources:
1.statistia.com 2.Weforum.org 3.indiawaterportal.org
4.worldbank.org 5.wbwaterdata.org 6. Insightsonindia.com

II. EXISTING METHODOLOGY

Below diagram shows the water quality assessment using physio-chemical and bacteriological parameters such as temperature, pH, dissolved oxygen, biochemical oxygen demand, total coliform, etc. This is the most common and conventional method used by various agencies such as CPCB, SPCBs, CWC, etc.

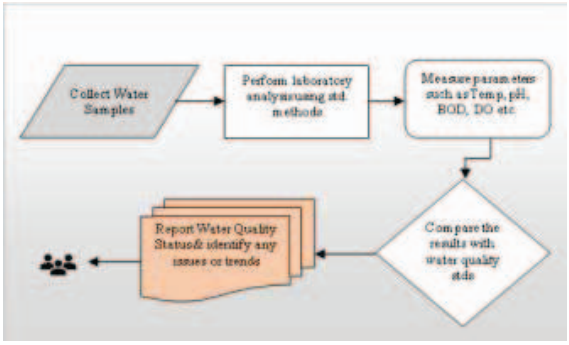

Fig 1. Conventional method Architecture

Some other existing methodologies for detecting lake water quality in India are as follow,

- Water quality assessment using biological indicators such as algae, macrophytes, zooplankton, fish, etc. This method is based on the principle that the presence and abundance of certain organisms reflect the ecological status and health of the lake.
- Water quality assessment using online water quality monitoring systems [1][3][4] that can provide real-time [5] data on various parameters using sensors and communication devices. This method is based on the principle that the continuous and automated measurement of water quality can provide more accurate and reliable information than discrete sampling and analysis.
- Water quality assessment using remote sensing and GIS based techniques that can provide spatial and temporal information on water quality using satellite imagery and geospatial analysis. This method is based on the principle that the spectral reflectance of water bodies can be related to their physical and chemical properties such as suspended solids, chlorophyll-a, dissolved organic matter, etc.

.

III. PROPOSED METHODOLOGY

The purpose of this paper is to develop an application capable of doing five major tasks listed below,

- Onsite Measurement at lake using sensors (TDS, Temperature, pH, Turbidity in our case)
- Live tracking of smart buoy/bot [10][11] using GPS
- To perform Image processing on collected water /lake images
- Applying Machine learning to detect WQI and classify water by feeding sensor values to model.
- Website development for above listed tasks

Smart buoy [10][11] makes use of sensors for data acquisition, 4G module for establishing communication between sensors and Website/Dashboard. This communication plays a vital role in case where bot [2] is located in some remote area. Smart buoy also uses ESP32 CAM to capture underwater images [14], these images can be processed and output can be displayed on website. Image processing [7][13] can help in getting values like Turbidity, Chlorophyll and can also provide information about biological indicators such as algae, macrophytes, zooplankton, fish etc. Moreover, biological indicators can provide information on the trophic state, nutrient enrichment, organic pollution, toxic substances, etc. of the lake.
Website deployed with Machine Learning [6][8][9][15] can detect WQI (Water Quality Index) and can classify water into following types that is Excellent, Good Water, Poor Water, Very Poor Water, Unsuitable depending on the values collected by sensors. Live dashboard can help user to visualize current values and historical values with timestamp.

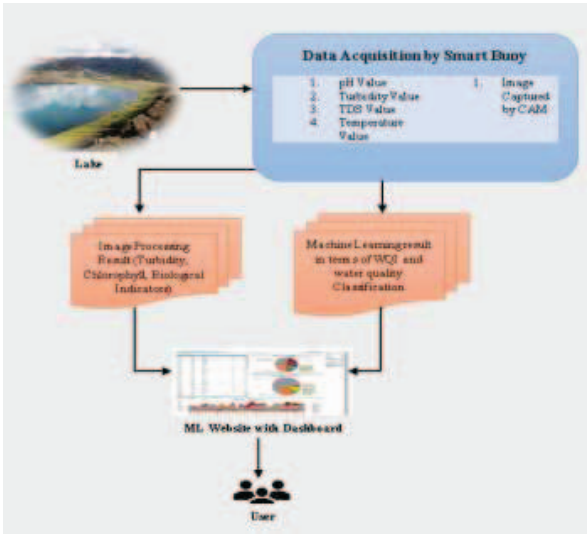

Fig 2. Proposed Architecture for our Application

A. Software Requirement

a. Backend - Flask, Stream lit
b. Database – MySQL
c. Programming Languages - Python, SQL
d. API - Google sheet API
e. Other Tools – GitHub, XAMPP

B. Data Acquisition

This is accomplished in following ways.

