A Low Cost Innovative Solution for Automatic Water Quality Monitoring using Machine Learning

A

Thesis

Submitted to

For the award of

DOCTOR OF PHILOSOPHY (Ph.D)

in

Electronics & Communication Engineering

By

Sandeep Bansal

41500062

Supervised By

Dr. G. Geetha

Professor & Dean

Division of Research & Development

LOVELY FACULTY OF TECHNOLOGY AND SCIENCES
LOVELY PROFESSIONAL UNIVERSITY
PUNJAB
2019

CANDIDATE DECLARATION

I hereby certify that the work, which is being presented in the thesis, entitled **"A Low Cost Innovative Solution for Automatic Water Quality Monitoring using Machine Learning**" in fulfilment of requirements for the award of degree of Doctor of Philosophy in Electronics and Communication Engineering is an authentic record of my own research work carried out under the supervision of **Dr. G. Geetha**. The matter presented in this thesis has not been submitted elsewhere in part or fully to any other University or Institute for the award of any degree.

Sandeep Bansal
(41500062)
School of Electronics & Electrical Engineering
Lovely Professional University
Phagwara, Punjab, INDIA

<u>CERTIFICATE</u>

I hereby certify that **Sandeep Bansal (41500062)** has prepared thesis entitled **“A Low Cost Innovative Solution for Automatic Water Quality Monitoring using Machine Learning”**, for the award of degree of Doctor of Philosophy in Electronics and Communication Engineering, under my guidance. The matter presented in this thesis has not been submitted elsewhere in part or fully to any other University or Institute for the award of any degree.

Dr. G. Geetha
Professor & Dean
Division of Research & Development
Lovely Professional University
Phagwara, Punjab, INDIA

ABSTRACT

Increasing rate of water pollution and consequently waterborne diseases are the engrossing evidence towards danger to living organisms. It becomes a great challenge these days to preserve our flora and fauna by controlling various unexpected pollution activities. Although the invention of many schemes and programmes regarding water purification has done a tremendous job, but still there is something that has been lagging. Still in few urban areas or forest areas, people do not receive proper aid for the water. The water quality is still in diffident position in such areas. With increase in population, industrialization and global warming situation is getting worse day by day. It becomes very difficult to get safe drinking water and appropriate quality water for other domestic usage and agriculture purpose. Major reasons for water pollution include undesirable increase in impurities. These may cause eutrophication of the water body, change in taste, discolouration & odour of water, water borne diseases and increase in water toxic nature etc. For water to be serviceable it should be aesthetically acceptable, chemically safe, absence of organic substances, bacteria free and radioactive elements should be absent. So, there is an urgent need to look into this situation and took the corrective and necessary actions to overcome this situation. The government is paying an attention to this problem and finding the ways to control the situation. However, major areas are not developed to the point and water quality estimation is totally dependent upon sampling at location and testing in laboratories. Manual sampling and measurements are prone to human errors and these techniques may create ambiguities in predicted output.

Engineering is always playing a vital role since decades to solve such natural disasters/issues in very efficient way. In this 21st century, there are lot of technological methodologies are inventing those can help humans to control these growing issues related to pollution. Due to consumption of polluted water there may be lot of water borne diseases; those can seriously affect the human health to dangerous levels. Also, India is a beautiful country having beautiful rivers, here humans are having tremendous religious emotions and attachments to rivers like River Ganga, but studies represent that these rivers are getting polluted due to many reasons like dumping of wastage from industries to rivers, deforestation, urbanization

etc. So, there is an urgent need to look into these problems and solve those using engineering methods.

Initially to measure water quality information we have selected the important water quality parameters as per guidelines from reputed organization like World Health Organization (WHO) and designed a sensor-based hardware to pick real time water quality information from the water resources. From this database we have calculated Water Quality Index (WQI) and also applied Machine Learning (ML) methods to further classification of water for usage and to reduce the calculation errors for water quality information. WQI is a single indicator representing the overall water quality information. Various works are presented in literature for gathering water quality information. These research studies conducted in India as well as the United Nations, and these entities finally introduced a term known as the water quality index, which simplifies the problem to measurement of the efficacy of water quality in standard terms. WQI is a water quality indicator with range 0-100 where 0–25 defines poor quality, 26–50 denotes below average quality, 51– 70 indicates average quality, 71–90 represents good quality and 91–100 defines excellent water quality.

This study includes monitoring of physiochemical parameters such as pH, Temperature, Dissolved Oxygen (DO), Turbidity and Total Dissolved Solids (TDS). Microcontroller based multi sensor system measures said parameters for detecting water contamination and incorporates communication technology for further processing and alerts. Data communication module transmits the data received from system to intended user for making alerts regarding water quality. User can check the water quality information perpetually even from far away and he can take several safety measures to evade health hazards. Facile design and low cost make the designed system captivating enough for large scale deployment.

Decision Tree approach is used for classification of water into class I-V, where class I represents excellent and class V represents extremely polluted water quality. Decision tree method gives appropriate information to estimate water quality. The permissible limits of parameters are considered as per WHO guidelines. Results calculated using ML techniques showed prominent accuracy over traditional methods. Accuracy achieved is also significant.

