IoT and Satellite Image Driven Water Quality Monitoring and Assessment Method in Coastal Region



Jasrin Shiddike
Department of Computer Science and Engineering
United International University

A project submitted for the degree of
MSc in Computer Science & Engineering
July 2023

Declaration

I, Jasrin Shiddike, declare that this project titled, IoT and Satellite Image Driven Water Quality Monitoring and Assessment Method in Coastal Region and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a MSc degree at United International University.
- Where any part of this project has previously been submitted for a degree or any other qualification at United International University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this project is entirely my own work.
- I have acknowledged all main sources of help.
- Where the project is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

Jasrin Shiddike

ID: 012211014

Certificate

I do hereby declare that the research works embodied in this project entitled IoT and Satellite Image Driven Water Quality Monitoring and Assessment Method in Coastal Region is the outcome of an original work carried out by Jasrin Shiddike under my supervision.

I further certify that the dissertation meets the requirements and the standard for the degree of MSc in Computer Science and Engineering.

Signed: _____

Date: _____

Dr. A.K.M. Muzahidul Islam

Professor & Director-CIAC

Department of Computer Science and Engineering,

United International University,

Dhaka-1209, Bangladesh.

Abstract

Water is a gift of creator and fundamental resource of human life, agriculture and industry. As a result of the combined impacts of growing population, climate change, pollution, high financial demands, and excessive salt extraction from the sea, safe water is becoming threat to get from natural resource. Some unambiguous and comprehensive coordination of planning and sustainable action must take to ensure, access to freshwater for urban and rural people. So, it must have a coordinated and sustainable system in place to periodically check on water purity. Monitoring water quality is essential, especially for remote areas. This research leads to monitor and identify water quality in real-time without human interaction to assess whether the water is drinkable and domestic usage. In this research, we have designed an IoT-driven system that consists of 5 water parameter sensors such as temperature, EC (Electric Conductivity), pH, turbidity, and TDS (Total Dissolved Solids) for assessing the water quality. The obtained IoT sensor data is sent to the cloud server so that we can monitor the water quality in real-time. We have also collected 200 satellite map data of both Sentinel-2 and Landsat-8 satellites jointly from EarthExplorer to characterize the water quality in two categories such as freshwater and seawater. So, we have applied the Deep Learning method CNN (Convolutional Neural Network) for the model build and predict the categories of water individually based on the map area. For analyzing the past data, we have taken 3180 ideal water data from the WHO to examine the water quality by considering five water parameters. Then we have built our model by applying some machine learning algorithms such as Decision Tree, Random Forest, Extra Tree, and KNN Classification to find good accuracy which are 85.29%, 85.66%, 84.75%, and 79.85% respectively. Following that, we cross-match the results depending on location by comparing the satellite data with the IoT real-time data which is shown in the web interface.

Acknowledgements

First and foremost, I would really like to express my gratitude to Almighty Allah (SWT) for giving me the strength to complete my task.

I want to express my admiration and acknowledgment to my supervisor Dr. A.K.M. Muzahidul Islam, Professor of the Department of Computer Science & Engineering, for his excellent supervision, direction, motivation, and support to make this project possible. My project and reports are supervised by his supervision and endorsement. I would also like to convey my sincere gratitude to my reviewer Dr. Md. Motaharul Islam, Professor of the Department of Computer Science & Engineering, for his time and effort in providing insightful comments which helps me to improve this report. The members of my committee are also to be commended for making the experience of my defense excellent and for their remarks and suggestions.

I want to express my thankfulness to all of the faculty members of United International University for supporting my efforts and accepting my work.

Last but not the least, I want to thank my family and friends for being so kind towards me and their continuous support for me while I conducted my research and wrote my project report.

Contents

List of Figures viii

List of Tables ix

1 Introduction 1
- 1.1 Meaning of Satellite image, IoT, and Data Analytics . . . 1
- 1.2 Present State of the Problem . . . 2
- 1.3 Motivation . . . 2
- 1.4 Study Region . . . 3
- 1.5 Challenges . . . 3
- 1.6 Objectives . . . 4
- 1.7 Our Contribution . . . 5

2 Related Work 7
- 2.1 Background . . . 7
- 2.2 Literature Review . . . 8

3 Methodology 15
- 3.1 Sensor Layer . . . 16
  - 3.1.1 Semantic Diagram . . . 18
- 3.2 Network Layer . . . 19
- 3.3 Data Processing Layer . . . 19
  - 3.3.1 Cloud Server . . . 19
  - 3.3.2 Data Pre-processing Techniques . . . 20
- 3.4 Application Layer . . . 24

4 Implementation & Results 25
- 4.1 IoT Result . . . 25
  - 4.1.1 Hardware Setup . . . 25
  - 4.1.2 ThingSpeak Server . . . 28
- 4.2 Analysis of Past Data . . . 29
- 4.3 Satellite Image Processing . . . 33
- 4.4 Web Interface . . . 35

4.5 Discussion 36

5 Conclusion and Future Work **37**

Bibliography **39**

List of Figures

1.1 System Overview 5

3.1 Taxonomy of our proposed system 15
3.2 pH Scale 16
3.3 Hardware Sensor 17
3.4 Schematic Diagram 18
3.5 ThingSpeak Channel Id & API key 19
3.6 Input Data of Water 20
3.7 Location of the study area 22
3.8 Input image 22
3.9 Flowchart of Proposed System 24

4.1 Initial Hardware Setup 27
4.2 Final setup for system testing 27
4.3 Sensor data in ThingSpeak server 28
4.4 Last IoT data 29
4.5 Mean, Standard Deviation, Q1, and Q3 29
4.6 Data Description 30
4.7 Processed input data 30
4.8 Target Value 31
4.9 Boxplot 31
4.10 Comparison accuracy of the model 32
4.11 Accuracy result of Random Forest Classification 32
4.12 Final output of IoT 33
4.13 Categorized Data 33
4.14 Reshape Data 33
4.15 Visualization of image data 34
4.16 Loss of validation data 34
4.17 Accuracy of the output image 35
4.18 Web Interface 35

List of Tables

2.1 Summary table of related research 12

3.1 Recommended water parameters by WHO 23

4.1 Connection of D1 mini and TDS sensor 25
4.2 Connection of D1 mini and Temperature sensor 26
4.3 Connection of D1 mini and Turbidity and ADS1115 sensor 26
4.4 Connection of D1 mini and pH and ADS1115 sensor 26

Chapter 1

Introduction

Water is one of the most significant natural resources that nature has provided to humanity. Water is a key component for both industrial and agricultural development that cannot be ignored. Water is also critically necessary for all living things, including humans, for economic growth, reduction of poverty, and food security, as well as for animals, plants, sustainability of ecological processes, and other life forms [1]. The availability of water is also essential due to the high demands brought on by the rapid population growth and global warming, groundwater has been decreasing [2]. Given that the global population is predicted to exceed eight billion people by 2025, there will undoubtedly be competition for scarce resources due to increased needs for drinking water supplies and water for crop production [3].

1.1 Meaning of Satellite image, IoT, and Data Analytics

Satellite image processing is a type of remote sensing that processes the captured images of the earth and satellites using artificial satellites which is a significant area of research and development. It is one of the important computing techniques used in the military, and agriculture, preventing natural disasters, identifying natural resources, and other fields. Digital photos are first captured, and then information is extracted from them through computer processing. After applying statistical techniques to digital photos, it is possible to identify the different discrete surfaces by looking at the pixel values and also using pixel resolutions to gather coherent data about the surface of the earth. As a result, satellite image processing has a wide range of uses in astronomy, remote sensing, research and development, and now large-scale cloud computing [4].

Another remote sensing technique is the IoT (Internet of Things) which is physical objects and things of a network, including machines, buildings, and vehicles that are equipped with sensors, micro-controllers, and network connectivity. It can be described as a wide network of embedded applications designed with wireless technologies which are already built in to enable monitoring, controlling, and linking inside the current internet infrastructure [5]. Each sensor must be able to autonomously collect real-time

data and have a distinct identity. IoT's fundamental building components include sensors, processors, gateways, and applications. The sensors' data is gathered, stored, and analyzed by the cloud services, which then empowers users to make informed decisions.

Data analytic use a variety of software tools, including spreadsheets, reporting tools, visualization of data, data mining software, or open-source software platforms, to modify data to the greatest extent possible. In the modern era, data analytic is the process of analyzing raw data to derive useful, and practical insights. It consists of a range of approaches, innovations, and tools for analyzing patterns in statistical data and resolving problems. It can also influence data for enhancing decision-making [6]. Data analytic can be applied to forecast, assess, and improve organizational performance. Advanced data analysis approaches are used to characterize, forecast, and address present and future difficulties, including data modeling, data mining, data transformation, etc.

1.2 Present State of the Problem

Water needs are anticipated to increase due to the expected repercussions of global climate change, including higher temperatures, drought, more unpredictable precipitation patterns, and more intense storms. Ground and surface water supplies are both being impacted by agricultural water withdrawal. Although the importance of reducing water consumption cannot be overstated, considerable efforts for sustainable water are focused on the development of novel cleaning and recycling technologies [7]. Another issue is the climate crisis which is one of the most repeated crucial moments that is happening all over the world, especially in the coastal region. Because the water level is rising day by day and the freshwater is converted into saline water which is unsuitable for human consumption and agricultural purposes and it also reduces the amount of freshwater availability in coastal areas. For that reason, people face water crisis issues globally. Due to high salt and turbidity, the water quality of rivers around the coastal region has deteriorated, making the water from those rivers frequently inappropriate for residential and agricultural use, particularly in the dry and winter season. Water quality is one of the biggest issues people will likely confront in the twenty-first century.

1.3 Motivation

Every year, the WHO conducts a survey to find out how many people are affected by drinking unhygienic contaminated water worldwide. There is a misconception that clean water is of high quality, which points to a knowledge gap regarding the existence of these compounds in water [8]. Contaminated water has an effect on all facets of human existence, including mental, social, economic, physical, and emotional growth. Every year, illnesses brought on by tainted drinking water claim a million lives, the bulk of them being children. As one of the key elements in regulating health and the status of diseases in people, aquatic animals, and many agricultural lands, it is important to determine the water quality [9].

Water availability and sustainable management are difficult to ensure in a resource-poor nation like Bangladesh, where development is one of the most difficult areas to achieve. However, Bangladesh has made significant progress toward ensuring that everyone has access to better water sources. Although coverage has considerably increased, issues with water quality and equal access in remote places still exist. The poorest and most vulnerable people in Bangladesh are those who continue to lack access to clean water. The three main problems with water quality in Bangladesh are salinity, microbial infections, and arsenic (As) in groundwater. It is impossible to decrease maternal, neonatal, and child morbidity and mortality, enhance nutrition, increase school attendance, or free up time for beneficial activities without access to safe drinking water [10]. Bangladesh is a low-lying deltaic nation with an extraordinarily high population density, making it vulnerable to various environmental pressures and natural disasters that might make it more difficult to acquire potable water. The provision of potable water to the general public is a crucial factor and the primary duty of modern public health engineering. Safe to drink, pleasant, and appropriate for home use are all characteristics of potable water. Low-quality water intake is responsible for around 80% of all diseases, two-thirds of fatalities, and 10% of lost productivity worldwide, particularly in developing nations like Bangladesh [11]. In addition to aquatic life, poor water quality affects the local environment. In Bangladesh, inadequate governance, poor infrastructure, and improper water management all have a significant impact on water security, making it difficult to meet both drinking water and agriculture needs.

1.4 Study Region

Our study regions are the Karnafuli river and its tributaries, such as the Halda river, the lakes of Rangamati, and the nearest area of the Bay of Bengal which is situated in Chattogram.
Karnafuli is the main river in Chattogram, which originates in the Lushai highlands of Mizoram state and flows for about 270 km across Bangladesh's southeast before emptying into the Bay of Bengal. This river's winding route passes multiple bridges, the Halda-Karnafuli confluence, and the Kaptai hydroelectric station. The river is approximately 45 km long from Kaptai Lake to the Halda-Karnafuli junction and 30 km long from the Halda-Karnafuli junction to BN Academy. The Karnafuli River has semi-diurnal features and is tidal. During times of flooding, the flow from the Halda River travels far upstream and passes extremely close to the Kaptai Dam on the upper Karnafuli River [12].

