Water Quality Analysis from Google Earth images

Department of Computer Engineering, Ramrao Adik Institute of Engineering, Dr. D. Y. Patil Deemed to be University, Nerul, Navi Mumbai.

Abstract:

Introduction:

Water quality generally refers to the extent to which water is fit for a specific use and whether or not pollutants in the water are a threat to the environment. Freshwater that is clean, safe, and sufficient is essential for human survival as well as the survival of all other living things in the environment. Complex and varied water quality challenges demand immediate worldwide attention and action. Due to its intrinsic potential to significantly affect the hydrological cycle, the reduction in water quality has become a global problem. Due to the enormous rise in population, the quick rate of urbanization, the intensification and extension of agricultural methods, and other factors, man has had a significant impact on the environment over the past 10 years. This has caused resources, particularly surface water, to degrade gradually and continuously. Diseases are spread largely through contaminated water. 1.8 million people residing in underdeveloped nations die every year due to water-borne diseases[1].

Water's spectral characteristics fluctuate with wavelength of incident radiation due to both its molecular makeup and contaminants in the water body. The assessment of water quality is used to establish if the water is fit for consumption or secure for the environment. This is often done by comparing the water's physio-chemical and biological qualities or parameters against a set of standards. Water quality assessment is defined as "the evaluation of the physical, chemical, and biological state of the water in relation to its natural state, anthropogenic effects, and future uses."

The link between water quality indicators and spectral reflectance must be established in order to monitor water quality by remote sensing. Particularly in the case of non-point source contamination, remote sensing plays a significant role in the evaluation of water quality and the creation of management plans. Water quality evaluation and monitoring can make good use of the benefits of remote sensing, such as synoptic coverage, the development of databases in close to real time, and the availability of multispectral, hyper spectral, and multi temporal data.

So to effectively track land use and document land change brought on by climate change, urbanization, drought, wildfire, biomass changes (carbon assessments), and a host of other natural and human-caused changes, Landsat satellites have the best ground resolution and spectral bands. Landsat satellites collect data by using passive sensors onboard the satellite that detect radiation emitted from the Earth in different bands of the electromagnetic spectrum.The Operational Land Imager Band 1 (0.433-0.453 µm) on Landsat 8 and Landsat 9 is useful for imaging shallow water and tracking fine atmospheric particles like dust and smoke.

Two sensors are carried by Landsat 8. The Ball Aerospace & Technology Corporation is the manufacturer of the Operational Land Imager sensor. NASA Goddard Space Flight Center is in charge of producing the Thermal Infrared Sensor. The visible, near infrared, and shortwave infrared (VNIR, NIR, and SWIR) regions of the spectrum are where the OLI measures.

There are various water quality indicators observable from Landsat 8:

- Turbidity and sediments
- Colored Dissolved Oxygen Matter(CDOM)
- Sea Surface Temperature(SST)
- Chlorophyll-a

- Salinity
- Total Suspended Solids(TSS)
- Fluorescence Line Height
- Euphotic Depth

Turbidity:

Water clarity is gauged by turbidity (i.e., transparency). Water can seem cloudy or murky when suspended particles, such as silt, algae, plankton, and sewage, are present. Instead of allowing light to pass through the water directly, these particles scatter and absorb light waves. A greater turbidity rating denotes cloudier, "thicker," and more particle-filled water. The turbidity of water is low when it is clear.

Salinity:

Salinity is an indicator of how much salt is in a body of water. The majority of the salt in seawater is sodium chloride (NaCl). A mixture of dissolved ions, including sodium, chloride, carbonate, and sulphate, can result in some lakes having a high salinity level. Water used for irrigation or drinking can have a worse quality due to salts and other impurities. Also, they have a significant impact on aquatic biota, each of which has a typical range of salinity tolerance. Moreover, the water's ionic makeup might be important.

Chlorophyll:

The concentration of phytoplankton is determined by the amount of chlorophyll present in a water sample. These metrics help us comprehend the system's overall biological "health," including its trophic status and primary output. Chlorophyll measurements can forecast hazardous algal blooms and identify algal bloom occurrences and their impacts on water quality.

