Assessing Water Quality through Image Recognition and Machine Learning

Mr. Anirudh Vaidison[1], Mentor: Dr. Liam Yasin[2]

[1]Student, Oberoi International School, Maharashtra, India
[2]Department of Design Engineering at Imperial College London, London, England

---

Abstract - *This research project explores the application of machine learning in assessing water quality through image recognition. The study leverages a diverse dataset of water samples collected from various sources across Mumbai, encompassing ponds, lakes, water outlets, and sewage points. Multiple parameters, including pH, conductivity, turbidity, and dissolved oxygen content, are examined in relation to the resulting HEX colour code, which serves as a visual representation of water quality. The central hypothesis posits that employing machine learning algorithms can reliably predict water safety based on these environmental parameters through image recognition of the colour gradient of the water samples. A systematic approach to data collection, standardisation, and logistic regression modeling has been employed. Results demonstrate the effectiveness of the logistic regression model in predicting water safety with an 85% accuracy rate, highlighting its potential for real-time water quality monitoring and risk assessment. Nevertheless, this study recognises its limitations and the need for further research to refine the model's predictive accuracy and address variations across different geographical regions and water sources. This research aims to contribute to the development of innovative approaches for water quality assessment and therefore, environmental preservation.*

Key Words: **Machine learning, water sources, pH, conductivity, turbidity, dissolved oxygen, HEX colour code, logistic regression**

1.INTRODUCTION

Clean water and adequate sanitation are vital for human existence, environmental sustainability, and economic development. However, the irresponsible disposal of industrial effluence has emerged as a significant threat to water quality in major world cities including Mumbai, a particularly urbanised region. This research paper aims to investigate how pollutants—measured through various water quality parameters such as pH, turbidity, dissolved oxygen and conductivity—affect the colour gradient of various water bodies and therefore, the overall water quality.

1.1 PROBLEM STATEMENT

While water quality using the aforementioned parameters (pH, conductivity, turbidity, and dissolved oxygen) has been tested and analysed in the past, this research paper looks into the feasibility of using image recognition and machine learning as a way of predicting water quality - in essence using technology to look at a very widely prevalent, yet fundamental community problem. The fusion of machine learning with water quality assessment has immense potential to make water safety evaluations more widely accessible.

2. BACKGROUND INFORMATION

The World Wildlife Fund (WWF) defines water pollution as "toxic substances entering water bodies such as lakes, rivers, oceans and so on, getting dissolved in them" ("Water
Pollution").

2.1 FACTORS THAT AFFECT WATER POLLUTION

The most common reason of poor water quality is a consequence of human activity and consumption. In addition to irresponsible disposal of industrial effluence, there are numerous other factors that contribute to widespread contamination of water.

1. Global Warming:
One of the most predominant factors that results in water pollution is global warming. Rising global temperatures caused by increased cardon dioxide emissions heat the water, reducing its oxygen content.

2. Deforestation
Felling forests can exhaust water resources and generate organic detritus which allows harmful microorganisms to develop and enter water ways. This leads to increased sediment and nutrient runoff into water bodies, further degrading water quality and allowing toxic substances to dissolve in the water.

3. Garbage and faecal water dumping

According to the United Nations (UN), more than 80% of the world's sewage ends up in seas and rivers untreated, meaning it cannot be recycled and put to better use ("UNEP - UN Environment Programme"). This contamination introduces pathogens and harmful chemicals into water sources, posing significant health risks to both humans and wildlife.

4. Maritime traffic and fuel spillages

Fishing boats, tankers and cargos tend to dispose of plastic and other pollutants in the sea. Even unintentional oil leakages are detrimental to marine life and water conservation efforts.