1.5 Challenges

To determine status and trends, monitoring of water quality is the standardized measuring, long-term, and observation of the aquatic environment [13]. In order to preserve human health and well-being and keep up with rapid environmental change, regular water quality monitoring is becoming more and more essential [14]. Because there are

so many aspects to take into account, monitoring water quality is still a very difficult procedure. The conventional techniques for water monitoring are accurate and trustworthy. The conventional method of assessing the quality of the water involves manually collecting samples and sending them to a lab for testing and analysis. This strategy is not only inefficient in terms of time, people, and resources, but also expensive. The use of remote sensing has been suggested as a possible method for keeping track of a number of water quality indicators without the need for on-site testing. In order to categorize water for its appropriate application and to minimize waste, continuous monitoring of the water quality is required. For instance, water that shouldn't be consumed can be utilized to clean. So, the present requirement is for a portable, sensitive, and selective sensor for continuous water quality monitoring.
Remote sensing technologies have advanced quickly in recent years. This implies that it is now much easier to acquire a large collection of high-resolution remote-sensing image data [15]. One remote sensing technique is satellite images which have been expanding use in key sectors including weather, forestry, and agriculture as well as surface changes and biodiversity monitoring, and users from a range of sectors, including industry and government, have begun to pay more attention to satellite imagery. [16].
Another remote sensing technique is an IoT-based approach for monitoring water quality that is quick, affordable, and provides results in real-time. Consequently, there is a growing demand to provide quick, cheap, portable sensors with great sensitivity and dependability. A good sensor would be reasonably priced, have excellent sensitivity and selectivity, have high throughput, be easy to use, and have the flexibility to operate in the field. Additionally, it must adhere to WHO regulations for portable sensors. Because of recent developments in the IoT, which can be used to create more effective, secure, and affordable systems with real-time capabilities. The goal of water quality monitoring using IoT is to deliver reliable decision support systems that direct societal and technological advancement in examining water quality and its utilization. One of the UN's sustainable development goals for 2030 is water security [17].
Some research challenges we have faced during our data collection and industrial sensors collection period. When we have collected satellite images the water map area was not visible because of cloud coverage. So, we need to reduce the cloud coverage between (0-5) ranges for the purpose of identifying the water area. We have also purchased waterproof industrial sensors which are rarely available in the Bangladeshi market. So, we have to wait for a long time to collect all the sensors and access the sensors which is another difficult task for our research.

1.6 Objectives

Our main goal is to design a system with great detection accuracy, low power consumption, time efficiency, and low cost for continuously monitoring water quality in remote places. Figure 1.1 shows the overview of our proposed system.

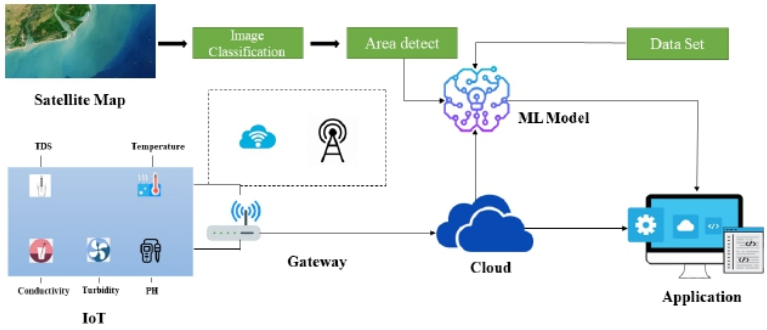


Figure 1.1: System Overview

The main objectives of this research are-

1. To apply Data analytics and Deep learning approach to monitor water quality by analyzing ideal water data and satellite images.
2. To predict the category of water by using IoT technology for taking measurements of water parameters such as temperature, turbidity, pH, conductivity, and TDS.
3. To find better accuracy and adaptation.

1.7 Our Contribution

- Analyzing the WHO ideal water quality index and building some ML techniques on statistical data to predict the category of water with good accuracy.
- CNN model is employed on satellite image data to build a model and classify the images with the category of freshwater and seawater regions.
- Combination of three different fields such as IoT, Image Processing, and Data Analytics techniques together to monitor water quality which is unique.
- Compare both the satellite and IoT results to analyze and crossmatch the result shown in web interface.

For our system, we have collected satellite map data including Landsat-8 and Sentinel-2 satellite images which are taken from "Earth Explorer". To process the image data and build a model we have applied a Deep Learning technique which is Convolutional Neural Networks (CNN) to obtain knowledge from image data and the result of this model predicts the map is freshwater or seawater based on their location. We have also applied data analytical techniques on ideal WHO water data to examine the water quality by analyzing the statistical view of water parameters. To process the water data and build models we have applied some machine learning techniques such as Decision Tree, Random Forest, Extra Tree and KNN classifier to predict the water quality

which is in two categories such as freshwater and sea water. We have also designed an IoT-based architecture where we have considered five water parameters such as temperature, turbidity, pH, conductivity, and TDS to examine and read the water quality live and also collect the real-time data by a centralized cloud-based system which is ThingSpeak. The real data is also tested on the model and classify the category of water quality. Then we compare both the satellite data and IoT real-time data to analyze and crossmatch the result based on location. The whole solution is based on the coastal area. The uniqueness of our work is that we have combined three different fields together to monitor water quality and also get real-time data without human interaction which is cost-effective and time-efficient.

The report's remaining sections are arranged as follows. The literature review is described in Chapter 2, and the methodology is shown in Chapter 3. The implementation & results are illustrated in Chapter 4 along with a discussion. After discussion, we have concluded or finished our work in Chapter 5.

Chapter 2

Related Work

2.1 Background

All of humanity's demands, including agricultural, domestic, industrial, municipal, and residential, use just 1% of the water that the globe has to provide.In recent decades, the quality of the water has been negatively impacted by a number of factors, including population growth, rapid industrialisation, increased urbanization, and careless use of natural resources. [18]. Water crisis is a worldwide issue because of global warming and pollution. Due to climate crisis sea water level is rising and freshwater converted into saline water in coastal areas. Water contamination is another big issue worldwide that may take human lives. Natural water has different types and quantities of contaminants depending on the location, season, and other variables. The availability of drinkable water is steadily declining as a result of pollution. There are several different causes of water contamination including wastes from domestic and industries. There are two main causes of the harmful effects of water pollution: first, the amount of toxins in the water is constantly rising, and there are currently no effective techniques to measure the natural changes; and second, human evolution, which usually implies the need for the creation of new water sources with uncertain quality [19]. Monitoring the water quality on a regular basis becomes crucial for this aim.
Bangladesh's environmental characteristics make it partially susceptible to water insecurity. Due to both man-made and natural reasons, coastal Bangladesh has long been known to have a serious salinity issue with its water supply. Bangladesh's rivers have a significant impact on both the way of life and the culture of the population. For the aforementioned reason, it is essential to frequently monitor water quality measurements in order to detect any changes in those parameters in real-time and confirm the water is safe. Water quality monitoring devices must quickly identify any changes in the water's quality so that authorities can take fast action [20].

2.2 Literature Review

In the literature review, we have chosen a large number of related research that are based on IoT, Satellite imageries, and smart water monitoring systems. We have discussed the article below-

Mohammed et al. [21] designed a deep learning-based system that was used to classify the saline particles of seawater. The main focus of this research was to desalinate and reuse water for the purpose of human consumption and agricultural sectors. They used current edge-based techniques such as Convolution Neural Networks with transfer learning so that they could classify the salinity from the water with less effort and also minimize the cost. They also applied data augmentation techniques for the purpose of robust output results. The research gap of this paper is that they were not discussed the detection of saline particles from seawater and also did not mention any accurate desalination techniques.

An interactive shower panel system was designed by Adrian et al. [22] for discussing the concept of global water crisis based on IoT and cloud computing and also provided solutions for water management. Basically, the system measured the liquid flow of water based on statistics of water consumption and sent the data to the user's smartphone using a cloud server. The main objectives of this research were to educate the people about the amount of water consumption and also save the resource for the sustainability purpose. They mainly focused on the economy of Poland and its key sectors, for example, agriculture and energy. This system monitors the real-time water consumption and also informs the user if the limit was exceeded. The limitation of this article is that they were discussed the protection of water resources using shower panels but running the system users should be well-educated and they also did not discuss the recycling process of water.

Freshwater availability and groundwater quality were measured by the association of some geomorphic features like paleochannels and sand dunes in coastal areas which were designed by Prusty et al. [23]. The major goals of this study were to establish a link between the mechanisms that regulate groundwater salinity in groundwater sources and the degree of seawater-freshwater mixing at the surface, as well as to identify the most likely freshwater source at a given well level in the seawater-affected locations. An innovative multi-dimensional approach that combines hydrochemical tools with GIS and remote sensing technologies was used to comprehend the distribution of freshwater and seawater. Additionally, geomorphic characteristics and their potential connection to the distribution of freshwater were found using satellite images of the region. The results of this study would contribute to better management of coastal groundwater resources, which would benefit inhabitants quality of life and health. The research gap of this article is that they only discussed the groundwater resources but there was no discussion about the process of saline water and converting the saline water into freshwater for human consumption and agricultural purposes.

The main goal of this paper [24] was to develop a Sentinel-2 and Landsat-8 satellite pre-processing framework for an optical multi-sensor technique in La Palma. Addition-

ally, they discussed the important water quality indicators of chlorophyll-a concentration, and turbidity at fine-scale bio-optical features and temporal gradients throughout the research phase. Using the retrieval products, examine and detect the main modifications and crucial areas by considering the most recent volcanic disaster. Their main contribution was the use of high-resolution satellite technology, which might be used to quickly and effectively lessen the harmful effects on coastal waters after these catastrophic events, to facilitate security and preventive measures, and to make vulnerable locations accessible in an emergency. The limitation of this research is analysis-based where they examined the water quality based on the satellite image.

The main goal was to develop a system for wireless sensor networks based on the Bangladeshi river within cheap cost, low power use, and had good detection accuracy for monitoring river water quality constantly in remote regions [25]. They measured several water characteristics, including pH, conductivity, dissolved oxygen, turbidity, and others using adjacent sensors to keep an eye on the condition of the water. They applied an IoT-integrated big data analytics approach which could offer dependability, scalability, speed, and permanence. It appeared to be a better solution. They also contain Deep learning neural network models, the Belief Rule Based (BRB) system, and Spark streaming analysis through Spark MLlib, all of which would be used during the phase of project development. The research's limitation is that it just assessed the quality of river water and did not go further to consider an efficient water management system.

Sathish et al. [26] developed an smart water quality monitoring system based on IoT that was more improved usability, cost-effective, and provides real-time feedback. The system that was suggested in this article makes use a number of sensors to monitor a wide range of variables, including water turbidity, pH level, temperature, tank level, and air humidity. Additionally, the Microcontroller Unit was connected to these sensors (MCU). As part of this operation to examine the properties of water, which was the research gap of this article, some more water parameters were also needed, such as dissolved oxygen, free residual chlorine, nitrates, and electrical conductivity of water.

A review paper was discussed by Joshua et al. [27] based on IoT for monitoring water quality and assessing whether the water was pure or not. This article's main contribution was a summary of research developments, a discussion of recent advancements, and a list of intriguing and difficult problems for future in-depth investigation. They highlighted the benefits of IoT for assessing and monitoring water quality, including the potential for real-time monitoring, automation for smart solutions, adaptable and responsive systems, and lower costs for monitoring water quality. This article summarizes different water parameter values in different locations based on real-time IoT data which was reviewed from different articles to measure and monitor the water quality. The lack of discussion of various sensor technologies to identify the vast range of characteristics that could accurately reflect water quality, as well as a lack of attention to design costs, constitutes a research gap in this paper.