(Remote Sensing):

Remote sensing provides water quality data with a high spatial and temporal resolution for thousands of water bodies at a time. It supports the evaluation of environmental problems and potential health risks through the analysis of changes in water quality. The two main subcategories of remote sensing are passive and active. Only naturally occurring electromagnetic radiation from the earth's surface, such as reflected sunlight or emitting thermal infrared rays that depend on the temperature of objects, is picked up by passive remote sensors. Photographic cameras, electro-optical sensors, and passive microwave sensors are examples of typical passive sensors. Active sensors, like laser radar, emit their own artificial radiation to illuminate the object of study and then collect the reflected portion.

(Regarding Bay of Bengal):

The Bay of Bengal is the north-eastern region of the Indian Ocean, bordered by India on the west and northwest, Bangladesh on the north, and Burma and India's Andaman and Nicobar Islands on the east. South and Southeast Asia have countries that rely on the Bay of Bengal. The Bay of Bengal covers a total size of 2,600,000 square kilometers (1,000,000 sq mi). The Ganges-Hooghly, Padma, Brahmaputra-Yamuna, Barak-Surma-Meghna, Irrawaddy, Godavari, Mahanadi, Brahmani, Baitarani, Krishna, and Kaveri are among the major rivers that flow into the Bay of Bengal. Sewage-borne pathogens and organic load; solid waste/marine litter; growing nutrient inputs; oil pollution; persistent organic pollutants (POPs) and persistent toxic substances (PTSs); sedimentation; and heavy metals are the key transboundary pollution and water quality challenges for Bay of Bengal. The transboundary character of these challenges is exemplified by the widespread discharge of untreated/partially treated sewage. The Ganges-Brahmaputra-Meghna River's sewage and organic emissions are anticipated to be transboundary. Plastics and abandoned fishing gear may go considerable distances across national borders. Distinctions between countries in terms of shipping discharge legislation and enforcement may encourage discharges across borders. Tar balls travel across vast distances. POPs/PTSs and mercury, notably organo-mercury, are transported across great distances. Most heavy metal pollution and sedimentation are confined and lack a significant transboundary dimension. The primary reasons for the problems are: rising coastal population density and urbanization; increased consumption, resulting in more garbage created per person; insufficient finances devoted to waste management; industrial relocation into BOBLME nations; and the growth of tiny companies. The fast expansion of the shrimp cultivation business, which necessitates the use of antibiotics and pesticides for export-quality food safety yet pollutes the Bay of Bengal, is an important concern[2].

(Regarding Arabian Sea):

The Arabian Sea is a region of the northern Indian Ocean bounded to the north by Pakistan, Iran, and the Gulf of Oman, to the west by the Gulf of Aden, the Guardafui Channel, and the Arabian Peninsula, to the southeast by the Laccadive Sea and the Maldives, to the southwest by Somalia, and to the east by India. It has a total size of 3,862,000 km2 (1,491,000 square miles) and a maximum depth of 4,652 meters (15,262 ft). The Arabian Gulf is one of the world's most polluted marine habitats, owing to a combination of pollution factors such as climate change, oil and gas operations, and coastal human disturbances.The Arabian Gulf's marine ecosystem is naturally stressed by the continual change in water temperature and high salinity. According to Joydas et al. (2015), water temperature can reach 36 °C in the summer and as low as 15 °C in the winter, with salinity exceeding 43 ppt[3]. Human activities place an even greater strain on the maritime ecosystem. In reality, the Arabian Gulf is one of the most anthropogenically altered areas on the planet. The Arabian Gulf's rising pollution presents a significant danger to the region's marine ecosystems and aquatic species[4].