- Onsite:

The process of on-site measurement starts with an independent buoy that has essential sensors and a camera installed, which is deployed in a lake to gather information on different water quality factors which are pH, turbidity, Total Dissolved Solids (TDS), temperature, and capture images [14] of the underwater environment. The buoy's movement is tracked using GPS, and all the water quality related data is sent back in real-time [5] to the cloud via a communication module for monitoring purposes.
Onsite measured data will be utilized to detect water quality parameters, while the captured images will undergo image processing to monitor the water quality.
The data does not already include the information that we wish to anticipate. Therefore, we need what we want to train before we can develop a model to train with. We shall now calculate WQI using method Weighted Arithmetic Index. The World Health Organization (WHO) and Bureau of Indian Standards (BIS) drinking water quality standards were used to calculate the WQI. The Water Quality Index was computed by linearly averaging the quality rating and weight.

$$WQI=\sum (qn \times Wn)$$

Where,
qn =Quality rating for the nth Water quality parameter
Wn= unit weight for the nth parameter

C. Model Training

It is critical that the model be validated in real-time to ensure compatibility.
Thus, model for the regression job, which is to produce Water Quality index using sensor data, is trained using Machine Learning [9][15] algorithms such as Multiple Linear regression, Random Forest Regression, Support Vector regressor, Gradient Boosting Regression, Decision Tree Regression, and Deep Learning algorithm ANN (Artificial Neural Network).
A total of five major categories were developed based on the WQI (Water Quality Index) values of the samples: Excellent, Good, Poor, Very Poor, and Unsuitable.
The classification process was conducted using machine learning algorithms, such as Logistic Regression, Random Forest Classification, Support Vector Machines (SVMs), Gradient Boosting Classification and Decision Tree Classification.

Algorithm: Water Quality Prediction and Classification Algorithm

---

Input: Sensor Values
Output: Prediction of Water Quality Index (WQI) and Classification of samples.

- Collecting the water sample from different identified stations.
- Feature Engineering: This involved selecting the relevant features that would be used to train the machine learning model.
  - Exploratory data analysis (EDA) is performed, and it reveals that the features are largely unrelated.
  - Boxplot is analysed.
- We used following models:
  - For Detecting WQI:
    1. Multiple Linear Regression:
       Accuracy: 48%, Loss: 15.58
    2. Random Forest Regression:
       Accuracy: 96.91%, Loss:3.8
    3. Support Vector Regressor:
       Accuracy: Poor
    4. Gradient Boosting Regression:
       Accuracy: 98%, Loss: 2.49
    5. Decision Tree Regression:
       Accuracy: 91.5%, Loss: 5.16
  - For Classification:
    6. Logistic Regression:
       Accuracy: 55%
    7. Random Forest Classification:
       Accuracy: 92%
    8. Support Vector Machine:
       Accuracy: 54%
    9. Gradient Boosting Classification:
       Accuracy: 91%
    10. Decision Tree Classification:
       Accuracy: 87%

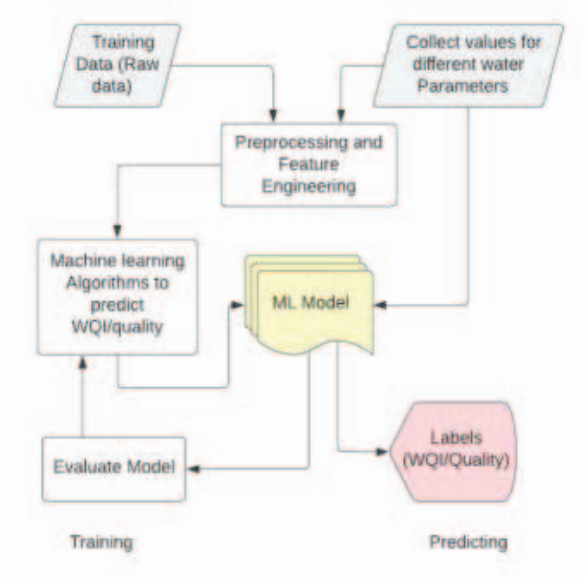

Fig 3. Proposed Architecture for Machine Learning Model

Given the foregoing, it is clear that the Ensemble learning approach outperformed other commonly used learning methods. Ensemble techniques are a type of machine learning technique that integrates numerous base models to create a single best predictive model.
Overall, the main purpose to detect WQI and categorise water samples is achieved.