Also, we have designed an artificial neural network (ANN) based method for calculating the WQI to estimate water pollution level hence its quality information for usage. The WQI is a single indicator representing an overall summary of various water test results. However, selection of the weight values of the water quality parameters for WQI calculation is a tedious task. There was no standard method suggested for choosing the appropriate weights for water quality parameters while calculating water quality index. Also, it is not possible all the times to include each and every water quality parameter for estimating water quality information of water resources. Hence this approach makes sure to choose right weights even with missing parameters through artificial neural network approach. Therefore, designed approach is found to be useful in this study for calculating the weight values and the WQI in an efficient manner. This work is novel because we have proposed a methodology that uses a Soft Max function to calculate the accurate and appropriate weight values of the water quality parameters regardless of missing parameters, which were randomly decided in previous work. The results of the proposed model show increased accuracy over traditional methods. The accuracy of the calculated WQI also increased to 98.3%. Additionally, we also designed a web interface and mobile app to supply contamination status alerts to the concerned authorities. Hence the proposed work is successfully able to remove the human intervention while measuring the water quality and also ensures the best results i.e. minimum possible error to estimate water quality information. This approach could be implemented to monitor the water resources across world by the administrators to ensure exact real time information about rivers or any intend application.

ACKNOWLEDGEMENTS

I am deeply grateful to my research advisor Dr. G. Geetha for her motivating guidance throughout the duration of this research work. She provided me great freedom to choose research topic of my interest and directed and critiqued me just in time. Without her inspiration, untiring efforts, enthusiastic attitude and personal care, the accomplishment of this dissertation would have been impossible. I benefited enormously from her valuable feedback, logical analysis and comments. Her input was vital in shaping this work. She has been great sources of inspiration to me and I would like to thank her from the core of my heart.

I would like to gratefully acknowledge the School of Electronics & Electrical Engineering, Lovely Professional University (Punjab), India for offering this doctorate program. The institute and the department were always very cooperative in extending all the resource and infrastructure facilities during my research span.

I am thankful to all the reviewers of my research manuscripts for providing valuable feedback and comments that helped me to raise the standard of this research work. I also extend my special thanks to all the cited authors in my thesis whose research findings provide me a battleground to choose upon the research directions.

Words are not adequate to express my gratitude to my loving wife Dr. Urvashi Bansal for always being there for me. She never failed to give me great encouragement and confidence so that I can finish my thesis. I am highly grateful to her for all the cooperation, affection, motivation, understanding, patience and sincere advice. Her support and care helped me overcome setbacks and stays focused on my thesis. I greatly value her friendship and deeply appreciate her belief in me.

I am also grateful to my father Mr. Vinod Bansal and my mother Mrs. Geeta Bansal for their best wishes and words of encouragement. They cheer me up during walk through thorny journey towards attaining the Doctorate Degree. My love and thanks also go to my father in law Mr. Sushil Garg and my mother in law Mrs. Sangeeta Garg for their support and care that motivated me to carry out this work. Also, I express the thanks to my other family members for always being with me at any stage during this work.

Above all, I firmly believe that everything in the world is happening as per the plan and will of the Almighty. So, I thank him for planning this and giving me an opportunity to do this thesis. I pray that he continues showering his grace and blessings onto me and all others.

Sandeep Bansal

Table of Contents

List of Figures

List of Tables

LIST OF ABBREVIATIONS

ADC	Analog to Digital Converter
ANN	Artificial Neural Networks
API	Application Programming Interface
BIS	Bureau of Indian Standards
BNC	Bayonet Neill Concelman
BOD	Biological Oxygen Demand
CLI	Command Line Interface
COD	Chemical Oxygen Demand
DC	Direct Current
DO	Dissolved Oxygen
EU	European Union
GIS	Geographical Information System
GUI	Graphical User Interface
ICMR	Indian Council of Medical Research
IoT	Internet of Things
LCD	Liquid Crystal Display
ML	Machine Learning
NESREA	National Environmental Standards and Regulations Enforcement Agency
NBP	Normalized Back-Propagation
NTU	Nephelometric Turbidity Unit
ORP	Oxidation Reduction Potential
PPM	Parts Per Million
Q-FACTOR	Quality Factor
RMS	Root Mean Square
STP	Sewage Treatment Plant
TDS	Total Dissolved Solids
TSS	Total Suspended Solids
USB	Universal Serial Bus

USEPA	United States Environmental Protection Agency
WEKA	Waikato Environment for Knowledge Analysis
WHO	World Health Organization
WQI	Water Quality Index
WQM	Water Quality Monitoring

Chapter 1
Introduction

Water is the predominant substance on earth as it used for agriculture, industry, commerce sector, livestock, generation of hydropower, drinking and household needs etc. The variation in the water quality levels also affects the climatic patterns. So, the availability and the quality of water lead to the immense development of the countries. In the last few decades, numerous issues have been shaped for environment. The problems faced in ozone layer and increasing pollution for either water, air or soil affected the complete earth. With increase in pollution, a major problem like global warming is getting worse day by day. Water is the major requirement for life which is also getting polluted to huge extent. Therefore, it is the demand of today's environment to get safe water either for drinking or for agriculture. At present, major diseases are caused only due to water pollution e.g. cancer is caused due to some chemicals deposited into water from industries. With the economic growth of society and urbanization number of pollutants deposited into water is also increasing which leads to degradation in water quality to large extent. With the increase in pollution and shortage of water, serious threats to mankind are increasing. Therefore, good methodologies are required to secure and prevent water pollution. Even the government is also paying attention to this problem now a day. However, major areas are not developed to the point and water quality estimation is totally dependent upon sampling at location and testing in laboratories. This technique may create ambiguities in predicted output.