2.3 EFFECT OF WATER ON FOUR WATER QUALITY PARAMETERS

Effect of Water Pollution on pH:

The safe range for human consumption of water lies between 6.5 and 8.5 on the pH scale [1]. The pH of water, however, says more about the water than just its alkalinity or acidity. It affects the taste, odour, and colour of water, making it a clear qualitative indicator of water quality. For example, water with a low pH indicates it is 'acidic' indicating the presence of heavy metals in the water that result in a more toxic sample (McGrane). Whereas, water with a high pH indicates it is 'basic' meaning there are less heavy metals available resulting in an overall less toxic sample, although too high a pH could also indicate the presence of bacteria and contaminated water ("Microbial Growth at Low or High pH").

Urban areas, such as Mumbai, contribute to water pollution through rainwater runoff that carries pollutants on the surface such as oil, heavy metals, etc into water bodies. These metals are usually toxic to marine life and can alter the pH of the water posing a threat to humans. Water with acidic pH levels can contain iron, manganese, copper, lead and zinc elements that may in turn cause "health problems such as cancer, stroke, kidney disease, memory problems and high blood pressure" (Culligan Nation).

Effect of Water Pollution on Electrical Conductivity:

Electrical conductivity is a measure of how well a substance allows the flow of electric charges, typically in the form of electrons or ions.

Water pollution often introduces various charges into aquatic ecosystems. For example, agricultural runoff, industrial discharges, and urban runoff can contain ions like sodium, chloride, nitrate, sulfate, and heavy metal ions. These ions increase the electrical conductivity of water as they are capable of carrying electrical charges [5].

Importantly, there is a relationship between pH and electrical conductivity. Heavy metals found in water pollution affect pH (discussed in earlier paragraphs), and consequently, these changes in pH can affect the dissociation of ions, potentially impacting electrical conductivity. Acidic conditions, for example, can lead to the release of hydrogen ions (H^+), which can alter the conductivity.

Effect of Water Pollution on Turbidity:

Turbidity is a measure of the cloudiness or haziness of a fluid caused by the presence of suspended particles, which scatter and absorb light. Turbidity is an essential parameter in environmental and water quality monitoring, as it can indicate the presence of contaminants or impurities in a liquid [3].

Pollutants such as sediment, clay, organic matter, and debris can be introduced into water bodies through various sources, including erosion, industrial discharges, and sewage runoff. These particles can increase turbidity, making the water appear cloudy or visibly spoilt. Industries that release effluents containing fine particles, chemicals, heavy metals, and other contaminants can significantly increase turbidity in nearby water bodies. Construction sites in Mumbai, are also the cause of such particles entering water bodies (Tembhekar et al.).

A point to note is that with greater suspended particles in a water sample, there is a higher chance of finding metals in the water. This is notable because with a higher amount of suspended particles, the impurities containing ions allow electricity to be easily conducted. Hence, turbidity and electrical conductivity have a direct relationship, wherein, the increase in the value of one will result in the increase of the other.

Effect of Water Pollution on Dissolved Oxygen:

Dissolved oxygen content, often referred to as dissolved oxygen (DO), is a crucial water quality parameter that measures the concentration of oxygen gas (O_2) that is present in a liquid, typically water.

The dissolved oxygen content in water is usually measured in units such as milligrams per liter (mg/L) or parts per million (ppm). It is a critical factor in assessing the health and quality of aquatic ecosystems and is affected by various environmental factors, including temperature, pressure, and the presence of organic matter and pollutants.

Water pollution, resulting from various contaminants such as organic matter, nutrients, toxins, and sediments from industrial, agricultural, and municipal sources, can significantly deplete dissolved oxygen (DO) levels in water samples (Britannica). This pollution disrupts the delicate balance of oxygen required for aquatic life, leading to oxygen-depleted "dead zones" in water bodies, reduced biodiversity, and altered nutrient cycling (US EPA). Pollution-induced eutrophication and increased biological oxygen demand (BOD) further exacerbate the problem, highlighting the critical need for effective pollution control measures and sustainable land use practices to protect the health of aquatic ecosystems and water quality.

Turbidity and dissolved oxygen show an inverse relationship; as the value of turbidity increases, there are less oxygen gas molecules present in the water sample, vice versa [4].