IV. DATASET AND PREPROCESSING

For dataset preparation, we have taken dataset from below sources,

- data.gov.in: The Indian government's open data platform is called data.gov.in. It offers access to datasets made available by numerous Indian government ministries, departments, states, organizations, and institutes. By making public government data available, data.gov.in seeks to encourage transparency, accountability, innovation, and citizen engagement.
- Kaggle dataset: Kaggle datasets are groups of files that include information about a given topic or domain. These datasets can be explored, downloaded, analyzed, and shared by users on Kaggle or elsewhere.
- Water Quality Database, NWMP: Is a website that provides water quality data collected under the National Water Quality Monitoring Program (NWMP) of India. The NWMP is a program initiated by the Central Pollution Control Board (CPCB) in collaboration with the State Pollution Control Boards (SPCBs) and other agencies to monitor the water quality of various water bodies across the country.

In order to create an expected dataset for our study, we sift through the data we obtained from the sources mentioned above.

V. RESULTS

Several evaluation metrics, including R2 score, RMSE, and MAE, were used to test the models. The results show that the Random Forest and Decision Tree are the most successful for predicting WQI and for classifying water sample and gives highest R2 scores. The model can now predict the Water Quality Index and can classify water samples with high accuracy.

Regression		**Classification**	
Machine Learning Models	**Accuracy**	**Machine Learning Models**	**Accuracy**
Random Forest Regression	96.1%	Random Forest Classification	92%
Gradient Boosting Regression	98%	Gradient Boosting Classification	91%
Decision Tree Regression	91.5%	Decision Tree Classification	87%

The above Table shows algorithms with highest accuracy for detecting WQI & classifying lake water quality.

Using data acquired manually using sensors, the trained model was near enough in learning from the behaviour of data, with an accuracy of 98% in a known environment and up to 96% in an unfamiliar environment.

This showed that the model was able to accurately predict the behaviour of the data in both a known and unfamiliar environment. The model had a high degree of accuracy and was able to learn from the data with minimal human intervention.

VI. FUTURE SCOPE

This paper has proposed a system for monitoring and assessing the water quality of lake using remote sensing images and biosensors. The system can provide real-time [5] and accurate information on various parameters of water quality and alert the users about any issues or trends. However, there are some possible directions for future work that can enhance the performance and functionality of the system. Some of them are:

Incorporating more parameters of water quality such as nutrients, metals, toxins, etc. and developing more robust and reliable algorithms to detect and quantify them from remote sensing images and biosensors.

Improving the user interface and user experience of the website and the application by adding more features such as interactive maps, graphs, charts, etc. that can display the water quality data in a more intuitive and informative way.

Comparing the current and future water quality data with historical records and using statistical models to predict and forecast the water quality variations under different scenarios and conditions. This can help in identifying the patterns and causes of water quality changes over time and planning appropriate management strategies.

VII. CONCLUSION

In this paper, a novel and cost-effective method for monitoring and detecting the water quality of lakes around the world has been presented. Various technologies have been developed for this purpose, among which machine

learning [9][15] and image processing algorithms are the most effective ones. Images captured by smart water buoy can be analyzed by image processing algorithms to detect and measure different water quality parameters such as chlorophyll, nitrate, turbidity, and suspended sediment. The overall water quality of the water body can then be determined and any areas of concern can be identified. Water bodies such as lakes and rivers can be detected and classified by machine learning algorithms, and changes in water quality can be detected. Turbidity, chlorophyll levels, nitrates, and suspended sediments are some of the important indicators of water quality that can be detected by the algorithms. The algorithms can also detect changes in surface temperature and water levels, which can indicate events such as floods. Changes in the aquatic ecosystem, such as changes in species populations, can also be detected by the algorithms, which can help understand the impact of human activities on water bodies. A valuable tool for monitoring and assessing waterway conditions can be provided by integrating remotely sensed data and image processing technologies.