There are numerous parameters to measure quality of water e.g. pH, Turbidity, Total Dissolved Solid (TDS), Dissolved Oxygen (DO), Temperature, Biological Oxygen Demand (BOD), nitrogen, phosphate, nitrate etc. The most widely recognized strategy to distinguish these parameters is to gather tests physically and after that send them to research center for identifying and examining. This strategy squanders an excessive amount of labor and material asset, in addition to has the restrictions of the examples gathering, long-term examining, the maturing of analysis gear and different issues. If this manual sampling and laboratory testing is converted into sensor-based hardware then accuracy, sensitivity and time management will increase. It will also lead to automatic water quality estimation without manual interference. The latest communication module systems can be used to alert the administrative department about any emergency.

Water being the most fundamental need in present age, it is as basic as that to keep it clean. Despite the fact that the development of numerous plans and projects in regards to water sanitization has completed a colossal activity, yet at the same time there's something that has been lacking. In the event that you take urban regions or forested regions, individuals over yonder don't get legitimate guide for the water. The water quality is still in reverse position in such territories. The primary driver of water contamination is modern dumping of waste legitimately into water, testing of undesirable weapons like bombs by straightforwardly tossing them into water and knowing their range, oil poison, marine and so forth. Also, for the most part water contamination is finished by unfriendly increment if polluting influences, these may cause eutrophication of the water body, and change is taste discoloration and smell of water, water borne ailments, water harmful nature increments and so forth. For water to be protected water or healthy water it ought to be tastefully adequate, artificially sheltered, natural substances ought to be missing, free from microorganisms and radioactive components to be missing. As indicated by insights the WHO reports express that roughly 36% of urban and 65% of provincial India were without access to safe drinking water. Also, as an Indian culture it is progressively imperative to have clean water since they rely on it for nourishment, just as social need. Amid the past investigations it's been appeared human exercises are progressively responsible of water gets contaminating.

The principle parameters in the nature of water is characterized by physical, inorganic or concoction, dangerous metals, natural supplement and request, bacteriological, science and radioactive components. Substance parameters like hardness, calcium, magnesium, chloride, sulphate, nitrate and so on. Physical parameters like temperature, pH, conductivity, turbidity, taste, scent and so on. Harmful metals like copper, chromium, zinc, lead, mercury and so on. Natural parameter has BOD, Chemical Oxygen Demand (COD), and phenols and so on. The water quality

defilement can be diminished by taking consideration and changing physio-compound parameters, for example, debased burden, natural focuses, and temperature and so on.
In our work we focus on water quality check by utilizing few of planned frameworks. This is useful in light of the fact that it tends to be cost effective in contrast to others. Among all these different parameters we focused on five important parameters that reason a noteworthy change in water are pH, temperature, DO, turbidity and TDS. In our examination we completed an exploration on water tests by utilizing their WQI estimations of every parameter and programming them under programming tool named as WEKA apparatus which utilizes the hypothesis of decision tree.
The major task of the proposed work is to design a cost-effective hardware to monitor, analyze and predict water quality class. We have used multi-sensor for five water quality parameters. Then some machine learning model is applied to get water quality index and class. Additionally, we have used communication module for disseminating alert information to higher authorities.

1.1 Water Quality Measurement

Water quality can be measured through some parameters or characteristics of water in various circumstances. This behavior can be changes by two types of contaminates.

1.1.1 Biological contaminates

It includes some micro-organism e.g. protozoa, bacteria, algae, virus etc. These contaminates are deposited in water through industry waste, domestic run off or agriculture. There are three steps for measuring this contamination i.e. collection, incubation and enumeration. But this process in quite slow and there is no rapid measurement technology for this measurement till now.

1.1.2 Chemical contaminates

The effects of chemical contaminate are worse than biological contaminates. It includes metals, nitrates, chlorines etc. The methodologies for measuring chemical contaminate are much direct than biological contaminates.

There are some conventional parameters which gives a general idea about the water quality. Those parameters which are measured to get a clear image of environment, including the impact of the local environmental conditions and drainage basin, annual variability etc. Following are the description of some of the most commonly measured conventional parameters.

1.1.2.1 *pH:* pH is an amount of the hydrogen ion concentration of a resolution. Solutions with a high absorption of hydrogen ions have a low pH and solutions with a little concentration of H+ ions have a high pH value.

Table 1.1: pH classification for water

Sr. No	pH value	Water condition
1	7	Neutral
2	>7	Alkaline
3	<7	Acidic

Now the question may rise, which is what does pH mean for water? The pH range is better indicator of whether hardness of water. Generally, 7 is the idle pH value. Normally, acidic is considered if the pH value is lower than 7 or else it is considered as basic if the pH value is greater than 7. Furthermore, the value of pH in surface water system is 6.5-8.5 while the pH value for ground water system is 6-8.5. To create the water more acidic, a method is used to measure of potential of the water to oppose a variation in pH which is named as alkalinity. To establish the corrosiveness of the water, two factors are required such as measurement of alkalinity as well as pH. Usually, water is acidic if pH < 6.5, it could also be considered as corrosive as well as soft. Some of the metals will include in acidic water like as iron, copper, zinc, manganese and lead. Furthermore, acidic water includes high dimensions of toxic metals. Premature damage will be caused by acidic water to metal piping; also, it has linked aesthetic problems like as metallic or else sour taste.

On sinks as well as drains, it could stain laundry and affect "blue-green" staining. There are health risks linked with the mentioned toxins significantly. To thwart the water from reacting by means of household plumbing or else from supplying towards electrolytic corrosion, a solution has been feed into the water through the neutralizer. Soda ash is the characteristic neutralizing chemical. Soda ash functions to enhance the sodium content which develops pH; also, it is called as sodium carbonate. If the pH level of water > 8.5, it could be depicted with the intention of water is hard. Hard water can able to cause aesthetic dilemmas, and it does not pretense a strength risk. An alkali taste to the water will be embraced in the mentioned issues, creation of scale deposits on utensils, laundry basins and dishes, complicated in obtaining soaps in addition to detergents towards lather, at last, the creation of insoluble precipitates on clothing.