(Regarding Lonar Lake):

Lonar Lake, also known as Lonar crater, is a designated National Geo-heritage Monument, salty, soda lake in the Buldhana district of Maharashtra, India. Lonar Lake lies roughly 137 meters (449 feet) below the crater rim and has a mean diameter of 1.2 kilometers (3,900 feet). The rim of the meteor crater is roughly 1.8 kilometers (5,900 feet) in diameter. The use of fertilizers, pesticides, and harmful compounds in the agricultural fields surrounding the lake pollutes the lake water. The perennial streams "Dhara" and "Sita Nahani" are one of the lake's water sources. Locals, pilgrims, and visitors utilise them for swimming, washing clothing and animals, and other domestic activities. Household effluents containing detergents are disposed of here on a regular basis. Illegal deforestation occurs in the surrounding area, and livestock grazing inside or near the crater's rim causes fecal contamination. The lake's environment is being harmed as a result of the sewage dump. Marauding pilgrims and growing pollution are disrupting its diverse flora and wildlife, which includes around 100 permanent and migratory species. The use of fertilizers, pesticides, and harmful compounds in the agricultural fields surrounding the lake pollutes the lake water. The perennial streams "Dhara" and "Sita Nahani" are one of the lake's water sources. Locals, pilgrims, and visitors utilize them for swimming, washing clothing and animals, and other domestic activities. Household effluents containing detergents are disposed of here on a regular basis. Illegal deforestation occurs in the surrounding area, and livestock grazing inside or near the crater's rim causes fecal contamination. The lake's environment is being harmed as a result of the sewage dump. Marauding pilgrims and growing pollution are disrupting its diverse flora and wildlife, which includes around 100 permanent and migratory species[5].

Literature Survey:

(Paper difficulties):

(Sep 2021) A. K. M. Azad Hossain, Caleb Mathias, and Richard Blanton proposed "Remote Sensing of Turbidity in the Tennessee River Using Landsat 8 Satellite," in which they created a numerical turbidity estimation model for the Tennessee River and its tributaries in Southeast Tennessee using Landsat 8 satellite imagery and near real-time in situ measurements. The restriction was that in-situ measurements for the research region would be gathered to increase model accuracy, however this was not achievable owing to cloud covering and limited resources. As a result, the data became more skewed and had less variety. In addition to quantitative research, qualitative studies are commonly used to estimate turbidity when there is little to no in situ data available. When quantitative forecasts are not required, these qualitative predictions might be useful for general assessment of water quality conditions. One researcher created a turbidity metric for qualitative analysis called the Normalized Difference Turbidity Index (NDTI).

(Jan 2020)Yuanhong Li, Zuoxi Zhao, Sunghwa Han, Xiao Wang, and Zong Liu submitted "Lagoon water quality monitoring based on digital image analysis and machine learning estimators," in which they monitored nitrogen, phosphorous, bacteria, and total solids parameters. To prove such correlations, they employed machine learning to train three types of estimators, with their primary methods being normal equation linear regression (LR), stochastic gradient descent (SGD), and ridge regression (R-PLS). This article also has limitations, such as the suggested method's application to just a subset of lagoon effluent contamination indicators. Heavy metal contamination indicators outside the visual range cannot be assessed for some substances.

(Jun 2021)Xidong Chen, Liangyun Liu, Xiao Zhang, Junsheng Li, Shenglei Wang, Dong Liu, Hongtao Duan, and Kaishan Song proposed "An Assessment of Water Color for Inland Water in China Using a Landsat 8-Derived Forel-Ule Index and the Google Earth Engine Platform," in which a cloud-free composite image of China for the summer of 2015 was generated using time-series The created BAP composite and the Google Earth Engine computer platform were then used to create the first Forel-Ule index (FUI) water color product with a resolution of 30 m for China. The paper's limitation was that the influence of bottom reflectance might change the perceived hue of water bodies, which is common in optically shallow environments. Because of the shallow depths, signals from the sediment and plants at the bottom of the sea may also contribute to the reflectance out of the water, as there is no mechanism for reliably classifying shallow pixels on a broad scale.