In order to forecast and assess water quality, Ping et al. [28] developed an IoT-based long short-term memory (LSTM) deep neural networks technique. The deep learning

(DL) theory was used in this study to design and construct a model to forecast water quality big data using advanced deep learning techniques that were highly performed in time-series prediction. In order to develop a forecasting approach for water quality using LSTM deep neural networks, the autonomous water quality monitoring station data from the Guangzhou Source Of water of the Yangtze River in Yangzhou was prepared. The drawback of this study was that it avoided multi-dimensional input records that could precisely forecast the target parameters and focused less on simulation by merging the several water quality measures.

In this research, [29] they introduced a new Supervisory Control and Data Acquisition (SCADA) system which was improved with more sensors and lower prices for monitoring water quality in real-time that combines with the Internet of Things (IoT) technology. The main objectives of this system were to record the precise, real-time sensor data such as the water flow, temperature, the color of water, the level of contamination of water, pipeline leaks, and automatic measurement of parameters (pressure, pH, water level, and energy sensors) to process sensor data in real-time Module for the Global System for Mobile Communications (GSM). Here, the SCADA via wireless GSM module was used to continuously monitor and manage the water quality effectively at analyzing and detecting water contamination by measuring physical and chemical characteristics of water across several stations to examine the water's condition. The lack of this research is unable to apply real-time data analytics techniques with an ON/OFF capability in water distribution stations for real-time data processing. Furthermore, the research did not address the need to secure wireless networks that were used for the transfer of sensor data.

A complete picture of the subject was provided through the use of remote sensing and GIS in associated with test datasets studies including vertical electrical soundings (VES) and tests of water quality to identify the routes for saline water incursion in the controlled aquifers which was discussed by Ratnakar et al. [30]. In this article, they showed some indirect evidence of land sinking provided by geomorphic units such as paleo-channel sinuosity, beach ridges, backwater inundation, and tidal flats which was confirmed by subsurface geo-electrical signals collected using VES and on-site measurements of groundwater quality. The goals of this study were to use GIS and remote sensing techniques to map the precise geomorphic units, pinpoint the main causes of saline water, and also delineate the freshwater zones in the river systems to prevent invasions. The research gap of this system is unable to make decisions, the protection of water, and could not give real-time data to manage groundwater.

This paper [31] investigates typical monitoring water quality metrics, their drinking water safety limitations, linked with smart sensors, a critical evaluation, and the approval of modern IoT-WQMS, used of an observational metric that is proposed, analysis, discussion, and, finally, design recommendations for an effective system. They discussed a brief overview of the freshwater situation, including its causes, pollution, subsurface water depletion, and saltwater desalination. Additionally, they provided information on the WHO-recommended water quality index, widely-used water quality index

parameters, safe water consumption guidelines, and a typical system which was an IoT-based water quality monitoring for the analysis of current IoT-WQMS for household water.

This article [32] used chemical sensors to monitor water quality which provides a real-time critical assessment of the applicability of various technologies. In particular, the effectiveness of molecularly imprinted polymer-based sensors was thoroughly assessed, together with information on their operating principle, sustainability in real-world on-site applications, and options for mass production. This paper offers a critical evaluation of recent developments in real-time chemical sensor-based water quality monitoring, which underwent online testing using actual water samples, and their viability for long-term use was taken into consideration. The limitation of this article is to only consider chemical sensors which were very costly and the process might take long time.

This research offered a real-time IoT-based framework to monitor water quality, analyze, and alert for taking action based on pollution and dangerous component levels, sensor, and application performance, as indicated by Akashdeep et al. [33]. In this proposed work, trends in water quality as well as the architecture of device administration and monitoring were analyzed by machine learning models. They applied different sensors, like a Raspberry Pi, and cloud servers were used to execute the suggested water quality and device monitoring framework for monitoring, data storing, processing, and alerting. In this work, water turbidity was forecasted and examined using a variety of variables utilizing machine learning techniques. This article's primary goal was to create accessible, efficient water management solutions employing specialized IoT architecture designs coupled with devices, sensors, and embedded probes of IoT with cutting-edge cloud-based technologies. The research gap of this article is to examine fewer sensor data for monitoring water quality.

This review article was suggested by Małgorzata et al. [34], to introduce the concept of IoT and water parameters for monitoring water quality. The purpose of this work was to introduce Internet of Things (IoT) principles for assessing water quality and potential components that might be linked to it. Here they explored the applications, advantages, solutions, challenges, and efficiency of IoT to determine water quality. They discussed the different types of water sensors to measure water quality which may provide real-time data to identify the water situations. Additionally, the Internet of Things (IoT) could be used to automatically collect data from a wide range of sources, which could assist give a more comprehensive view of water quality. They were also talking about the costs of sensors, standard datasets, and social awareness which may increase the efficiency of IoT. The research gap in this article is that they were not implementing the system.

Peng et al. [35] proposed a system to monitor the water quality of the urban river. This paper's main goal was to provide a technique for realizing the total inversion of urban rivers using only a few measured samples, in order to establish a solid foundation for pollution traceability. They developed a self-optimizing ML technique to achieve

automated adjustment of the model parameters in order to increase the model's inversion accuracy and predictive power which was the main contribution of this article. The input portion of the model used feature augmentation to improve the suitability of the input features for model training. Due to the various nonlinear characteristics of these data, which further complicated the issue of the quantity and quality of multi-source data, the spatial mapping technique was also used to ensure consistency in the water quality information. The limitation of this article is that they only discussed the urban river water quality which may be more required for rural areas.

A combined model was developed by Guimei et al. [36] to evaluate and predict water quality. The main goal of this article was to check the water quality status by proving the efficiency of the system and generalizing the prediction and evaluation model. The evaluation model and the significant information on the water quality of the two basins were established using a T-S fuzzy neural network. In this research, they suggested the ARIMA-WNN model which was a novel combined model to establish the combined prediction system for each water quality index of the two basins and evaluate the prediction outcomes against the combined model for the prediction of water quality. By using this combined model, they achieved good accuracy than the original one which was the main contribution of this article.

The previous research work has been summarized and we have specified their main findings which is shown in Table 2.1.

Table 2.1: Summary table of related research

Author	Title	Source	Findings
Mohammed et al. [2021]	Deep Learning Based Approach to Classify Saline Particles in Sea Water	MDPI (water)	To classify the salinity from the water with less effort and minimize the cost and also desalinate and reuse water for the purpose of human consumption and agricultural sectors
Adrian et al. [2021]	Global Water Crisis: Concept of a New Interactive Shower Panel Based on IoT and Cloud Computing for Rational Water Consumption	MDPI (Applied Sciences)	To educate the people about the amount of water consumption and also save the resource for the sustainability
Prusty et al. [2020]	Association of geomorphic features with groundwater quality and freshwater availability in coastal regions	International Journal of Environmental Science and Technology (Springer)	To establish a link between the mechanisms that regulate groundwater salinity in groundwater sources and the degree of seawater-freshwater mixing at the surface, as well as to identify the most likely freshwater source at a given well level in the seawater-affected locations

Isabel et al. [2022]	Water quality monitoring with Sentinel-2 and Landsat-8 satellites during the 2021 volcanic eruption in La Palma (Canary Islands)	ELSEVIER (Science of The Total Environment)	Use of high-resolution satellite technology, which might be used to quickly and effectively lessen the harmful effects on coastal waters after these catastrophic events, to facilitate security and preventive measures, and to make vulnerable locations accessible in an emergency
Mohammad et al. [2019]	IoT Based Real-time River Water Quality Monitoring System	ELSEVIER (Procedia Computer Science)	To create a wireless sensor network with low cost, low power consumption, and high detection accuracy to continuously monitor river water quality remotely
Sathish et al. [2020]	Smart water quality monitoring system with cost-effective using IoT	Heliyon (CellPress)	To create a smart monitoring system that is cost-effective and more efficient to use and gives real-time feedback
Joshua et al. [2020]	Internet of Things for Water Quality Monitoring and Assessment: A Comprehensive Review	Springer Nature (Studies in Computational Intelligence)	highlight the benefits of IoT for assessing and monitoring water quality, including the potential for real-time monitoring, automation for smart solutions, adaptable and responsive systems, and lower costs for monitoring water quality
Ping et al. [2019]	Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment	MDPI (Sustainability)	To create and construct a model to detect water quality using big data with the advanced deep learning techniques which is strongly performed in time-series prediction
Krishnann et al. [2018]	Real-time water quality monitoring using Internet of Things in SCADA	Springer (Environmental Monitoring and Assessment)	To record the precise, real-time sensor data to process data in real-time Module for the Global System for Mobile Communications (GSM)
Ratnakar et al. [2016]	Demarcating saline water intrusion pathways using remote sensing, GIS and geophysical techniques in structurally controlled coastal aquifers in Southern India	Springer (Environmental Earth Sciences)	Use GIS and remote sensing techniques to map the precise geomorphic units, pinpoint the main causes of saline water, and also delineate the freshwater zones in the river systems to prevent invasions
Farmanullah et al. [2021]	IoT Based Smart Water Quality Monitoring: Recent Techniques, Trends and Challenges for Domestic Applications	MDPI (Water)	Discuss the summary of different water quality monitoring systems for the analysis of current IoT-WQMS for household water
Irina et al. [2020]	Real-Time Water Quality Monitoring with Chemical Sensors	MDPI (sensors)	To monitor water quality which gives a critical evaluation of the suitability of different approaches in real-time

Akashdeep et al. [2022]	Smart IoT and Machine Learning based Framework for Water Quality Assessment and Device Component Monitoring	Springer Nature (Environmental Science and Pollution Research)	To create accessible, efficient water management solutions employing specialized IoT architecture designs coupled with devices, sensors, and embedded probes of IoT with cutting-edge cloud-based technologies
Małgorzata et al. [2023]	IoT in Water Quality Monitoring—Are We Really Here?	MDPI (Sensors)	To introduce Internet of Things (IoT) principles for assessing water quality and potential components that can be connected with IoT
Peng et al. [2022]	Urban river water quality monitoring based on self-optimizing machine learning method using multi-source remote sensing data	ELSEVIER (Ecological Indicators)	To provide a technique for realizing the total inversion of urban rivers using only a few measured samples, in order to establish a solid foundation for pollution traceability
Guimei et al. [2023]	Water Quality Evaluation and Prediction Based on a Combined Model	MDPI (Applied Sciences)	To check the water quality status by proving the efficiency of the system and generalizing the prediction and evaluation model

The main difference between the previous work and our work is that previous work only focused on IoT or data analytics or satellite image to measure the water quality where we have combined three different fields to measure the water quality based on coastal areas of Chittagong. We have also considered five basic water parameters e.g., temperature, TDS, pH, conductivity and turbidity to measure water quality which is very essential for our work. Here we have measured the IoT-integrated sensors data to get real-time feedback which is in low-cost and time-efficient and do not need any human interaction.

Chapter 3

Methodology

Climate change is most vulnerable global crisis in recent and particularly coastal people is in top risk to source fresh water due to increase of salinity in coast and its surface area. From North to South block of land area, most south is Chittagong and our research area concentrates on Chittagong coastline. Our primary goal is to develop a system that can continuously monitor water quality remotely by engaging wireless sensor networks with high detection accuracy, cheap cost, and low power consumption. We have considered five parameters for our water quality monitoring system which are temperate, TDS (Total Dissolved Solids), pH, turbidity, and conductivity to determine whether the sample water is freshwater or sea water. Our proposed system is divided into four layer architecture where we have discussed about the IoT sensors, past data processing, machine learning models, and satellite image processing technique to classify the water quality. Figure 3.1 represents the taxonomy of our proposed system.