2.4 INSTRUMENTS MEASURING PARAMETERS

1. pH Meter – pH: scale from 1-14
2. Conductivity Meter - Electrical Conductivity: expressed in microsiemens per centimeter (µS/cm).
3. Turbidimeter - Turbidity: expressed in Nephelometric Turbidity Units (NTU) or Parts Per Million (PPM).
4. Nephelometer (DO meter) - Dissolved Oxygen: expressed in milligrams per liter (mg/L) or parts per million (ppm)

2.5 HOW PARAMETERS AFFECT HEX COLOUR

pH can influence the colour of water indirectly by affecting the solubility and speciation of certain substances. For example, water with a low pH (acidic) can cause the leaching of metals like iron and manganese, leading to discolouration in the form of reddish-brown or yellowish tints. Extreme pH values can also impact the colour perception, making the water appear more or less translucent.

Conductivity measures the concentration of ions in water, mainly due to the presence of salts and minerals. While high conductivity can be indicative of dissolved ions, it does not influence the colour of the water directly. High ion concentrations can affect the taste and odor of water, but they do not define the colour.

Turbidity is a measure of the cloudiness or haziness of water due to suspended particles. The presence of suspended particles in water can scatter and absorb light, leading to variations in the water's appearance.

High turbidity can make water look cloudy or murky, which can affect the perceived colour.

The dissolved oxygen content does not have a direct impact on the colour of water. However, it can indirectly affect the health of aquatic ecosystems and the growth of algae or other organisms that may impart different colours to water, such as green or brown hues. Changes in dissolved oxygen levels can influence the biological activity and the presence of organic matter in water, which may impact the water's overall appearance.

3. SCIENTIFIC LITERATURE REVIEW

Groundwater quality is assessed through various parameters such as turbidity, pH, dissolved oxygen (DO), and conductivity. Recent studies have also explored the application of machine learning in predicting and managing water quality. This literature review categorizes the selected research into these themes to provide a cohesive understanding of the topic.

<u>Turbidity and Its Correlation with Water Quality</u>

Turbidity, a measure of water clarity, is influenced by suspended particles and has significant implications for water quality. Huey et al. [3] assessed turbidity alongside total suspended solids (TSS) and microbial concentrations in various watersheds. Their study found turbidity to be a reliable indicator of TSS and microbial pollution. Similarly, Argenal et al. [4] examined the impact of turbidity on dissolved oxygen levels, finding an inverse relationship between the two. High turbidity was associated with lower DO levels, indicating poorer water quality.

Another study [7] reinforced this inverse relationship, highlighting the detrimental effects of high turbidity on aquatic ecosystems. Rusydi [9] explored the correlation between electrical conductivity and turbidity as indicators of water salinity. This study underscored the complexity of this relationship, emphasizing the need for precise turbidity measurements to accurately assess water quality.

<u>pH and Dissolved Oxygen (DO) Relationships</u> The relationship between pH and DO is crucial for understanding aquatic ecosystem health. Zang et al. [2] investigated this relationship in aquaculture and nonaquaculture waters, finding significant positive correlations between pH, DO, and chlorophyll in eutrophic nonaquaculture waters. Seasonal variations in pH and DO were also studied in Asa Lake, Nigeria [6], revealing fluctuations that impact fish diversity.

In contrast, a study [8] in Vancouver's Salish Creek and Canyon Creek found no significant relationship between pH and DO concentrations, suggesting that this correlation may vary based on local environmental factors.

Conductivity as an Indicator of Water Quality

Conductivity, which measures the ability of water to conduct electrical current, is influenced by the presence of inorganic dissolved solids. Dr. Abha Mathur [5] discussed how conductivity is determined by the geology of the surrounding area, with granite bedrock areas yielding lower conductivity and clay soils resulting in higher conductivity due to ionizing substances. The study also highlighted the impact of human activities, such as sewage system failures and oil spills, on conductivity levels.