()Ersan Batur and Derya Maktav proposed "Assessment of Surface Water Quality by Using Satellite Images Fusion Based on PCAMethod in the Lake Gala, Turkey" in which in order to monitor the quality of the surface water in Gala Lake, the PCA model is described in this research. It integrates surface water reflectance values from satellite images. The model developed in this research was computed for the surface water quality characteristics of Lake Gala measured exclusively during the summer months because surface water quality models can only be applied to the relevant region and season. For a model that covers all times, longer measurements should be taken and examined.

With novel models developed employing band ratios and long-term data, surface water quality characteristics can be more precisely predicted for further research.

(Jan 2018) Hendrik jan van der Woerd and marcel Robert Wernand proposed “Hue-Angle Product for Low to Medium Spatial Resolution Optical Satellite Sensors” in which it is possible to determine the hue color of the body of water, but not all of the colors can be determined with equal accuracy. Despite having only four small bands in the visible spectrum and one large band in the IR, the CZCS sensor proved remarkably good in retrieving the hue angle for the majority of angles.

(Dec 2015) Shungudzemwoyo P. Garaba, anna Friedrichs, Daniela Vob and Oliver Zielinski proposed “Classifying Natural Waters with the Forel-Ule Colour Index System: Results, Applications, Correlations and Crowdsourcing” in which a change in field observation procedures is one potential source of uncertainty;for instance,personnel can affect FUI observations because the human eye's perception of color varies greatly.Water depth and FUI colors don't appear to be correlated in any obvious way. Due to the FUIscale's restriction to discrete numerical values,clustering will have an impact on correlations with other continuous water constituents.

(Objective):

This dissertation focuses on developing a water quality assessment model that takes into account many factors and determines water quality with the following goals in mind:

- To collect Landsat 8 satellite imagery of the Arabian Sea, Bay of Bengal, and Lonar vLake.
- To extract numerous metrics like as salinity, turbidity, and chlorophyll-a from Landsat 8 images
- To forecast whether the water is pure or needs treatment, algorithms such as Decision Tree and K-nearest neighbors are used.
- To increase the accuracy of the results in comparison to existing systems.

Flow of the project:

The paper explains the features and methodology of the proposed Water Quality Analysis using Google earth images. The paper gives a detailed explanation of each and every aspect of the project from introduction to references. Chapter1 presents basic information about the project that is Introduction. Chapter2 describes the literature survey and limitations of some existing systems which are overcome by the proposed system. Chapter 3 describes the proposed system, its architecture,algorithm and flow of work. Chapter 4 gives details about the result and conclusion by describing some future work which can be done to make the system more reliable and flexible.

Methodology:

The proposed work focuses on deriving qualitative water quality parameters of four regions which are the Arabian sea along the western coastal region surrounding Mumbai, Bay of Bengal, Lonarlake and coastal water near the southern Indian region. Strong seasonal oscillations in biological production are a feature of the Arabian Sea, one of the significant water bodies on earth. Goes and Gomes found that chlorophyll-a levels in the Arabian Sea have been rising gradually during the 1990s—as much as four times higher —using ocean color data from NASA in a 2020 research study.According to the literature survey, the Arabian Sea has a higher concentration of chlorophyll in comparison to the Bay of Bengal region. However, water is generally more turbid in the Indian ocean including Bay of Bengal and Arabian sea than in the Pacific and Atlantic oceans. The salinity parameter has a slight variation among each other as all the regions possess high alkaline levels.

Lonar Lake which is an inland saline crater in Maharashtra's Buldhana district has long drawn the attention of scientists working in a variety of sectors. Many researchers examined the physicochemical characteristics of water samples, which found that the water is alkaline with a high pH and classified by high concentrations of alkalinity, sulfate, sodium, total dissolved solids, magnesium, chloride, and dissolved oxygen.