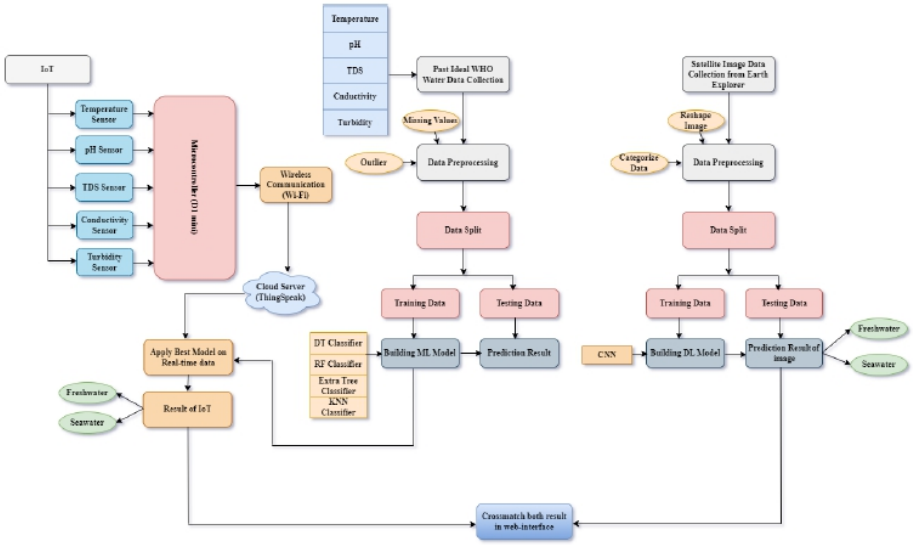


Figure 3.1: Taxonomy of our proposed system

3.1 Sensor Layer

In order to ensure that the water is safe to drink, numerous water parameters have been examined using various types of sensors. These parameters are detailed as follows:

Temperature: The DS18B20 is a specific type of waterproof electronic temperature sensor that measures air or water temperature and displays values in the 9- to 12-bit range. Using a single data line, this sensor can communicate with an internal CPU that uses the one-wire bus protocol. Because it receives its power straight from the data line, this sensor doesn't need an external power source. There are three wires on this sensor: output, power, and ground —the sensor's output voltage changes when the water's temperature changes. The sensor's output voltage is inversely proportional to temperature [37]. As a result, sensor output voltage increases along with the temperature.

pH: The term "pH" stands for "power of hydrogen," and it refers to a measurement of the body's hydrogen ion concentration. The overall pH scale has a range of 1 to 14, with 7 being regarded as neutral. A solution is described as acidic if its pH is less than 7, and primary or alkaline if its pH is greater than 7 [38]. Figure 3.2 displays the pH scale.

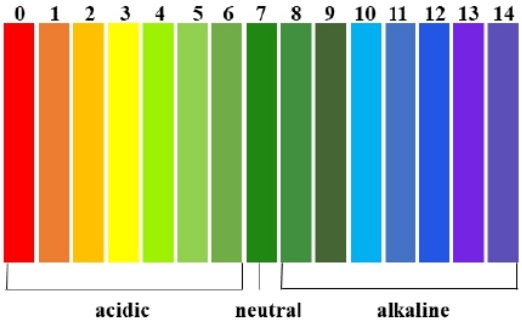


Figure 3.2: pH Scale

Here we have used the SEN0161 pH sensor to determine the pH level of water. The three pins on the pH sensor are for power, ground, and output. The pH sensor and the WeMos D1 mini development board are connected by the pin.

TDS: TDS is a metric used to determine how many soluble substances are dissolved in water. The TDS value and the quantity of soluble particles dissolved in water are typically used as measures of water purity. As a result, one benchmark for assessing the purity of the water is the TDS value. TDS is compatible with both 5V and 3V control systems and circuit boards due to its broad voltage input range of 3.3 to 5.5V. For prolonged measurement, the TDS probe can be submerged in water because it is

waterproof. This sensor can be used in water quality applications including hydroponics and domestic water analysis [39]. Here we have used the SEN0244 TDS sensor to determine the solid particle in water.

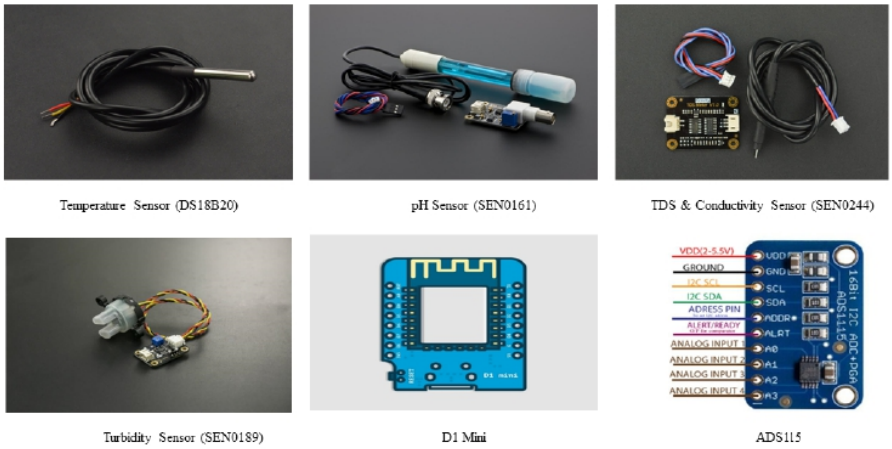


Figure 3.3: Hardware Sensor

Conductivity: Conductivity estimates the amount of electricity that might potentially pass through the water (in Siemens) in order to determine the concentration of soluble ions in water. The ion concentration in the solution is assessed by this type of analysis. The conductivity rises when there are more ions in the solution. Ions with both positive and negative charges can be created when salts and other substances dissolve in water. The concentration of free ions in the water conducts electricity, hence the electrical conductivity of the water is dependent on the concentration of ions [40]. Here we have used the SEN0244 TDS sensor to determine the conductivity of water.

Turbidity: A turbidity sensor is an optical feature that can continuously monitor and measure the turbidity of water and also describes the level of clarity of a liquid. The turbidity sensor detects the levels of turbidity to assess water quality. It analyzes the light transmittance and reflection rate, which vary with the total suspended solids (TSS) content of the water, in order to detect suspended particles present in the water. The SEN0189 water turbidity sensor monitors the amount of infrared light that enters the phototransistor, which generates the output voltage on the sensor [41].

D1 mini: The D1 Mini is a low-profile, budget-friendly ESP8266-based Wi-Fi board that is just as potent as any NodeMCU on an ESP8266-based microcontroller. It will perform the role of a web server, allowing any Wi-Fi-enabled device to interact with the board and wirelessly control its pins. Due to its low cost, Wi-Fi capability, and complete compatibility with the Arduino platform, the D1 Mini is very versatile [42].

ADS1115: The ADS1115 is an analog-to-digital converter (ADC) with a built-in reference voltage and 16-bit accuracy. It is also extremely small and low-power. It is generally used in situations requiring high-precision collection, such as battery voltage measurement, automotive electronics, and high-precision instrumentation [43].

3.1.1 Semantic Diagram

Firstly, we have designed our IoT devices based on the Arduino IDE platform to get real-time data. As previously discussed, we have DS18B20 sensor for temperature, SEN0161 sensor for pH level, SEN0244 sensor for TDS and conductivity, and SEN0189 sensor for turbidity of water. These sensors are linked with the D1 mini, which is a sort of Arduino board with built-in Wi-Fi.

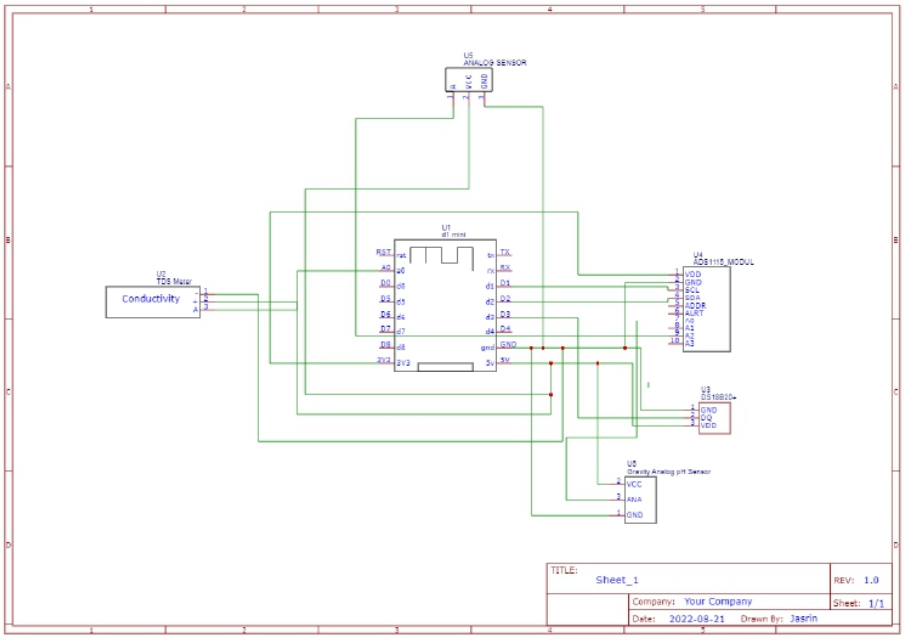


Figure 3.4: Schematic Diagram

Figure 3.4 is a schematic representation of the system we have suggested. As we can see, the esp8266 device, which serves as the core microcontroller of the system we have proposed, has wi-fi which is IEEE 802.11 capabilities. The device has both I2C and SPI interfaces, as we have already mentioned. Our suggested system includes the ADS1115 since it has four sensors, three of which use the analog pin of the microcontroller. Therefore, a converter is required, that's why we have utilized ADS1115.

3.2 Network Layer

After connecting all the devices, we have used Arduino (IDE) platform for programming purposes. Firstly, we have configured the D1 mini which is an ESP8266 Wi-Fi development board by connecting with a laptop using a USB cable and installing packages of ESP8266 Board Manager. We have also installed packages of Wi-Fi manager for Wi-Fi connection, Adafruit_ADS1X15 for ADS1115, OneWire, and DallasTemperature for temperature. In order to upload programs and communicate with central server, it connects to the Arduino hardware by using API key then we verify the program. Gateways can be utilized in the context of water quality to assemble sensor data and transmit it to a central server for processing and analysis.

3.3 Data Processing Layer

Data processing layer is combined with cloud server, data storage and data processing so that it can handle data pre-processing and analysis.

3.3.1 Cloud Server

A cloud server is a dispersed, centralized server source that is hosted and increasing the supply across an Internet-based network and may be used as needed by a variety of users. Cloud servers can offer storage, processing power, and applications in the same ways that a conventional physical server could [44]. In our proposed system, we have used ThingSpeak as our central server. It is an IoT-based cloud platform where we have uploaded our sensor data. With MATLAB or other tools, we have also performed data analysis and data visualization. MathWorks runs the ThingSpeak platform [45]. After accessing the ThingSpeak site we have created a new channel where our channel id is 1823180.

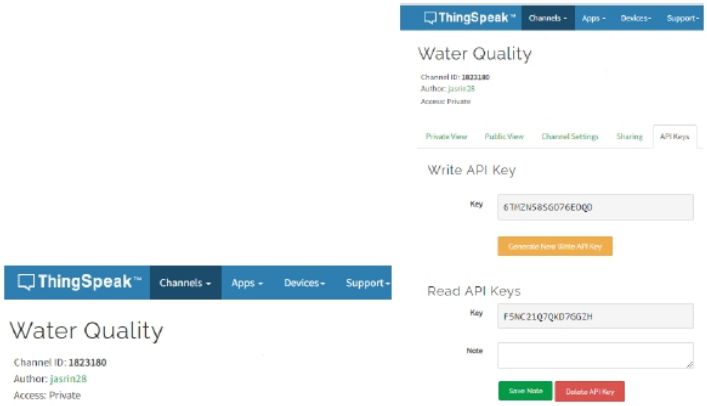


Figure 3.5: ThingSpeak Channel Id & API key

After creating a new channel, we then utilize the API key for communicating with sensor data into both Python and Arduino and viewing the data that has been saved which is shown in figure 3.5.