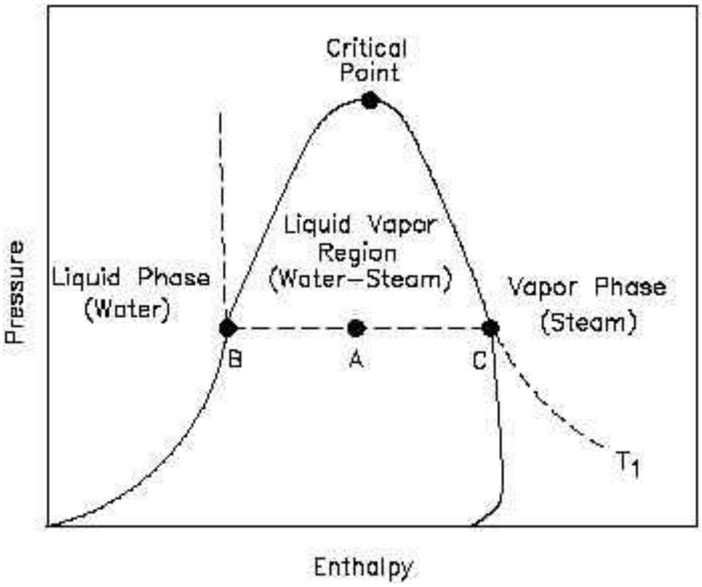

Figure 1.1: pH schematic diagram

Corresponding to a study of Wilkes University, the linking of pH by means of atmospheric gases as well as temperature is the initial cause why water samples ought to be verified on a regular basis. The strength of the acidic or else basic solution was not measured by the pH value of the water which will be explained in the study purpose, furthermore, it alone cannot supply an entire picture of the features or else restrictions with the afforded water.

The human body sustains pH equilibrium on a standard basis whereas the ideal pH dimension of drinking water should be 6 to 8.5, in addition to it will not be caused through water expenditure. For instance, there is a low pH level of 2 for our stomachs; hence it is a beneficial acidity with the intention of aids us by means of food digestion.

Scientist Sorenson has proposed the scale in the year 1909. The water depicts with the intention of H + and OH ions are equivalent in quantity which will be involved in the beneath reaction. The value of pH is 1-7 for acids and pH value for alkaline is between 7 and 14.

There are some factors that can influence pH. Some of these are:

- **Carbon dioxide:** Owing to the global warming, the rate of carbon dioxide is developing in atmosphere nowadays. At the time of rate of pH enhances, the quantity of carbon dioxide occurred in water also enhances and it depicts with the intention of pH as well as carbon dioxide are directly relative to each other. Aquatic bodies are entered by carbon dioxide through the following factors such as respiration of water bodies, marine, runoff next to rains, dumping of waste and so on; the variation in pH of water will cause the hydrogen levels of water. Carbonic acid is referred as the substances begin to pollute the carbon dioxide as well as water gathered in addition to produce an acid. The pH of water falls down owing to the occurrence of the carbonic acid. Furthermore, it will create water more acidic in nature because of the accessibility of the mentioned acid.
- **Acid rain:** Owing to the more acidic in nature, acid rains are additionally dependable for pH. Hence, the rate of pH has turned into acidic nature which causes water to be impure at the time of rain is forced onto the water.
- **Dissolved minerals:** Previously, minerals are melted within the rocks or else bottom surface of water are also dependable for pH development. They discharge the mentioned minerals which make water high contaminated.
- **Temperature:** Here, temperature as well as water is indirectly related to each other. pH decreases at the time of temperature enhances and this process will be done by vice versa. The water turns into more acidic nature at the time of temperature goes high as compared to the water.
- **Waste water:** Waste water becomes contaminated lake at the time of it is directly discarded into the lakes or else ponds from the entire parts of the systems. It includes chemicals as well as metals at the time of water are discarded with the industrial water and it causes to enhance in acidity of water.

1.1.2.2 Temperature: Temperature is an important parameter; it is the reason for bio chemical reactions in water life. Based on the water temperature, survival rate of aquatic animals can be estimated. Temperature plays a major role in aquatic life and its growth. Higher temperature causes rise in chemical and metabolic rate and converse as well. The changes in water temperature mostly happens due to the change in meteorological conditions. It may also occur due to the variation in day timings. It is usually high in morning time and stumpy at nights. Temperature not only affects marine lifespan, but it also effects other constraints in water like conductivity and pH.

Temperature is important because it impacts water chemistry. At higher temperature, the rate of chemical reactions generally increases. Its effects in water are dissolved oxygen, diffusion rate or gases, marine life. The DO concentration is affected by temperature greatly. Rather than cold water, some doctors recommend that room or body temperature water is better to drink. Warm water helps aid in digestion, help your body detox, settle an upset stomach. Cold water taste better, raise metabolism, lower fever, etc.

Temperature requirements are different for different plants. Some plants can survive in cool water while other in warm water. Tropical plants like hibiscus, orchids, gardenia etc. shows decrease in their growth, if the water temperature is below 21°C.

Aquatic life will also be affected regarding changes in water temperature. Water temperature also influences the movement of fish, their metabolism and their respiration. If the temperature of water is more (warm water) there will be an increase in metabolism, movement and respiration rate of the fish. If the temperature of the water is less, the fish will become inactive. Temperature below 5°C can be harmful to the species. This is also applicable to other insects and small aquatic organisms.