VIII. REFERENCES

[1] K. Shanmugam, M. E. Rana, D. Tan Zi Xuen and S. Aruljodey, “Water Quality Monitoring System: A Smart City Application with IoT Innovation,” 2021 14th International Conference on Developments in eSystems Engineering (DeSE), 2021, pp. 1-6, doi: 10.1109/DeSE53231.2021.00032.

[2] M. Melo, F. Mota, V. Albuquerque and A. Alexandria, “Development of a Robotic Airboat for Online Water Quality Monitoring in Lakes,” Robotics, vol. 8, no. 1, p. 19, 2019. https://doi.org/10.3390/robotics8010019

[3] H. Haque, K. Labeeb, R. B. Riha and M. N. R. Khan, "IOT Based Water Quality Monitoring System by Using Zigbee Protocol," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 1-6. https://doi.org/10.1109/ESCI51835.2021.9394930

[4] V. Lakshmikantha, A. Hiriyannagowda, A. Manjunath, A. Patted, J. Basavaiah and A. A. Anthony, "IoT based Smart Water Quality Monitoring System," Global Transitions Proceedings, vol. 2, no. 2, pp. 1-6, 2021. https://doi.org/10.1016/j.gltpro.2021.11.001

[5] O. V. Pharande, S. S. Patil, S. S. Burse and R. A. Patil, "Real Time Water Quality Monitoring and Management System," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 1, pp. 1-5, 2022. https://www.ijert.org/real-time-water-quality-monitoring-and-management-system

[6] H. A. N. Silva, A. Rosato, R. Altilio and M. Panella, "Modelling and Prediction of Water Quality by Using Artificial Intelligence," in IEEE Access, vol. 9, pp. 41701-41711, 2021. https://doi.org/10.1109/ACCESS.2021.3066423.

[7] J. Chen, D. Zhang, S. Yang and Y. A. Nanehkaran, "Intelligent monitoring method of water quality based on image processing and RVFL-GMDH model," in Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp. 1091-1104, 2021. https://doi.org/10.1007/s12652-020-02635-6.

[8] S. A. Vergina, S. Kayalvizhi, R. M. Bhavadharini and K. Devi, "A Real-Time Water Quality Monitoring Using Machine Learning Algorithm," 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 1-5. https://doi.org/10.1109/ICOSEC49089.2020.9211236

[9] K. Ashwini, J. J. Vedha and M. D. Priya, "Intelligent model for predicting water quality," International Journal of Advance Research, Ideas and Innovations in Technology, vol. 5, no. 2, pp.1-5,2019. https://www.academia.edu/38812197/Intelligent_model_for_predicting_water_quality

[10] A. S. Alaboudi, A. A. Anthony and S. Wang, "A Water Quality Research Platform for the Near-real-time Buoy Sensor Data," 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), 2020, pp. 1-6. https://doi.org/10.1109/IRI49571.2020.00048.

[11] M. R. M. Alves, J. C. B. da Silva and E. A. B. da Silva, "Developing the concept of multifunctional smart buoys," OCEANS 2021: San Diego – Porto, 2021, pp. 1-6. https://doi.org/10.23919/OCEANS44145.2021.9705916.

[12] M. Aminuzzaman, M. Rana and M. Rahman, "Mobile buoy for real time monitoring and assessment of water quality parameters," 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, pp. 1-6. https://doi.org/10.1109/ICEEICT.2015.7307372.

[13] J. Kukal, J. Havel and V. Beran, "Water quality assessment by image processing," 2015 38th International Conference on Telecommunications and Signal Processing (TSP), 2015, pp. 1-6. https://doi.org/10.1109/TSP.2015.7296329

[14] Y. Liu, Y. Zhang and J. Yang, "Underwater Image Quality Assessment: Subjective and Objective Methods," in IEEE Transactions on Multimedia, vol. 24, pp. 1980-1989, 2022. https://doi.org/10.1109/TMM.2021.3074825.

[15] S. R. Patil and S. S. Patil, "Predicting Water Quality Parameters Using Machine Learning," 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2019, pp. 1-5, doi: 10.1109/RTEICT46194.2019.9016825.