3.3.2 Data Pre-processing Techniques

Past Data Collection: Firstly, we have collected 3180 data from WHO ideal water data with five water parameters such as temperature, TDS, conductivity, pH, and turbidity which are the independent variables, and portability of water is the dependent variable. Figure 3.6 shows the input data of water.

inputData

	Temparature	ph	Solids	Conductivity	Turbidity	Potability
0	20.0	9.092223	17978.98634	398.410813	4.075075	0.0
1	18.0	8.099124	19909.54173	418.606213	3.055934	0.0
2	20.0	8.316766	22018.41744	363.266516	4.628771	0.0
3	20.0	9.092223	17978.98634	398.410813	4.075075	0.0
4	48.0	5.584087	28748.68774	280.467916	2.559708	1.0
...	...	...	...	...	...	...
3175	26.0	7.808856	17329.80216	392.449580	2.798243	1.0
3176	12.0	9.419510	33155.57822	432.044783	3.298875	0.0
3177	12.0	5.126763	11983.86938	402.883113	4.708658	0.0
3178	34.0	7.874671	17404.17706	327.459761	2.309149	1.0
3179	24.0	6.800119	23403.63730	475.573562	2.747313	1.0

3180 rows × 6 columns

Figure 3.6: Input Data of Water

After collecting the past data, we have pre-processed our data sets by removing noise and outliers. Then we split our data sets and build the model on training data.

Machine Learning Model: We have applied different types of machine learning algorithms such as Decision Tree, Random Forest, Extra Tree, and K-NN (K-Nearest Neighbour) Classification on the dataset to build a prediction model. Some descriptions of those algorithms are given below-

Decision Tree Classifier: A decision tree is a supervised learning technique that may be used to address both classification and regression models, while it is most frequently used to handle classification models. It is a tree-structured classification algorithm, with internal nodes denoting data set features, branches identifying decision rules, and each leaf node representing the classification result. A decision tree has a

decision node and a leaf node as its two nodes. Whereas Leaf nodes are the results of decisions and have no more branches, Decision nodes are used to make decisions and have numerous branches [46].

Random Forest Classifier: The Random Forest classifier employs a variety of decision trees to different subsets of the input data in order to increase the projected accuracy of the data set. It employs predictions rather than just one decision tree from all of the trees to make educated judgments about the outcome based on which forecasts have gained the greatest support. Because there are more trees in the forest, accuracy is increased and overfitting is prevented. It takes less time to train than other algorithms, and accuracy can still be kept even when a sizeable chunk of the data is missing [47].

Extra Tree Classifier: Extremely Randomized Trees Classifier, another name for the Extra Trees Classifier, is a kind of ensemble learning method that combines the outcomes of various de-correlated decision trees assembled in a "forest" to produce its classification output. It differs from a Random Forest Classifier only in terms of how the decision trees in the forest are constructed theoretically. In Extra Tree Classifier, the root node or any other node is constructed using a subset of randomly selected attributes of size sqrt, and the best split is discovered by searching through this subset. Each feature's split is determined at random and the decision stump can go only as deep as one [48].

KNN Classifier: Problems involving classification and regression prediction can be resolved using KNN (K-nearest Neighbors), a type of supervised machine learning technique. KNN has two characteristics, including being non-parametric and a lazy learner. Due to the lack of a distinct training phase, KNN employs all of the data for training during classification, making it a lazy learning method. KNN does not generate any presumptions on the underlying data [49].

Satellite Data Collection: As we have mentioned earlier, Chittagong is our study area. For that reason, we have collected 200 satellite images in the different parts of Chittagong such as the Karnafuli river, the Halda river, the lakes of Rangamati, and the nearest area of the Bay of Bengal. We have taken satellite map data of Landsat-8 and Sentinel-2 satellites jointly as our image data sets from EarthExplorer (https://earthexplorer.usgs.gov/). Figure 3.7 shows the location of our study area.

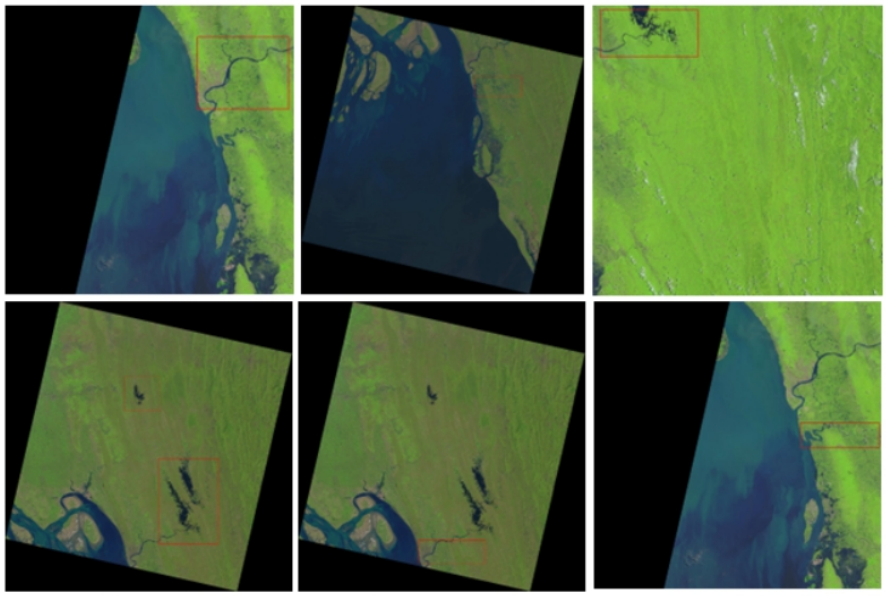


Figure 3.7: Location of the study area

After selecting our study area, we have cropped those images by using a window and also removed unnecessary information and take our sample area as an input image to assess the water quality which is shown in figure 3.8.

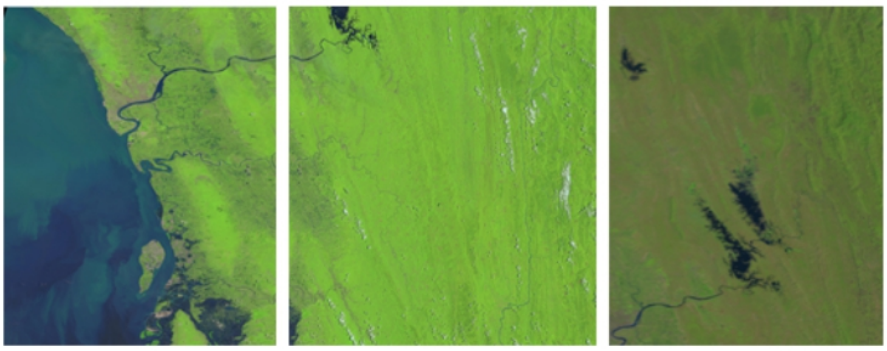


Figure 3.8: Input image

After collecting image data, we have preprocessed the data and build a model on datasets to classify the images and extract knowledge from those images.
DL Model: For image classification, we have applied the Deep Learning technique such as the CNN model on our image data. We have chosen CNN to classify the image data because it has a deep network architecture with weighted layers, which makes it

suitable for accurately categorizing images and makes it possible to capture robust and distinctive properties. We have also modified our architecture by altering the number of activation functions, hidden layers, and other hyperparameters. Here we have used the sigmoid and ReLU function as the activation function based on our system. The description of the activation function is given below-

Sigmoid Function: Sigmoid function is a mathematical function that is used to add nonlinearity to a classification model. In simple words, it chooses which values are appropriate for output and which are not. It converts any real value to a number between 0 and 1 [50]. The equation of the sigmoid function is given below-

$$Y = 1 \ / \ 1+e\hat{}z$$

Where Y = Sigmoid function.

ReLU Function: The rectified linear activation function, also referred to as ReLU, is a piecewise linear function that outputs directly if the input is positive or provides outputs zero when the input is not positive. In neural networks, especially CNN, it is the most often employed activation function [51]. The equation of the ReLU activation function is given below-

$$f(x) = \max(0, x)$$

Where f(x) = ReLU function.

WHO Approved Water Limit: Some statistical summary has been recommended by WHO [52] for pure drinking water parameters which can help to examine water quality. They also provide a guideline for drinking water limits which is shown in Table 3.1

Table 3.1: Recommended water parameters by WHO

Water Parameters	WHO approved the limit for drinking water
Temperature	30°C
pH	6.5-8.5
TDS	500-1000 mg/l
Conductivity	400 S/cm
Turbidity	1-5 NTU

All those values are provided by WHO so that we can examine the water quality by considering all the ideal values of water parameter. If our sample water is near to ideal values then we consider the water as "freshwater" otherwise the water is "seawater". After configuring our target features, we have represented freshwater as "1" and seawater as "0".

3.4 Application Layer

We have considered all the image input data and WHO water parameter input data, to preprocess the data and build the model to get the output result. Then we have collected sample water and examine the water quality by considering 5 input water variables like temperature, TDS, pH, turbidity, and conductivity for monitoring water quality in real-time. After getting all the results, we crossmatch the predicted result of both IoT and image data to compare the actual result and identify the current situation of water quality. Figure 3.9 shows the flow chart of our proposed system.

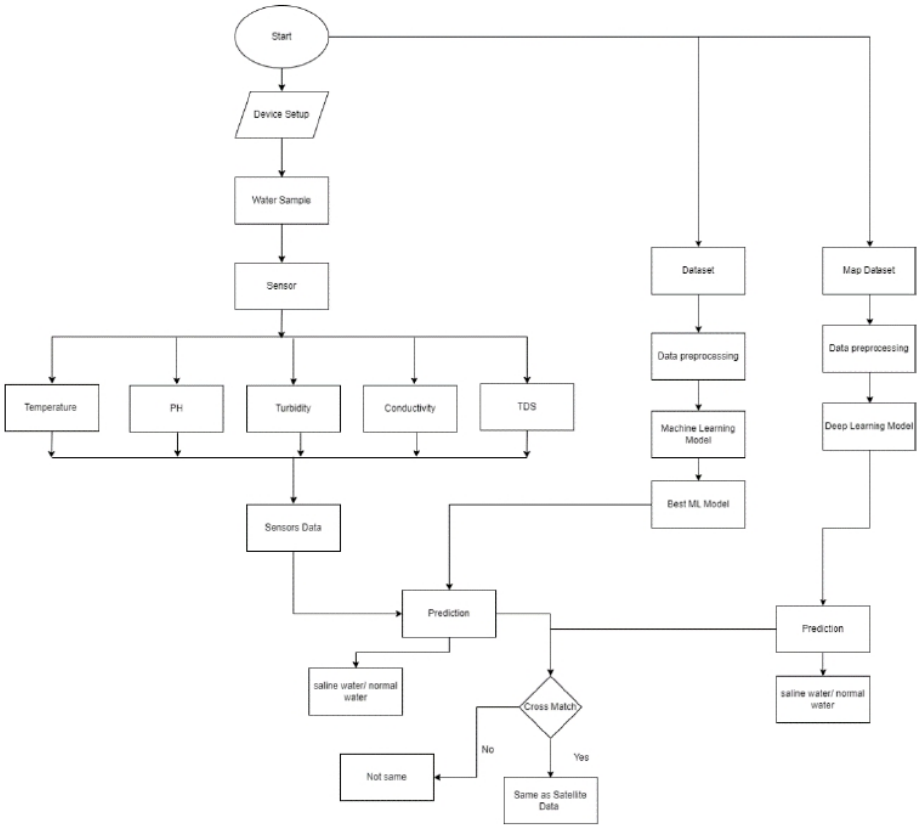


Figure 3.9: Flowchart of Proposed System

In short, we have taken water parameters from IoT-based sensors in real-time to monitor water quality remotely. We have also collected satellite imageries from Landsat-8 and Sentinel-2 satellites to extract the knowledge of images and detect the region of water quality based on their location. Finally, we have collected past water data from WHO ideal water data to analyze the water parameter and examine the water quality of our study area. Then we crossmatch both the result of IoT data and image data to compare the result which is shown in web interface.