Causes of changes in water temperature are:

- Sunlight/Solar radiation
- Deforestation
- Seasonal changes
- Turbidity

Unusual water temperature may lead to:

- Dissolved Oxygen solubility greatly affects variations in water temperature. Cold water leads to high solubility of gas as compared to warm water. Hence, we can say that DO and temperature are dependent.
- Water temperature is inversely proportional to oxygen level in water and hence warm water decreases oxygen in water which leads to threat for aquatic organisms.
- Highly cold or warm water strongly affects human health and may lead to increase in diseases.

1.1.2.3 Dissolved Oxygen (DO): Dissolved oxygen is a free non-compound oxygen occurred in water or else another liquid. Dissolved oxygen is significant factor in calculating water quality. Dissolved oxygen is a necessity parameter in limnology (study of lakes), which is second to the water itself. Dissolved oxygen can cause aquatic life as well as water quality if the oxygen level is too high or else too low.

Two factors are not bonded to few other factors in oxygen such as free oxygen (O_2) and non-compound oxygen. Inside the water, dissolved oxygen is the occurrence of the mentioned free O_2 molecules. In a compound, the oxygen molecule in water has been bonded as well as it does not calculate toward dimensions of dissolved oxygen. At the time of salt or else sugar is stirred, the free oxygen molecules melt in water as much.

Surface water is generally saturated or supersaturated with oxygen due to the sufficient light for photosynthesis. A reduction is seen in the dissolve oxygen during winter season. For a healthy aquatic ecosystem dissolve oxygen is essential. Dissolve oxygen is required for the aquatic animals to survive. Different species and life stage require different amount of dissolve oxygen. Some of the requirements are given below:

For the following creations of life, dissolved oxygen is needed such as bacteria, invertebrates, fish and plants. Alike to organisms on land, the above-mentioned organisms employed oxygen in respiration. By using their gills, fish and crustaceans can able to acquire oxygen for their respiration whereas plant life as well as phytoplankton need dissolved oxygen for respiration at the time of photosynthesis.

Furthermore, one more living thing has needed dissolved oxygen such as microbes like bacteria and fungi.

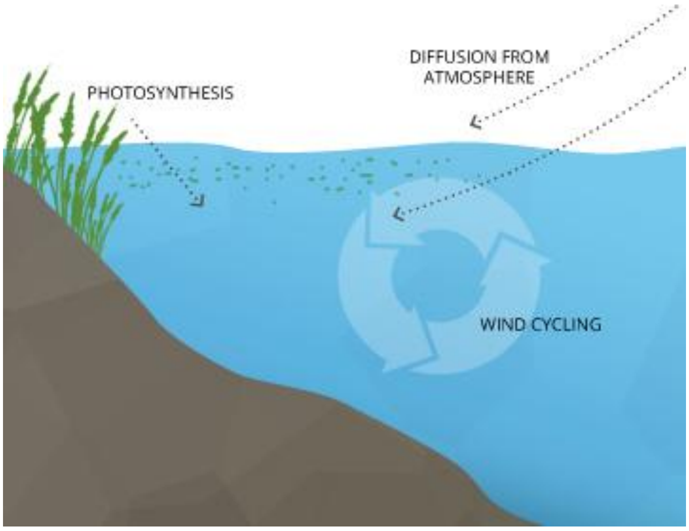

Figure 1.2 : Dissolved oxygen in water process

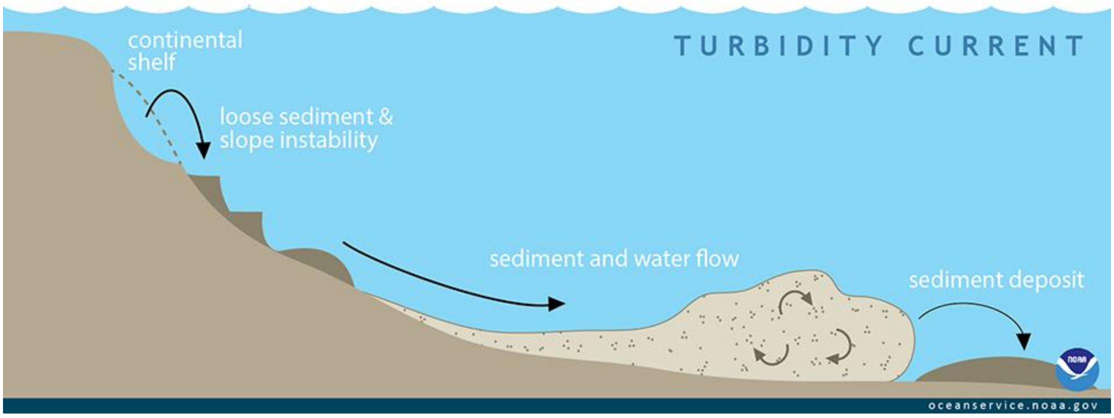

Figure 1.3 : Turbidity current

Enhance in temperature, organic contents in water, runoff after rain, high growth in aquatic life and so on are the major parameters which authorize DO of water. At that stage, a case will present such as super saturation at the quantity of progress of aquatic plant life enhances. At the time of oxygen by plant respiration goes beyond the atmospheric value, scarcity of oxygen will take place. In winter duration only scarcity of oxygen problem will happen.