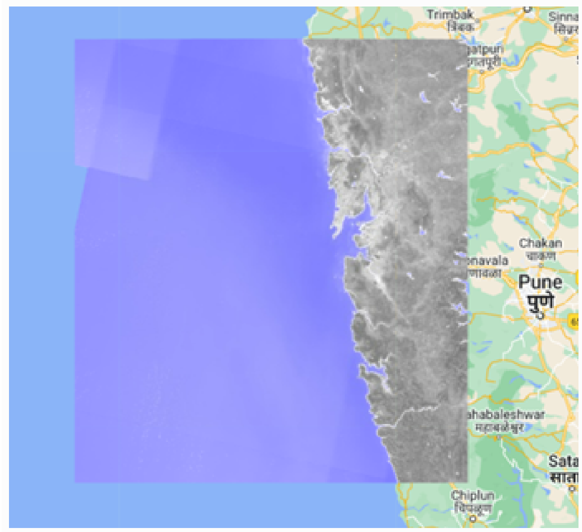

Figure 1 - Arabian Sea

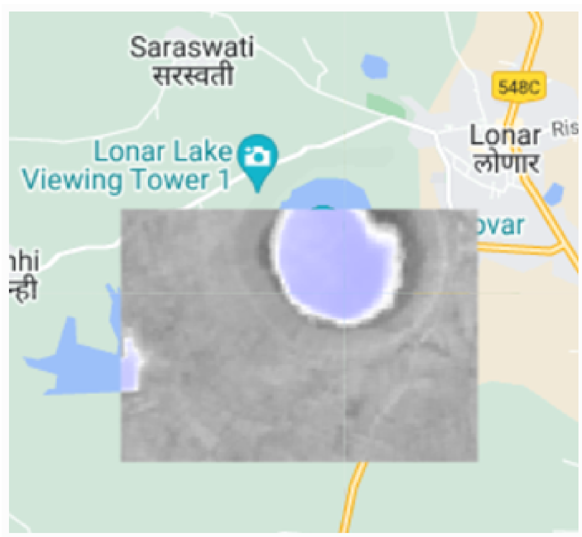

Figure 2 - Lonar Lake

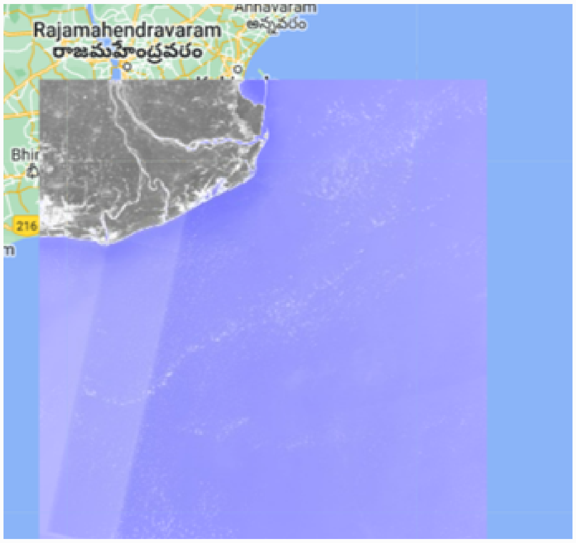

Figure 3 - Bay of Bengal

Landsat-8 data collection

Typically, satellites utilize sensors to gather electromagnetic radiation reflected from land and oceans as they orbit the Earth, some of which pass over a certain region every day. The most pertinent wavelengths for measuring the quality of water bodies are visible and near-infrared light. Following table describes various satellites with their respective sensors.

Satellites	Sensors	Resolution
Landsat 7	Enhanced Thematic Mapper (ETM+)	185 km swath; 15 m, 30 m, 60 m; 16 day revisit
Landsat 8	Operational Land Imager (OLI)	185 km swath; 15 m, 30 m, 60 m; 16 day revisit
Terra & Aqua	MODerate Resolution Imaging Spectroradiometer (MODIS)	2330 km swath; 250 m, 500 m, 1 km; 1-2 day revisit
Suomi NPP	Visible Infrared Imaging Radiometer Suite (VIIRS)	3040 km swath; 375 m – 750 m; 1-2 day revisit
Sentinel 2A and 2B	Multi Spectral Imager (MSI)	290 km swath; 10 m, 20 m, 60 m; 5 day revisit
Sentinel 3A	Ocean and Land Color Instrument (OLCI)	1270 km swath; 300 m; 27 day revisit