Chapter 4

Implementation & Results

Our Project discover facts and factors which we have made plan of action to framework of implementation in the system with advance analysis of probability of all aspect and this research has well findings with result which may accept as final output of this research

4.1 IoT Result

In the IoT part, we have discussed the device connection, uploaded sample water data in the cloud server, and shown the last IoT data.

4.1.1 Hardware Setup

The conductivity sensor, sometimes known as the TDS sensor, has three pinouts. Analog, VCC, and GND. The connection is shown in below table 4.1.

Table 4.1: Connection of D1 mini and TDS sensor

D1 mini	TDS sensor
5 volts	VCC
GND	GND
A0	A

And the other sensor which is DS18B20 also has three pinouts and the connection process of this sensor to D1 mini is given in the following table 4.2.

Table 4.2: Connection of D1 mini and Temperature sensor

D1 mini	Temperature sensor(DS18B20)
5 volts	VCC
GND	GND
D3	Digital Pin

And the other sensors are the turbidity sensor and PH sensor. Both sensors are analog sensors and the D1 mini does not have another analog pin as well. That's why we need ADS1115 as early said. So first we see the turbidity connection with D1 mini and ADS1115 in table 4.3.

Table 4.3: Connection of D1 mini and Turbidity and ADS1115 sensor

D1 mini	Turbidity sensor	ADS1115
5 volts	VCC	
3.3 volts		VCC
GND	GND	GND
	A	A1
D1		SCL
D1		SDA

The connection of the pH sensor with D1 mini and ADS1115 is shown in table 4.4.

Table 4.4: Connection of D1 mini and pH and ADS1115 sensor

D1 mini	pH sensor	ADS1115
5 volts	VCC	
3.3 volts		VCC
GND	GND	GND
	A	A0
D1		SCL
D1		SDA

In the above table, we can see the connection of each sensor and converter with D1 mini and with one another as well. In the last two tables, we can see three device connections for one feature which is necessary as we have used ads1115 as analog to digital converter. In the first two tables, there is no need for these converters, and the sensor interface connection is one-to-one with the D1 mini and the other sensor. The overall hardware setup of our system is shown in Figure 4.1.

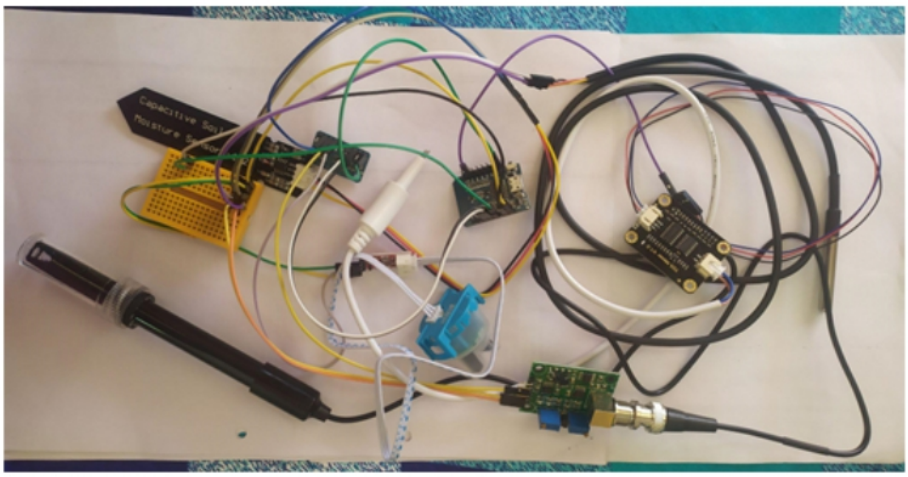


Figure 4.1: Initial Hardware Setup

Firstly, we have implemented the system with a sample water in small plastic mug to test the performance of the sensors and other devices which is shown in Figure 4.2. After establishing the connection, we communicate with the cloud server through the internet. Hardware and sensors are in good working order. Here, sensors can detect values from given water samples in real-time, and these results have sent to an IoT-based cloud platform called ThingSpeak.



Figure 4.2: Final setup for system testing

4.1.2 ThingSpeak Server

An open-source IoT application called ThingSpeak can save and retrieve data from sensors across a LAN or the internet using an HTTP server. For that purpose, we have used the ThingSpeak server as a central server to store our sensor data, analyze the real-time streaming data, communicate with data, and access the data by using the API key. In order to supply sensor data, ThingSpeak can access MATLAB. In order to use the data in a software environment for past data analysis, the data collection method also allows data manipulation. Here, we can view both recent data and all previous data. Figure 4.3 shows the sensors data of IoT which we have sent to the cloud server through the IoT gateway by using the API key of our ThingSpeak channel.



Figure 4.3: Sensor data in ThingSpeak server

We have also created five field in our ThingSpeak channel where filed1 represents the "Temperature", field2 represents the "pH", field3 represents the "Solids", field4 represents the "Conductivity" and field5 represents the "Turbidity". We can also view the last data of IoT from ThingSpeak using API key where temperature is 26, pH is 6.10, TDS is 11321.00, conductivity is 123, and turbidity is 2.23 which is shown in figure 4.4.

```
{
    "created_at": "2022-10-01T08:32:20Z",
    "entry_id": 198,
    "field1": "26",
    "field2": "6.10",
    "field3": "11321.00",
    "field4": "123.00",
    "field5": "2.23"
}
```


Figure 4.4: Last IoT data

4.2 Analysis of Past Data

We have also gathered 3180 past data from the WHO ideal water data, where the portability of the water is the dependent variable and the five water parameters such as temperature, TDS, pH, turbidity, and conductivity are the independent variables. After taking input data we have described our data sets to identify the outliers and important features. For that reason, we have determined the mean, standard deviation, min, Q1, Q2, Q3, and max values for our independent and dependent variables which are shown in figure 4.5.

	Temparature	ph	Solids	Conductivity	Turbidity	Potability
count	2846.000000	2843.000000	3180.000000	3180.000000	3180.000000	2964.000000
mean	26.187280	7.193335	21729.252252	426.719282	3.952928	0.612348
std	11.186063	1.496249	8583.192591	79.486804	0.778158	0.487297
min	1.000000	0.000000	320.942611	181.483754	1.450000	0.000000
25%	20.000000	6.261955	15297.682445	368.971151	3.441806	0.000000
50%	24.000000	7.139671	20562.104060	420.541695	3.928896	1.000000
75%	32.000000	8.076694	26869.085315	479.510333	4.490177	1.000000
max	57.000000	14.000000	61227.196010	753.342620	6.494749	1.000000

Figure 4.5: Mean, Standard Deviation, Q1, and Q3

Here we represent the descriptive statistics for our input data which is shown in figure 4.6.

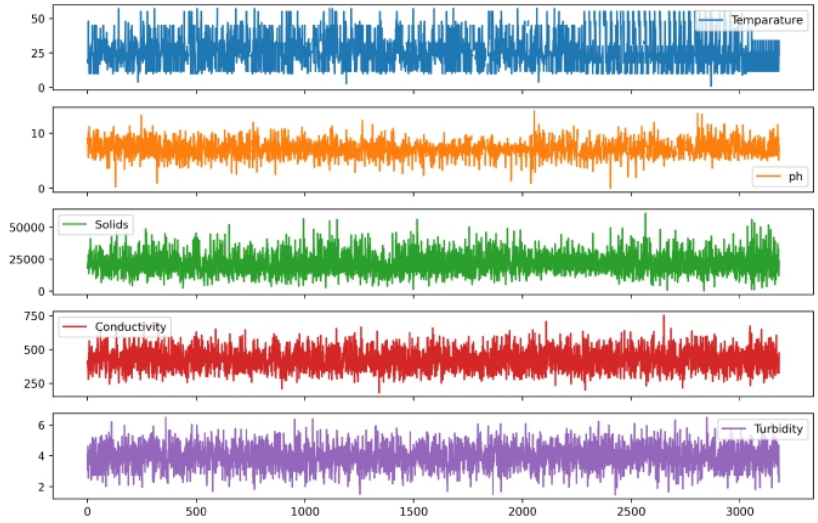


Figure 4.6: Data Description

For the preprocessing purpose, we find out how many missing data and outliers are present in our dataset. Then we have removed all the missing data and outliers from our datasets so that model cannot be troubled by some unwanted noise. After doing the filtration process, we have 2751 data with five independent variables and one dependent variable where our actual data was 3180 which is shown in figure 4.7.

```
filter_data
```

	Temparature	ph	Solids	Conductivity	Turbidity	Potability
0	20.0	9.092223	17978.98634	398.410813	4.075075	0.0
1	18.0	8.099124	19909.54173	418.606213	3.055934	0.0
2	20.0	8.316766	22018.41744	363.266516	4.628771	0.0
3	20.0	9.092223	17978.98634	398.410813	4.075075	0.0
4	48.0	5.584087	28748.68774	280.467916	2.559708	1.0
...	...	...	...	...	...	...
3175	26.0	7.808856	17329.80216	392.449580	2.798243	1.0
3176	12.0	9.419510	33155.57822	432.044783	3.298875	0.0
3177	12.0	5.126763	11983.86938	402.883113	4.708658	0.0
3178	34.0	7.874671	17404.17706	327.459761	2.309149	1.0
3179	24.0	6.800119	23403.63730	475.573562	2.747313	1.0

2751 rows × 6 columns

Figure 4.7: Processed input data

We have developed our classification model on WHO ideal water data to show the statistical view of our dataset after preprocessing where figure 4.8 represents the bar graph of the target values. Here, the value "1" represents the "freshwater" and the value "0" represents the "seawater" against the input features of our datasets.

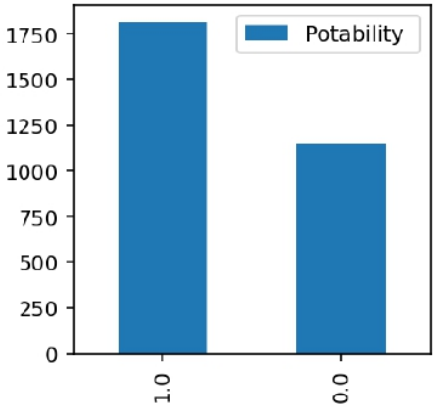


Figure 4.8: Target Value

We have also found out the outliers for all individual independent variables such as temperature, pH, TDS, conductivity, and turbidity by using a boxplot which is shown in figure 4.9.

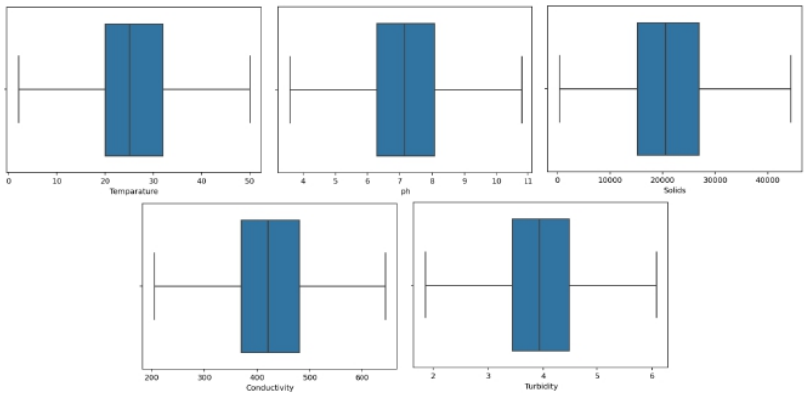


Figure 4.9: Boxplot

After preprocessing we split our data sets into 80% as training data and 20% as testing data. Then we have applied several machine learning algorithms such as Decision Tree Classification, Random Forest Classification, Extra Tree Classification, and KNN classification for the model building where we have achieved good accuracy which is almost

86%. We have got accuracy from Decision Tree Classification is 85.29%, Random Forest Classification is 85.66%, Extra Tree Classification is 84.75%, and KNN Classification is 79.85%. The comparison accuracy of our machine learning model is shown in figure 4.10.