1.1.2.4 *Turbidity:* Turbidity is the muddiness or fogginess of a liquid produced through the large amounts of distinct elements that are generally invisible to the naked eye. The dimension of turbidity is an important test of water quality measurement. Factors Influencing Turbidity: High water flow, soil erosion, urban runoff, wastewater pollutants, decaying plants and animals, algal blooms contribute to cloudiness of water.

The measure of comparative clearness of a liquid is said to be as turbidity. Turbidity is said to be as optical features of water as well as a term of amount of light with the intention of it is spread through material in the water at the time of a light is shined by the sample water. Turbidity becomes high if the depth of the spread light is high. The following organisms will be penetrated into the material which affects water to be turbid such as slit, algae, plankton, and clay etc.

Cloudy or else opaque will be created by turbidity. A United States Geological Survey hydrologist sampling huge turbid water in the Colorado River in Arizona has been depicted to the left picture. To detect the turbidity, the gathered water in the bottle has been utilized in which it is measured through shining a light by the water as well as it is descript in Nephelometric Turbidity Units (NTU). Turbidity's are in short, less than 10 NTU generally; because of several rivers are a clear green color and low flow (base flow). Owing to a rainstorm, particles from the surrounding land are cleaned into the river creating the water a dirty brown color, depicts water with the intention of it has huge turbidity values. Furthermore, water potentials are quicker as well as water are higher owing to high flows, in which it can able to more ease stir up in addition to suspend material from the stream bed, affecting higher turbidity.

The following factors will be caused by high concentrations of particulate matter such as recreational values, light entering and productivity, habitat quantity, and affect lakes to fill in quicker. Enhanced sedimentation as well as siltation could be presented in streams, in which it can able to conclude in damage to habitat zones for fish and other aquatic life. For some other pollutants, bacteria as well as notably metals, the attachment places will be generated by the particles. Hence, turbidity readings could be employed as an indicator of capable pollution in a water body [1].

In drinking water, high turbidity or else cloudiness is aesthetically unpleasant in addition to it might indicate a health concern. Food as well as shelter for pathogens will be generated by turbidity. Turbidity could able to endorse progress of pathogens in the allocated network if it is not eliminated, lead to waterborne disease outbreaks, in which it affects important cases of gastroenteritis by the US and the universe. Shelter will be generated by the particles intended for microbes through diminishing their revelation towards assault by disinfectants. To help in endurance of microbe, particulate material has been determined by the direction of microbial attachment. Providentially, classical water treatment processes have the potential to efficiently eliminate turbidity at the time of operated correctly.

Effects of unusual turbidity:

- It may increase the cost of treating water.
- Gastrointestinal diseases may elevate.
- High turbidity may fill the taps with mud and silt which can damage the pipes and valves.
- Some diseases e.g. nausea, cramps and headache may increase if high turbidity water is consumed.

1.1.2.5 Total Dissolved Solid (TDS): TDS is an acronym of Total Dissolved Solids, and which tells the total concentration of dissolved particles in water. TDS is a grouping of inorganic salts and organic matter. The inorganic salts present in the water are salts of magnesium, potassium, calcium which are wholly in the cation group. On the other hand, carbonates and bicarbonates contribute to the group of anions. These minerals are the results of both natural and human interventions. High levels of dissolved solids are present when the water has flooded through an expanse where the rocks have a great salt content. Surplus minerals are transferred into the water bodies due to runoff. Factors Influencing TDS are:

- Concentration of ions tells about the TDS in the water. Temperature, Vaporization (evaporation), Nutrient runoff, and several other causes associated with human activity also account for higher TDS concentrations.
- Temperature has a prominent effect on the concentration of ions, as when the temperature is high, higher is the movement of ions and hence, a larger generation of ions takes place in the waterbodies.

1.2 Machine Learning

A technique of information analysis is said to be as machine learning in which it mechanizes logical method constructing. We can also say it as a branch of artificial intelligence on the basis of concept with the intention of machines ought to be study as well as accept by practice. A region of computer science with the purpose of providing computers the potential to study in the absence of clearly programmed is said to be as machine learning.

Based on their studying “signal” or else “feedback” accessible to a learning network, machine learning objectives are generally categorized which are given below as:

- **Supervised learning:** This type of learning is used with the situations when we have been given with input and output and we have to find a mapping function for testing the data. We have used this learning mechanism in the proposed work.
- **Semi-directed learning:** This is similar to supervised learning with a difference that the dataset consists of some missing output labels.
- **Reinforcement learning:** This learning mechanism is environment specific learning. This is done to ensure maximum reward to a particular situation.
- **Unsupervised adapting:** This is the type of learning mechanism which is used when there is no output label. One of the examples of this learning is clustering technique. This is the technique when similar items are clustered in one group depending upon their properties and behavior irrespective of the output label.

Out of above four different approaches, we have selected supervised learning approach as our data set consists of water quality labels and supervised learning is best suitable and provide effective results for the datasets consisting of class labels. In the next step, we need some algorithm that can work well with parameter weights. There are numerous supervised learning algorithms available e.g. Support Vector Machine, K Nearest Neighbor, Neural Network etc. and we found neural network approach is the best suited for such cases.

1.2.1 Types of Supervised Machine Learning algorithms used in proposed study

1.2.1.1 Artificial Neural Network: Computers are good at explaining algorithmic and mathematical problems, but frequently the world can't simply be well-defined with a mathematical process. For instance, Language Processing and Face recognition is a pair of illustrations of difficulties that can't simply be calculated into an algorithm, yet these jobs are negligible to human beings. The significance to Artificial Neural Networks (ANN) is that it processes the information in a similar way to our own biological brains, by drawing encouragement from how our own nervous system function. This makes them beneficial implementations for solving problems like language processing and facial recognition, which our genetic brains can do without any effort.