Machine Learning Applications in Water Quality Assessment Machine learning has emerged as a powerful tool for analyzing and predicting water quality. One study [10] reviewed the application of machine learning in water environment research, highlighting its effectiveness in improving water treatment, pollution control, and ecosystem management. Sami et al. [11] used an artificial neural network (ANN) model to predict DO levels in Taipei City's water sources. The model demonstrated high accuracy, showcasing the potential for real-time water quality monitoring.

In India, a study [12] utilized smartphone cameras and machine learning to assess water turbidity, resulting in a user-friendly app for rapid analysis. This innovation provides a practical solution to the challenge of unsafe water consumption. Silveira et al. [13] combined remote sensing and machine learning techniques to map suitable habitats for coral reefs, demonstrating the broader applicability of these technologies in environmental monitoring.

4. VARIABLES

Independent Variables:

The independent variables in this study are represented by the various water bodies from which water samples are procured. These water bodies serve as distinct categories or groups for the machine learning algorithm's analysis. Within each category, the parameters, including pH, conductivity, turbidity, and dissolved oxygen content, are measured and act as the features or characteristics used for the machine learning algorithm. These parameters are the focus of the study particularly concerned with their relationship to the HEX colour code, and the different water bodies, from different locations, are the independent variable. The machine learning algorithm employs these features within each category to make predictions, classifications, or other data-driven decisions. This research design allows the algorithm to learn patterns and relationships within and between the water bodies and the parameters, enabling it to make predictions or assessments relevant to the study's objective of assessing water quality through image recognition through supervised learning.

Dependent Variable:

The dependent variable in this study is represented by the HEX colour code generated as a result of the various water samples procured, each of which is characterized by different parameter values. The hex colour code is the output or outcome of interest in this research. It is a visual representation that encapsulates the information derived from the measured parameters, including pH, conductivity, turbidity, and dissolved oxygen content. HEX codes are obtained through a colour extraction algorithm. The machine learning algorithm processes and quantifies these parameters to correlate with the HEX colour code, which serves as the dependent variable for analysis and assessment. The colour code reflects the comprehensive data transformation achieved by the algorithm, making it a key focus in understanding the relationships and patterns uncovered in the study.

5. HYPOTHESIS

The utilization of a machine learning algorithm for predicting water safety based on the parameters of pH, conductivity, turbidity, and dissolved oxygen content, within various water bodies, will result in a reliable and efficient method for assessing water quality, as represented by the corresponding HEX colour code.

Supporting Evidence:

pH levels influence the colour of water indirectly by affecting the solubility and speciation of certain substances. For instance, low pH (acidic) can lead to the leaching of metals like iron and manganese, which can result in reddish-brown or yellowish tints, thus affecting the HEX colour code. Conductivity measures ion concentration in water, impacting its taste and odor. High conductivity does not directly affect the HEX colour code, but it can indirectly influence water quality which in turn may have an effect on the colour composition of water. High ion concentrations can alter water characteristics, potentially impacting the HEX colour code. Turbidity, which affects water clarity and appearance, can directly impact the HEX colour code. High turbidity can make water appear cloudy, leading to differences in the HEX colour code. While dissolved oxygen content does not directly affect the HEX colour code, it can impact the health of aquatic ecosystems, potentially influencing the growth of algae or other organisms that may impart different colours to water. This, in turn, may indirectly affect the HEX colour code.

6. METHODOLOGY

The research employed a systematic approach to collect and analyse water samples from diverse sources across Mumbai and develop a logistic regression model for water quality assessment.

1. Sample Collection:
60 water samples were collected from various locations, including ponds, lakes, water outlets, and sewage sources throughout Mumbai, to ensure a broad spectrum of water sources and conditions. The collection was carried out using a bucket and a rope. Each sample was collected in individual sample cups to maintain sample integrity. 3-5 samples of the same source were collected.