The proposed work focuses on data collection from Landsat 8 which contains Operational Land Imager (OLI) sensor. The image data is collected from USGS Earth Explorer. It contains Landsat 8 Collection 2 level 1 (L1) and level 2 (L2)data. While there is no need for any correction techniques, Level-2 atmospherically corrected (Surface Reflectance) data is the "ready to use data".Tier 1 data is used for processing as the USGS recommends using Tier 1 data for analysis because of its bestpositional and radiometric quality. They have been inter calibrated among the Landsat sensors and have high precision terrain processing. Landsat 8 satellite acquires data in 11 bands from two separate sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS).

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Coastal aerosol	0.43-0.45	30
Band 2 - Blue	0.45-0.51	30
Band 3 - Green	0.53-0.59	30
Band 4 - Red	0.64-0.67	30
Band 5 - Near Infrared (NIR)	0.85-0.88	30
Band 6 - Shortwave Infrared (SWIR) 1	1.57-1.65	30
Band 7 - Shortwave Infrared (SWIR) 2	2.11-2.29	30
Band 8 - Panchromatic	0.50-0.68	15
Band 9 - Cirrus	1.36-1.38	30
Band 10 - Thermal Infrared (TIRS) 1	10.6-11.19	100
Band 11 - Thermal Infrared (TIRS) 2	11.50-12.51	100

Table 2 - Landsat 8 bands

Result and Analysis:

In this section, we illustrate the results of the Water Quality Prediction using Satellite Images.

Firstly, we shortlisted algorithms based on our problem statement, which is a classification problem. For classification problems, Naive Bayes and Decision Tree/Random Forest algorithms are widely used.

Two well-known machine learning techniques for categorization issues are Naive Bayes and Decision Tree. Because even little changes in the data may result in totally different trees being formed, decision trees have the potential to produce complicated trees that do not generalize effectively and can be unstable. The Naive Bayes method relies on Bayes' theorem and makes the assumption that every pair of characteristics is independent of one another. When there are several characteristics, and they all independently contribute to the final result, Naive Bayes performs well. Compared to Naive Bayes, decision trees perform better when there is a lot of data. The Decision Trees model offers superior average values (i.e., more accuracy) for forecasting true positives and true negatives when compared to the Naive Bayes model.

We have compared Naive Bayes and Decision Tree algorithms with each other to see which algorithm performs well on the dataset we provide. We have judged the performance on the basis of accuracy. For Naive Bayes, we got the accuracy on the training dataset around 86% and 83% on the testing dataset. In the Decision Tree case, the training set is around 93% and 81% on the testing dataset. To choose the most optimal algorithm for our project, we have calculated the difference between the accuracy of training and testing dataset for both the algorithm. Since the Naive Bayes has less difference, that means it can generalize better compared to Decision Tree.

Conclusion:

In this study, we have examined the importance of high-quality water and developed methods to determine if the water is good or bad by considering many factors like pH, turbidity, chlorophyll-a. The water potability dataset was used to train the model, provide predictions, and get an understanding of how it functions. The following algorithms' effectiveness on our dataset:

1. Decision Tree:

2. Naive bayes:

In order to determine if the parameters of the water sample being tested make it excellent quality water or not, we successfully developed a model and tested it. By increasing the algorithm's precision, the project can be further improved. Also, it has been highlighted that there is an increase in the release of sewage from homes, industrial effluents, and air inputs from burning fossil fuels and bushfires, all of which contribute to the decline in water quality. In order to achieve accurate water quality prediction, more research is needed to enhance the system's performance and to concentrate on the availability of well-labeled information.

References:

1. https://www.intechopen.com/chapters/77416
2. https://en.wikipedia.org/wiki/Bay_of_Bengal
3. https://www.sciencedirect.com/science/article/abs/pii/S1385110115000143?via%3Dihub
4. https://en.wikipedia.org/wiki/Arabian_Sea#Arabian_Sea_warming
5. https://en.wikipedia.org/wiki/Lonar_Lake#Threats_to_Lonar_lake
6.