Benvenuto has proposed the back-propagation algorithm [50]. This algorithm was explained in four major parts. These are illustrated as:

- Feed-forward computation
- Back-Propagation to output layer
- Back-Propagation to hidden layer
- Weight updates

The network topology shown in figure 1.5 is used as an artificial neural network in the proposed work. It consists of three layers namely input layer, hidden layer and the output layer. At input layer our inputs or information is applied for the further processing. From the input nodes the information is passes to the hidden layer. In hidden layer, as it is placed in between input layer and output layer it has no connection with the outside world. So, it performs its process and sends the output to the output layer. In output layer the information forwarded from hidden layer gets collected and it will transfer the information to the outside world. In-between the network the connection between the layers is done through communication lines which carries some weights. The result of output layer is compared with the desired output of the circuit and error is calculated. In the next step, this error is used for weights updating and result of output layer is feedback to hidden layer. This process goes on till the error becomes very minute. In this work, five input nodes have been used because five water parameters are considered for work. Along with that it contains 5 hidden nodes and one output node showing water quality level. The weights are being carried through the neural network and the evaluation is done within it.

1.2.1.2 Decision Tree: To predict values of a continuous dependent variable from one or more descriptive variables, decision tree analysis method is used as a recursive partition procedure which is the major deciding factor behind proposed scheme. Decision tree makes decisions which are more accurate as of regular systems. Decision tree consists of chance nodes, decision node and end nodes. Chance node is represented with circle and commonly used to show probability of certain results. Decision node is represented by square and it is used to make the decisions, and the finally end node represents output. Basically, a decision tree is a tree like structure which is sub divided into branches. These branches carry decision node along with them to take a decision for the end node. Our approach starts with identification of water quality class from various sensed parameters which takes decision at each step of identified parameter value. Furthermore, classification of water quality based on the sensed parameters has to be done, so decision tree is best fitted model or algorithm to serve the purpose in proposed work. Figure 1.4 represents the general structure of decision tree model.

1.3 Approach for Water Quality Measurement

WQI is a single indicator showing entire quality status of water. It is measured over a scale from 0-100, as discussed in Table 1.2. Water quality is defined by several parameters which are divided into three sub categories namely physical, chemical, and biological. Each sub category is again divided into insignificant categories e.g. physical parameters include pH, temperature, conductivity etc. chemical parameters include hardness, calcium, magnesium, chloride levels etc. and biological properties are divided into Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), phenols etc.

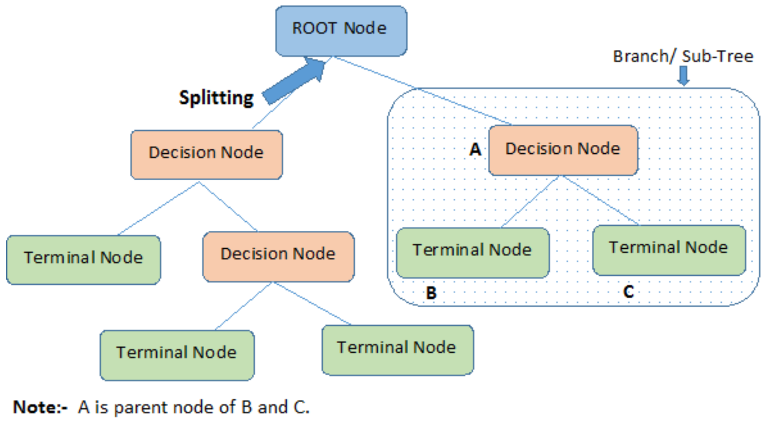

Figure 1.4 : Decision tree learning in machine learning algorithm

So, water quality can be estimated by measuring large number of parameters having some specified range defined by reputed organizations like World Health Organization [2] etc. However, the limits specified for all parameters may vary as per intended application.

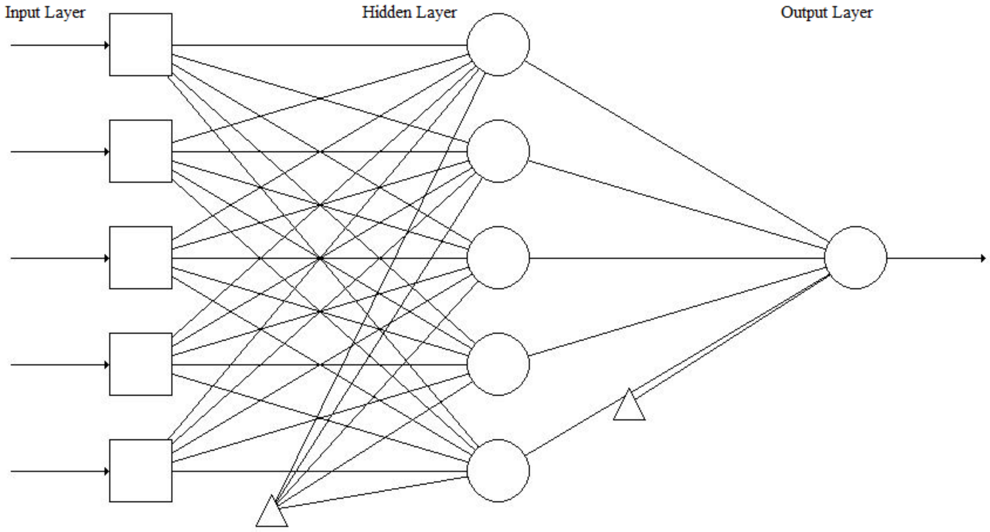

Figure 1.5 : ANN network topology used for proposed work

It is somewhat easier to measure individual parameter value and decide that this parameter is within defined range or not. However, it is difficult to decide quality of water considering interdependent effect of parameters upon each other. Numerous researches is being done in India as well as United Nations and they finally introduced a term called as Water Quality Index which simplifies the problem to measure efficacy of water quality in standard terms [3]. Water Quality Index can be estimated by using equation 1 given below. Its range lies between 0-100 as shown in Table 1.2 [4].