Fig. 6.1-6.6: Pictures of Various Ponds/Industrial Sites Across Mumbai from which Samples were Collected

2. Data Collection:
Multiple water quality parameters were measured for each collected sample. The IONIX 7 in 1 multiparameter instrument was used to gather data on pH, electrical conductivity, and turbidity. The Lutron Do-5509 Dissolved Oxygen Meter was used to collect data for dissolved oxygen. Both instruments were calibrated and standardized before use. Three trial measurements for each sample were taken.

Fig. 6.7: Data Collection using a Digital Multiparameter and a Nephelometer

3. Data Analysis & Standardization: The collected data, including pH, conductivity, turbidity, and DO values, was transferred to an Excel spreadsheet for analysis. This data was organized and averaged to prepare it for the development of a logistic regression model. Standard values for drinking water from secondary sources, mentioned in the background research, allowed for a human decision on whether the water was potable or contaminated beyond saving. This was indicated by a binary factor: if the water sample's data was within the standard drinking water ranges, it would yield a potability factor 1; however, if the data reflected digits outside of the standard drinking water range, a potability factor 0 was attained. The two factors that were given highest priority were conductivity and turbidity since they showed a positive correlation, hence, the portability factor was based off of the standard drinking water values for these parameters

Sample No.	Average pH	Average turbidity	Average conductivity	Average DO	Hex code w/o#	Potability (0=nonpotable, 1=potable))
1	6.95	5327	10623	4.7	9a9e9a	0
2	6.99	5320	10610	4.7	8e908b	0
3	6.99	5317	10610	4.5	6b7f98	0
4	6.99	5317	10617	3.9	81827e	0
5	6.99	5317	10603	3.9	90918c	0
6	6.95	3267	6593	3.45	9c9e97	0
7	6.94	3283	6593	4.1	979890	0
8	6.93	3290	6593	3.5	747168	0
9	6.94	3290	6607	3.9	91928c	0
10	6.96	3290	6603	3.75	7d7c74	0
11	6.88	330	659	3.15	777872	1
12	6.84	320	638	2.9	9d9f9b	1
13	6.85	320	639	2.6	777772	1
14	6.86	320	638	2.75	787971	1
15	6.85	320	639	2.25	565550	1
16	6.94	235	461	1.75	83847e	1
17	6.93	232	472	2.05	8b8c86	1
18	6.95	251	498	1.5	767671	1
19	6.95	252	501	2.35	8f9494	1
20	6.95	251	503	1.75	929797	1

Fig. 6.8: Digitisation of data and standardization of samples

Simultaneously, using the colour extraction algorithm the HEX code values and the respective intensity percentage of those HEX code values were extracted. These values were accounted for in the data sheet as they play a crucial role in the development of a supervised machine learning algorithm. Below is an example of one of the samples that underwent this process.

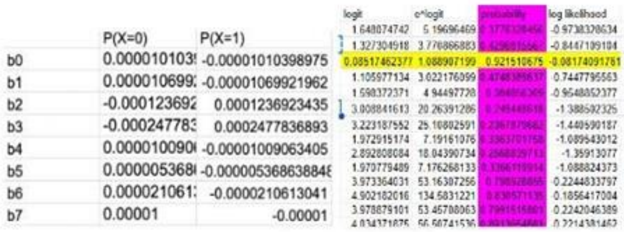
Fig. 6.9: Colour extraction of water samples

4. Model Development:

A logistic regression model was created using the processed data. Logistic regression is a type of regression analysis used to predict a dependent variable's outcome based on one or more independent variables. The dependent variable in logistic regression is binary, meaning it has only two possible outcomes (e.g., safe or unsafe water). Unlike linear regression, which predicts a continuous outcome, logistic regression predicts the probability of a certain class or event existing. In logistic regression, each feature (water quality parameter) is assigned a coefficient, which represents its contribution to the logit (log-odds) of the outcome. The logit function is given by:

$$logit(p) = ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

Where p is the probability of the outcome being 1 (potable), $\beta_0\beta_0$ is the intercept, $\beta_i\beta_i$ are the coefficients for the variables X_i X_i . The model coefficients ($\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$) were estimated using the maximum likelihood estimation method. This involves finding the values of the coefficients that maximize the likelihood of an accurate model as per standard drinking water conventions. In other words, reducing the number of outliers.