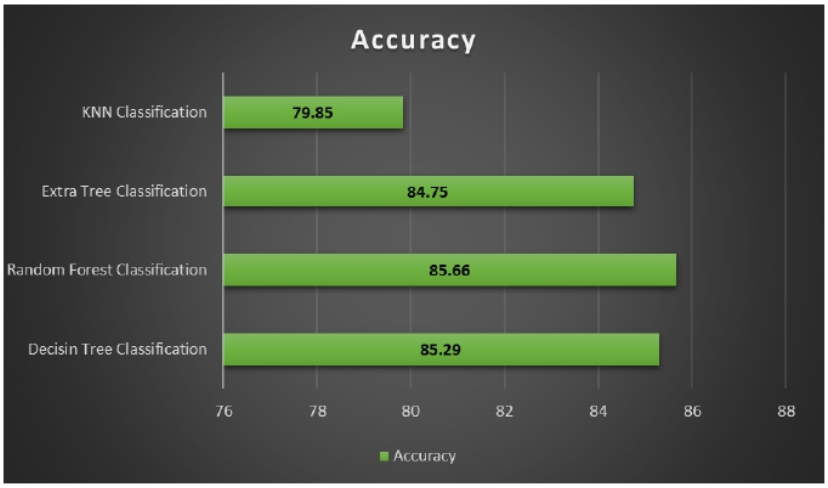


Figure 4.10: Comparison accuracy of the model

From accuracy result of above model, we can see that Random Forest Classification gives good accuracy than other classification models which is shown in figure 4.11.

```
[52] from sklearn.metrics import accuracy_score
     a=accuracy_score(y_test,pred1)
     print("The accuracy of Random Forest Classification is: ", a*100)

     The accuracy of Random Forest Classification is:  85.66243194192378
```

Figure 4.11: Accuracy result of Random Forest Classification

From past data analysis we find out the best machine learning model for our datasets which is Random Forest Classification. Then we fetch the last IoT data from ThingSpeak to Google Colab by using the read API key and to test the data into the best model which is Random Forest classifier to show the water quality category of last IoT data. Here we have got "1" which means "freshwater" based on the last data of IoT. The overall result is shown in figure 4.12.

```
[1.]
[1.]
```

Figure 4.12: Final output of IoT

4.3 Satellite Image Processing

We have also collected satellite image data from EarthExplorer. After collecting image data, we have modified the dataset and transformed it into a structured format, which now includes a folder with the images and a CSV file with binary labels. We have categorized the binary labels into water and sea whose target values are 1 and 0 based on their location. Figure 4.13 shows the categorized CSV data.

```
train.head(100)
```

	image	Water	Sea
0	1.png	1	0
1	2.png	1	0
2	3.png	1	0
3	4.png	1	0
4	6.png	1	0
...	...	...	...
95	98.png	1	0
96	99.png	0	1
97	100.png	0	1
98	101.png	0	1
99	102.png	0	1

100 rows × 3 columns

Figure 4.13: Categorized Data

Then we load and preprocess the image data and reshape those images into 400x400 pixels and 3 dimensions for the purpose of model building which is shown in figure 4.14.

```
X.shape
```

```
(198, 400, 400, 3)
```

Figure 4.14: Reshape Data

Then we plot and visualized the image data which is shown in figure 4.15.

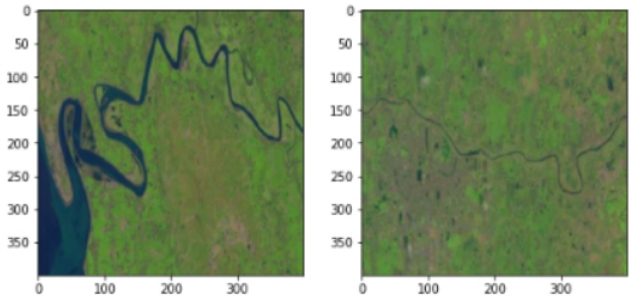


Figure 4.15: Visualization of image data

Then we split our data sets into 80% as training data for model building and 20% as testing data for validation purposes. After splitting data, we have applied deep learning techniques such as CNN (Convolutional Neural Network) model to classify the images and extract knowledge from those images. To evaluate the model's effectiveness, we have trained our model for 10 epochs and also pass the validation data which we have generated earlier. In figure 4.16 we can observe that the training loss has been decreased to 0.1 and the validation loss has been decreased to 0.2.

```
model.fit(x_train, y_train, epochs=10, validation_data=(x_test, y_test))

Epoch 1/10
5/5 [==============================] - 57s 11s/step - loss: 0.8061 - accuracy: 0.5253 - val_loss: 0.6841 - val_accuracy: 0.9500
Epoch 2/10
5/5 [==============================] - 51s 10s/step - loss: 0.6213 - accuracy: 0.6582 - val_loss: 0.6056 - val_accuracy: 0.8000
Epoch 3/10
5/5 [==============================] - 51s 10s/step - loss: 0.4196 - accuracy: 0.8797 - val_loss: 0.4357 - val_accuracy: 0.9250
Epoch 4/10
5/5 [==============================] - 51s 10s/step - loss: 0.2098 - accuracy: 0.9304 - val_loss: 0.1958 - val_accuracy: 0.9500
Epoch 5/10
5/5 [==============================] - 52s 10s/step - loss: 0.1501 - accuracy: 0.9684 - val_loss: 0.3307 - val_accuracy: 0.8750
Epoch 6/10
5/5 [==============================] - 52s 10s/step - loss: 0.1057 - accuracy: 0.9747 - val_loss: 0.1300 - val_accuracy: 0.9750
Epoch 7/10
5/5 [==============================] - 52s 10s/step - loss: 0.1475 - accuracy: 0.9620 - val_loss: 0.2574 - val_accuracy: 0.9250
Epoch 8/10
5/5 [==============================] - 52s 11s/step - loss: 0.1390 - accuracy: 0.9620 - val_loss: 0.2324 - val_accuracy: 0.9500
Epoch 9/10
5/5 [==============================] - 52s 10s/step - loss: 0.1194 - accuracy: 0.9810 - val_loss: 0.2109 - val_accuracy: 0.9250
Epoch 10/10
5/5 [==============================] - 51s 10s/step - loss: 0.0988 - accuracy: 0.9620 - val_loss: 0.2012 - val_accuracy: 0.9500
<keras.callbacks.History at 0x7f9baf483e90>
```

Figure 4.16: Loss of validation data

Again, we load and preprocess the image data individually and predict these images by using our training model. Then we have determined the probability of target images which shows the accuracy of being water or sea. Figure 4.17 shows the accuracy of image data.

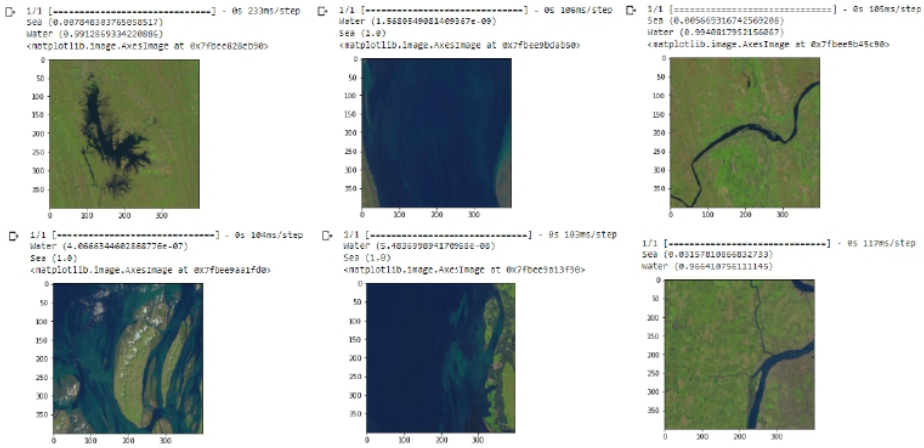


Figure 4.17: Accuracy of the output image

After getting the accuracy of image data, we crossmatch the results of image and sample data of IoT to compare both the results. The overall output is shown in our web-interface.

4.4 Web Interface

For our web interface, we have installed "Sublime text" software as a backend platform, and the language we have used for the backend is "PHP" for the development part. Firstly, we have designed a dashboard where we have shown the graphical view of our last IoT data. Then we have shown the IoT last data which we have sent to the cloud server and the IoT result with the single output image result in the tabular format. Then we crossmatch both the IoT and image result where the actual result will be "TRUE" if both data is matched otherwise "FALSE" which is shown in our web interface. We have also represented the graphical view of all IoT past data in our interface which we have sent to ThingSpeak previously. Figure 4.18 shows the web interface of our proposed system.



Figure 4.18: Web Interface

4.5 Discussion

Our main effort is to design an IoT-based system to measure water quality remotely in real-time. The challenging part of this work is to collect industrial sensors. Following that, we have used sensors for measuring water parameters such as temperature, pH, TDS, conductivity, and turbidity sensors. Then we have collected satellite data jointly from Landsat-8 and Sentinel-2 satellites and pre-processed the data and applied deep learning method such as CNN on image data to characterize the water quality based on their location. In this part, we have faced an issue which is cloud cover. We have to take images by selecting additional features of cloud cover below five. Otherwise, the study area was not identified. Then we have collected WHO ideal water data for analyzing water quality by building model and classifying the categories of freshwater and seawater. For that purpose, we have applied several machine learning algorithms such as DT Classification, RF Classification, Extra Tree Classification, and KNN Classification. All these model gives good accuracy where RF Classification performs well than other models in our data sets. Finally, we have crossmatched both the IoT and image results to find the actual water quality result.

Chapter 5

Conclusion and Future Work

Water is not only necessary for humanity's survival but also important for human health, economic growth, food security, the reduction of poverty, and the sustainability of ecological processes. Because of the climate crisis water level is rising and the availability of freshwater is reduced and saline water is increasing day by day worldwide. Saline water is unsuited for human food and household usage due to its salinity and high salt content. For that reason, desalination plants are encouraged to be built as a means of preserving the region's freshwater supply and, consequently, the viability of the agricultural system. It is well acknowledged that there is a need to provide innovative platforms such as remote sensing technologies that can be useful for the monitoring of various water contaminants. Our main goal is to monitor water quality remotely, to determine whether the water is safe for human consumption and domestic use. In this work, we have collected satellite images where Landsat-8 and Sentinel-2 satellites are jointly used for extracting the water land area of the sea and river. The map data is taken from the Earth Explorer by selecting our study area. In this research, our study area is the southeast part of Bangladesh which is the Chittagong region. To process map data and extract knowledge from those data we have applied Deep Learning based technique which is CNN on satellite images to identify whether the water area is freshwater or sea water based on their location. We have also applied the data analytic techniques on the past ideal water data which is formulated by WHO to examine the water quality by considering some water parameters. We have removed unwanted noise, outliers, and missing values to process data and build the model by applying some Machine Learning techniques such as Decision Tree Classification, Random Forest Classification, KNN Classification, and Extra Tree Classification, where all the algorithm works well but Random Forest Classification gives better accuracy than other models which is 85.66%. Then we have designed an IoT-based architecture to examine the water quality by considering some water parameters such as temperature, electrical conductivity, pH, TDS, and turbidity to get real-time data and also monitor water quality without human interaction. We have also sent the data to the ThingSpeak server, which is a centralized cloud-based system with a free database platform. The outcome of water quality is cross-matched using comparisons from both satellite

images and IoT data. The model is also be tested using actual data, which is used to categorize the category of water quality as freshwater or sea water.

Future Work: In the future, we will consider both physical and chemical water parameters and categorize our dependent variables into multiple categories to examine water quality and try to improve our accuracy. We will also develop an alarming system for unwanted occurrences and people's awareness when the quality of water is not suitable for human consumption and we will use this system for residential and industrial use as a product to serve the people of Bangladesh. We will also convert our centralized system into a decentralized one for security purposes and to reduce the burden on the centralized cloud system.