	P(X=0)	P(X=1)
b0	0.000010103!	-0.00001010398975
b1	0.000010699:	-0.00001069921962
b2	-0.000123692	0.0001236923435
b3	-0.000247783	0.0002477836893
b4	0.0000100901	-0.00001009063405
b5	0.0000053681	-0.00000536863884£
b6	0.000021061:	-0.0000210613041
b7	0.00001	-0.00001

logit	e^logit	probability	log likelihood
1.648074742	5.19696469	[illegible]	-0.9738328634
1.327304918	3.770866883	[illegible]	-0.8447109104
0.08517462377	1.088907199	0.921510675	-0.08174091781
1.105977134	3.022176099	[illegible]	0.7447795563
1.598372371	4.94497728	[illegible]	-0.9548852377
3.008841613	20.26391286	[illegible]	-1.388592325
3.223187552	25.10802591	[illegible]	-1.440590187
1.972915174	7.19161076	[illegible]	-1.089543012
2.892808084	18.04390734	[illegible]	-1.35913077
1.970779489	7.176268133	[illegible]	-1.088824373
3.973364031	53.16307256	[illegible]	0.2244833797
4.902182016	134.5831221	[illegible]	0.1856417004
3.978879101	53.45708063	[illegible]	0.2242046389
[illegible]	[illegible]	[illegible]	[illegible]

Fig. 9-10: Set values for parameters that yield probabilities

The predicted probability of water safety (p) is then obtained by applying the logistic (sigmoid) function to the logit:

$$p = \frac{1}{1 + e^{-logit(p)}}$$

This converts the logit value into a probability ranging between 0 and 1. A threshold of 0.5 was used to classify the water as safe or unsafe. Probabilities greater than 0.5 were classified as potable (1), while probabilities less than 0.5 were classified as non-potable (0).

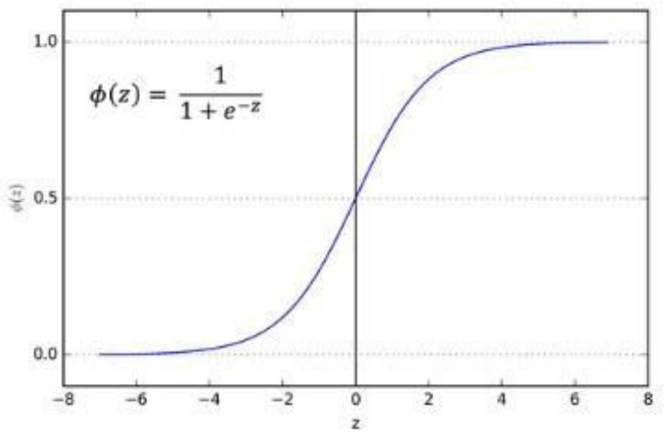


Fig. 11: Sigmod function (logistic regression curve)

5. Model Validation:

Samples that show discrepancies between their probability (0 or 1) and their respective potable factor are highlighted in yellow and are considered anomalies in this model. The developed logistic regression model was tested and validated using an unknown sample and predicting its quality. An 85% accuracy rate was attained with the numericals of each parameter described above. These anomalies are highlighted in yellow as shown in Fig. 6.10.

7. RESULTS

The results of this study demonstrated the feasibility of employing a logistic regression model to assess water quality and predict potability based on water parameters within various sources. Water samples from a diverse range of locations across Mumbai, including ponds, lakes, water outlets, and sewage sources, were collected and analysed. Key water quality parameters, namely pH, conductivity, turbidity, and dissolved oxygen (DO), were measured using specialized instruments. Additionally, the HEX colour code values and their respective intensity percentages were extracted using a colour extraction algorithm, providing crucial information for the machine learning model.