Bibliography

[1] K. Charles, L. Ong, N. Achi, K. Ahmed, M. Khan, S. Hoque, and S. Nowicki, "Bangladesh mics 2019: Water quality thematic report," *Bangladesh Bureau of Statistics, Ministry of Planning, Government of the People's Republic of Bangladesh and UNICEF Bangladesh: Dhaka, Bangladesh*, 2021. 1

[2] A. Motevalli, H. R. Pourghasemi, H. Hashemi, and V. Gholami, "Assessing the vulnerability of groundwater to salinization using gis-based data-mining techniques in a coastal aquifer," in *Spatial modeling in GIS and R for earth and environmental sciences*. Elsevier, 2019, pp. 547–571. 1

[3] L. Benneyworth, J. Gilligan, J. C. Ayers, S. Goodbred, G. George, A. Carrico, M. R. Karim, F. Akter, D. Fry, K. Donato *et al.*, "Drinking water insecurity: water quality and access in coastal south-western bangladesh," *International journal of environmental health research*, vol. 26, no. 5-6, pp. 508–524, 2016. 1

[4] A. Kumawat, "Satellite image processing." 2022. [Online]. Available: https://www.geeksforgeeks.org/satellite-image-processing/ 1

[5] J. Clark, "What is the internet of things (iot)?" 2016. [Online]. Available: https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/ 1

[6] "What is data analytics?" [Online]. Available: https://aws.amazon.com/what-is/data-analytics/ 2

[7] A. Salam and A. Salam, "Internet of things for water sustainability," *Internet of Things for Sustainable Community Development: Wireless Communications, Sensing, and Systems*, pp. 113–145, 2020. 2

[8] T. Akter, F. T. Jhohura, F. Akter, T. R. Chowdhury, S. K. Mistry, D. Dey, M. K. Barua, M. A. Islam, and M. Rahman, "Water quality index for measuring drinking water quality in rural bangladesh: a cross-sectional study," *Journal of Health, Population and Nutrition*, vol. 35, pp. 1–12, 2016. 2

[9] N. R. Moparthi, C. Mukesh, and P. V. Sagar, "Water quality monitoring system using iot," in *2018 Fourth international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB)*. IEEE, 2018, pp. 1–5. 2

[10] M. R. Karim, H. R. B. Khan, A. Ahsan, M. Rahadujjaman, S. Alam, R. Hasan, T. T. Sharif, and S. Farooq, "Potential health hazard of drinking water in restaurants and tea stalls," *Environmental Quality Management*, 2022. 3

[11] G. M. Munna, S. Islam, N. Hoque, K. Bhattacharya, and S. Nath, "A study on water quality parameters of water supply in sylhet city corporation area," *Hydrology*, vol. 3, no. 6, pp. 66–71, 2015. 3

[12] S. Alam and M. A. Matin, "Application of 2d morphological model to assess the response of karnafuli river due to capital dredging," *Journal of Water Resources and Ocean Science*, vol. 2, no. 3, pp. 40–48, 2013. 3

[13] S. Behmel, M. Damour, R. Ludwig, and M. Rodriguez, "Water quality monitoring strategies—a review and future perspectives," *Science of the Total Environment*, vol. 571, pp. 1312–1329, 2016. 3

[14] D. C. Barrett and A. E. Frazier, "Automated method for monitoring water quality using landsat imagery," *Water*, vol. 8, no. 6, p. 257, 2016. 3

[15] M. A. Kadhim and M. H. Abed, "Convolutional neural network for satellite image classification," *Intelligent Information and Database Systems: Recent Developments 11*, pp. 165–178, 2020. 4

[16] M. Alkhelaiwi, W. Boulila, J. Ahmad, A. Koubaa, and M. Driss, "An efficient approach based on privacy-preserving deep learning for satellite image classification," *Remote Sensing*, vol. 13, no. 11, p. 2221, 2021. 4

[17] A. Otte, D. Coates, R. Connor, G. Roder, D. Hebart-Coleman, M. Klimes, E. Yaari, M. Gutierrez, N. Crawhall, R. Kinna *et al.*, "Culture and the values of water," *cultural values*, vol. 46, p. 41Z, 2021. 4

[18] H. Tazoe, "Water quality monitoring," *Analytical Sciences*, vol. 39, no. 1, pp. 1–3, 2023. 7

[19] S. A. Jaywant and K. M. Arif, "A comprehensive review of microfluidic water quality monitoring sensors," *Sensors*, vol. 19, no. 21, p. 4781, 2019. 7

[20] B. Das and P. Jain, "Real-time water quality monitoring system using internet of things," in *2017 International conference on computer, communications and electronics (Comptelix)*. IEEE, 2017, pp. 78–82. 7

[21] M. Alshehri, M. Kumar, A. Bhardwaj, S. Mishra, and J. Gyani, "Deep learning based approach to classify saline particles in sea water," *Water*, vol. 13, no. 9, p. 1251, 2021. 8

[22] A. Czajkowski, L. Remiorz, S. Pawlak, E. Remiorz, J. Szyguła, D. Marek, M. Paszkuta, G. Drabik, G. Baron, J. Paduch *et al.*, "Global water crisis: Concept of a new interactive shower panel based on iot and cloud computing for rational water consumption," *Applied Sciences*, vol. 11, no. 9, p. 4081, 2021. 8

[23] P. Prusty, S. Farooq, D. Swain, and D. Chandrasekharam, "Association of geomorphic features with groundwater quality and freshwater availability in coastal regions," *International journal of environmental science and technology*, vol. 17, pp. 3313–3328, 2020. 8

[24] I. Caballero, A. Román, A. Tovar-Sánchez, and G. Navarro, "Water quality monitoring with sentinel-2 and landsat-8 satellites during the 2021 volcanic eruption in la palma (canary islands)," *Science of the Total Environment*, vol. 822, p. 153433, 2022. 8

[25] M. S. U. Chowdury, T. B. Emran, S. Ghosh, A. Pathak, M. M. Alam, N. Absar, K. Andersson, and M. S. Hossain, "Iot based real-time river water quality monitoring system," *Procedia Computer Science*, vol. 155, pp. 161–168, 2019. 9

[26] S. Pasika and S. T. Gandla, "Smart water quality monitoring system with cost-effective using iot," *Heliyon*, vol. 6, no. 7, p. e04096, 2020. 9

[27] J. O. Ighalo, A. G. Adeniyi, and G. Marques, "Internet of things for water quality monitoring and assessment: a comprehensive review," *Artificial intelligence for sustainable development: theory, practice and future applications*, pp. 245–259, 2021. 9

[28] P. Liu, J. Wang, A. K. Sangaiah, Y. Xie, and X. Yin, "Analysis and prediction of water quality using lstm deep neural networks in iot environment," *Sustainability*, vol. 11, no. 7, p. 2058, 2019. 9

[29] K. Saravanan, E. Anusuya, R. Kumar, and L. H. Son, "Real-time water quality monitoring using internet of things in scada," *Environmental monitoring and assessment*, vol. 190, pp. 1–16, 2018. 10

[30] R. Dhakate, S. Sankaran, V. S. Kumar, B. Amarender, P. Harikumar, and S. Subramanian, "Demarcating saline water intrusion pathways using remote sensing, gis and geophysical techniques in structurally controlled coastal aquifers in southern india," *Environmental Earth Sciences*, vol. 75, pp. 1–19, 2016. 10

[31] F. Jan, N. Min-Allah, and D. Düştegör, "Iot based smart water quality monitoring: Recent techniques, trends and challenges for domestic applications," *Water*, vol. 13, no. 13, p. 1729, 2021. 10

[32] I. Yaroshenko, D. Kirsanov, M. Marjanovic, P. A. Lieberzeit, O. Korostynska, A. Mason, I. Frau, and A. Legin, "Real-time water quality monitoring with chemical sensors," *Sensors*, vol. 20, no. 12, p. 3432, 2020. 11

[33] A. Bhardwaj, V. Dagar, M. O. Khan, A. Aggarwal, R. Alvarado, M. Kumar, M. Irfan, and R. Proshad, "Smart iot and machine learning-based framework for water quality assessment and device component monitoring," *Environmental Science and Pollution Research*, vol. 29, no. 30, pp. 46 018–46 036, 2022. 11

[34] M. Miller, A. Kisiel, D. Cembrowska-Lech, I. Durlik, and T. Miller, "Iot in water quality monitoring—are we really here?" *Sensors*, vol. 23, no. 2, p. 960, 2023. 11

[35] P. Chen, B. Wang, Y. Wu, Q. Wang, Z. Huang, and C. Wang, "Urban river water quality monitoring based on self-optimizing machine learning method using multi-source remote sensing data," *Ecological Indicators*, vol. 146, p. 109750, 2023. 11

[36] G. Jiao, S. Chen, F. Wang, Z. Wang, F. Wang, H. Li, F. Zhang, J. Cai, and J. Jin, "Water quality evaluation and prediction based on a combined model," *Applied Sciences*, vol. 13, no. 3, p. 1286, 2023. 12

[37] M. INTEGRATED, "Ds18b20 programmable resolution 1-wire digital thermometer," *Rev*, vol. 4, p. 22, 2008. 16

[38] A. Oberoi, S. Basavaraju, and S. Lekshmi, "Effective implementation of automated fertilization unit using analog ph sensor and arduino," in *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. IEEE, 2017, pp. 1–5. 16

[39] Z. Abidin, E. Maulana, M. Y. Nurrohman, F. C. Wardana *et al.*, "Real time monitoring system of drinking water quality using internet of things," in *2022 International Electronics Symposium (IES)*. IEEE, 2022, pp. 131–135. 17

[40] "Electrical conductivity (ec) sensor." [Online]. Available: https://www.aquaread.com/sensors/ec 17

[41] W. Hakim, L. Hasanah, B. Mulyanti, and A. Aminudin, "Characterization of turbidity water sensor sen0189 on the changes of total suspended solids in the water," in *Journal of Physics: Conference Series*, vol. 1280, no. 2. IOP Publishing, 2019, p. 022064. 17

[42] J. Hrisko, "Wemos d1 mini esp8266 arduino wifi board." 2019. [Online]. Available: https://makersportal.com/blog/2019/6/12/wemos-d1-mini-esp8266-arduino-wifi-board 17

[43] L. Gong, "Analog-to-digital converter——ads1115," 2021. [Online]. Available: https://www.seeedstudio.com/blog/2020/09/30/analog-to-digital-converter-ads1115/ 18

[44] "What is a cloud server?" [Online]. Available: https://www.vmware.com/topics/glossary/content/cloud-server.html 19

[45] H.-P. Halvorsen, "What is thingspeak?" [Online]. Available: https://www.halvorsen.blog/documents/technology/iot/thingspeak/thingspeak.php 19

[46] "Decision tree classification algorithm." [Online]. Available: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm 21

[47] "Random forest algorithm." [Online]. Available: https://www.javatpoint.com/machine-learning-random-forest-algorithm 21

[48] F. Ceballos, "An intuitive explanation of random forest and extra trees classifiers," 2019. [Online]. Available: https://towardsdatascience.com/an-intuitive-explanation-of-random-forest-and-extra-trees-classifiers-8507ac21d54b 21

[49] "K-nearest neighbor(knn) algorithm for machine learning." [Online]. Available: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning 21

[50] D. Kumawat, "Introduction to logistic regression - sigmoid function, code explanation," 2021. [Online]. Available: https://www.analyticssteps.com/blogs/introduction-logistic-regression-sigmoid-function-code-explanation 23

[51] V. Praharsha, "Relu (rectified linear unit) activation function." [Online]. Available: https://iq.opengenus.org/relu-activation/ 23

[52] W. H. Organization *et al.*, "A global overview of national regulations and standards for drinking-water quality," 2021. 23