Upon standardising the collected data and applying a binary classification for potability (1 for safe, 0 for unsafe), the logistic regression model was developed to predict water safety. The model's construction involved assigning values to each parameter and calculating a logit value using a fundamental formula. These logit values were then transformed into probabilities using the sigmoid function (logistic regression curve).

Throughout the data collection, where turbidity was high conductivity reflected the same increase in magnitude. For example, in sample 4 there was a conductivity of 5317 µs/cm while there was a turbidity of 10617 ppm. Contrastingly, in sample 31 there was a conductivity of 357 µs/cm with a proportionately low turbidity of 712 ppm. Regardless of the sample number, there is a direct proportion between the two parameters. This can be quantitatively backed by dividing the two values resulting in a constant. For any sample, the average conductivity divided by the average turbidity results in the numerical constant ~2. This was one such prediction that the hypothesis predicted correctly.

In the model validation phase, the logistic regression model demonstrated an 85% accuracy rate in predicting the quality of

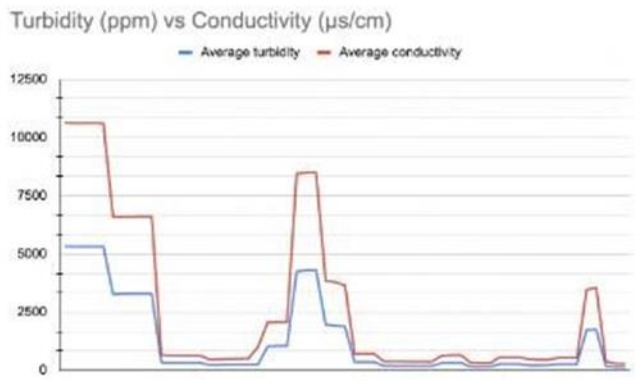


Fig. 7.1: Graph of turbidity vs conductivity

unknown water samples. There is some degree of error that mainly stems from the fact that the values of dissolved oxygen were inconsistent with that of turbidity and conductivity.

In some instances where the turbidity was extremely high, the dissolved oxygen was low. While in other instances where the turbidity was relatively low, the dissolved oxygen stayed low. The dissolved oxygen showed no clear relationship with any of the other such parameters, with the pH being a dormant factor in this study. This was one such case wherein this mutually exclusive dissolved oxygen relationship was overlooked in the background information.

The HEX code correlation may induce unwarranted errors due to the nature of obtaining the colour code. Since the method involves extracting the colour based on an image, factors such as lighting, camera quality, time of the day, and background colour may affect the colour code obtained. This definitely increases the uncertainty of the methodology, although the more quantitative aspects of the research had lesser uncertainty. For example, the pH, and dissolved oxygen had uncertainties of ± 0.01 which does not significantly affect the binary outcome of the potability of the sample.

8. CONCLUSION

The findings of this research highlight the potential of utilizing a supervised machine learning algorithm, specifically a logistic regression model, to assess water quality based on key water parameters and the corresponding HEX colour code. The study leveraged water samples from diverse sources to ensure a comprehensive evaluation of water quality. pH, conductivity, turbidity, and DO were identified as critical factors for water quality assessment, as they influence the HEX colour code. The observed relationships between these parameters and the HEX colour code align with previous research findings, emphasizing the importance of these variables in determining water safety.

The logistic regression model's development and validation indicate its capability to accurately predict water potability, with an 85% accuracy rate in classifying samples as safe or unsafe. This outcome supports the hypothesis that utilizing a machine learning algorithm for water quality assessment is an efficient and reliable method. The logistic regression model offers a practical tool for decision-makers and water quality managers to swiftly evaluate water safety in different locations and make informed decisions to safeguard public health.

It is crucial to note that this research design and model rely on the relationships observed within the specific water bodies in Mumbai. Variations in geological and environmental conditions in other regions may influence the parameters' relationships differently. Hence, while this logistic regression model is promising for Mumbai, its generalizability to other locations necessitates further investigation.

Moreover, the findings of this study have practical implications for water quality management and public health. The ability to rapidly and accurately assess water safety using machine learning algorithms can facilitate timely actions to protect communities from potentially contaminated water sources. The implementation of this model in real-time monitoring and decision support systems could significantly enhance the safety of water supplies and improve the overall well-being of communities. Future research should focus on validating the model in diverse geographical and environmental contexts to ascertain its broader applicability.

While water quality using the aforementioned parameters (pH, conductivity, turbidity, and dissolved oxygen) has been tested and analysed in the past, this research paper looks into the feasibility of using image recognition and machine learning as a way of predicting water quality - in essence using technology to look at a very widely prevalent, yet fundamental community problem. The fusion of machine learning with water quality assessment has immense potential to make water safety evaluations more widely accessible.

REFERENCES

[1] US EPA, O. (2015, November 4). *Dissolved Oxygen*. Www.epa.gov. https://www.epa.gov/caddis/dissolvedoxygen#:~:text=colour%3A%20The%20colour%20of%20water

[2] Electrical Conductivity and Electrical Resistivity - Resistivity of Materials, Formula, Unit, Examples and FAQs. (n.d.). BYJUS. Retrieved July 31, 2024, from https://byjus.com/physics/resistivity-variousmaterials/#:~:text=Electrical%20Conductivity%20is%20an%20intrinsic%20property%20of%20a

[3] Huey, G. M., & Meyer, M. L. (2010). Turbidity as an Indicator of Water Quality in Diverse Watersheds of the Upper Pecos River Basin. Water, 2(2), 273–284. https://doi.org/10.3390/w2020273

[4] Kang, C., Lee, T. M., Wong, B., & Yoo, J. (2018). Relationship Between Soil pH and Dissolved Oxygen (DO) Concentration in Salish Creek and Canyon Creek. The Expedition, 8. https://ojs.library.ubc.ca/index.php/expedition/article/view/191429

[5] pH Measurement | Electrical Instrumentation Signals | Electronics Textbook. (n.d.). Www.allaboutcircuits.com. https://www.allaboutcircuits.com/textbook/directcurrent/chpt-9/ph-measurement/

[6] Mathur, A. (2018). Conductivity: Water Quality Assesment. International Journal of Engineering Research & Technology, 3(3). https://doi.org/10.17577/IJERTCONV3IS03028

[7] Pathak, A. K., Sharma, M., Katiyar, S. K., Katiyar, S., & Nagar, P. K. (2020). Logistic regression analysis of environmental and other variables and incidences of tuberculosis in respiratory patients. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-79023-5

[8] Water Science School. (2019, October 22). pH and Water | U.S. Geological Survey. Www.usgs.gov. https://www.usgs.gov/special-topics/water-scienceschool/science/ph-and-water#:~:text=Pollution%20can%20change%20a%20water

[9] Taylor, M., Elliott, H. A., & Navitsky, L. O. (2018). Relationship between total dissolved solids and electrical conductivity in Marcellus hydraulic fracturing fluids. Water Science and Technology, 77(8), 1998–2004. https://doi.org/10.2166/wst.2018.092

[10] Iberdrola. (2021). Water pollution: how to protect our source of life. Iberdrola. https://www.iberdrola.com/sustainability/waterpollution

[11] Ltd, N. W. S. P. (n.d.). What are Conductivity in the water and its relationship with water TDS. Water Treatment Company in India - Buy STP, ETP, RO Plant | Netsol Water. https://www.netsolwater.com/whatare-conductivity-in-the-water-and-the-relationshipbetween-conductivity-and-tds-in-thewater.php?blog=4555

[12] Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. EcoEnvironment & Health, 1(2). https://doi.org/10.1016/j.eehl.2022.06.001