$WQI = \sum W_Y Q_Y / \sum W_Y$ (1)

Where, y= Available parameters

Qy= Q-value of available parameters

Wy= Weighting factors of available parameters

Table 1.2 : WQI classification range with effectiveness of water usability [45]

WQI Range	Rating of Water Quality
0-25	Poor
26-50	Below average
51-70	Average
71-90	Good
91-100	Excellent

Many authorities like World Health Organization (WHO) and Bureau of Indian Standards (BIS) have set water quality standards for drinking water and other intended uses [2][5].

Table 1.3 : Standards for Drinking Water

Parameters	Units	WHO [38]	EU [6]	USEPA [9]	Canada [9]	NESREA [8]	ICMR [5]
pH	-	6.5-8.5	6.5-9.5	6.5-8.5	6.5-7.5	6.5-8.5	6.0-8.5
Temperature	^{o}C				<15	ambient	
Total Solids	mg/l	500	500	500	500	500	500
Turbidity	NTU	<5	4	0.5-1.0	0.1-1.0	<5	2.5-9
DO	mg/l						5

These organizations set the permissible limits of various parameters for the consumer usage to reduce consumer health risk and to make sure clean water throughout world. For developing our system, after intensive research we have targeted on some important physiochemical parameters like pH, Temperature and Turbidity for Water Quality Monitoring (WQM) purpose. Table 1.3 shows various water quality parameters and their permissible limits for usage described by various organizations [6]–[10]. Where "-" represents unit less quantity/parameter. It can be observed from equation 1, there are two major factors required for calculating WQI:

- Quality Factor (Q-Factor).
- Weighting Factor

1.4 Motivation of the Work

Increasing rate of water pollution and consequently waterborne diseases are the engrossing evidence towards danger to living organisms. It becomes a great challenge these days to preserve our flora and fauna by controlling various unexpected pollution activities. Although the invention of many schemes and programs regarding water purification has done a tremendous job, but still there is something that has been lagging. Still in few urban areas or forest areas, people do not receive proper aid for the water. The water quality is still in diffident position in such areas. With increase in population, industrialization and global warming situation is getting worse day by day. It becomes very difficult to get safe drinking water and appropriate quality water for other domestic usage and agriculture purpose. Major reasons for water pollution include undesirable increase in impurities. These may cause eutrophication of the water body, change in taste, discoloration & odor of water, water borne diseases and increase in water toxic nature etc. For water to be serviceable it should be aesthetically acceptable, chemically safe, absence of organic substances, bacteria free and radioactive elements should be absent. So, there is an urgent need to look into this situation and took the corrective and necessary actions to overcome this situation. The government is paying an attention to this problem and finding the ways to control the situation. However, major areas are not developed to the point and water quality estimation is totally dependent upon sampling at location and testing in laboratories.

Manual sampling and measurements are prone to human errors and these techniques may create ambiguities in predicted output.

1.5 Research Gap

From the existing literature, we found following research gap that are addressed in the thesis.

- Water quality monitoring is likely to be affected by operator's ability.
- Obscure knowledge may exert influence on scalability and reliability of designed system.
- Inappropriate sampling rate may engender either increase in cost or decrease in accuracy.
- Water sample movement to labs for testing may increase impurities in the pathway. It will also increase delays.
- Sometimes management is not alerted on time about status of hardware system and hence the complete process is affected.
- There is a dearth of research on real time analysis which is the heart of water quality monitoring.
- Water quality weights were chosen randomly by different administrators which greatly affects water quality index value.
- Adequate and accurate sensors costs thousands of dollars and such a costly product may not be feasible for small scale system.

1.6 Objectives

Following are the objectives of this study:

1. To design and develop a cost-effective sensor-based hardware to collect real time water quality data.
2. To design and implement algorithms to automatically monitor the Water Quality Index (WQI) using Machine Learning approach.
3. To develop a Geographic Information System (GIS) web interface for data representation from various locations.

4. To develop an automatic alert system for common public and administrative departments.
5. To create a database to store all sensed data for analysis and further can be processed/used for determining health hazards.

1.7 General Approach of the Study

For the proposed study, we have employed following steps:

- **Parameter selection according to requirement:** Initially parameters are selected depending upon the purpose of water use e.g. for agriculture, drinking etc.
- **Define parameter range as per norms:** It is very important to decide the permissible range of values for individual parameters as per norms of various organization.
- **Selection of precise water parameter sensors:** Since a lot of sensors are available in the market with different level of accuracy. We need to select proper sensor according to the accuracy required.
- **Device implementation:** In the next step, proper circuit is implemented which may be installed at inlet and outlet of required location so that we can get proper readings of all set parameters.
- **Measurement of parameters:** After implementation and installation, we will note the values of parameters at different times under the various environmental conditions, so as to verify and validate our circuit.
- **Implementation of communication module:** This module is responsible for sending parameter values to the mobile device using Wi-Fi technology. This will ensure the proper monitoring of water resources from any location. Also, mobile app technology is used for monitoring the water quality.

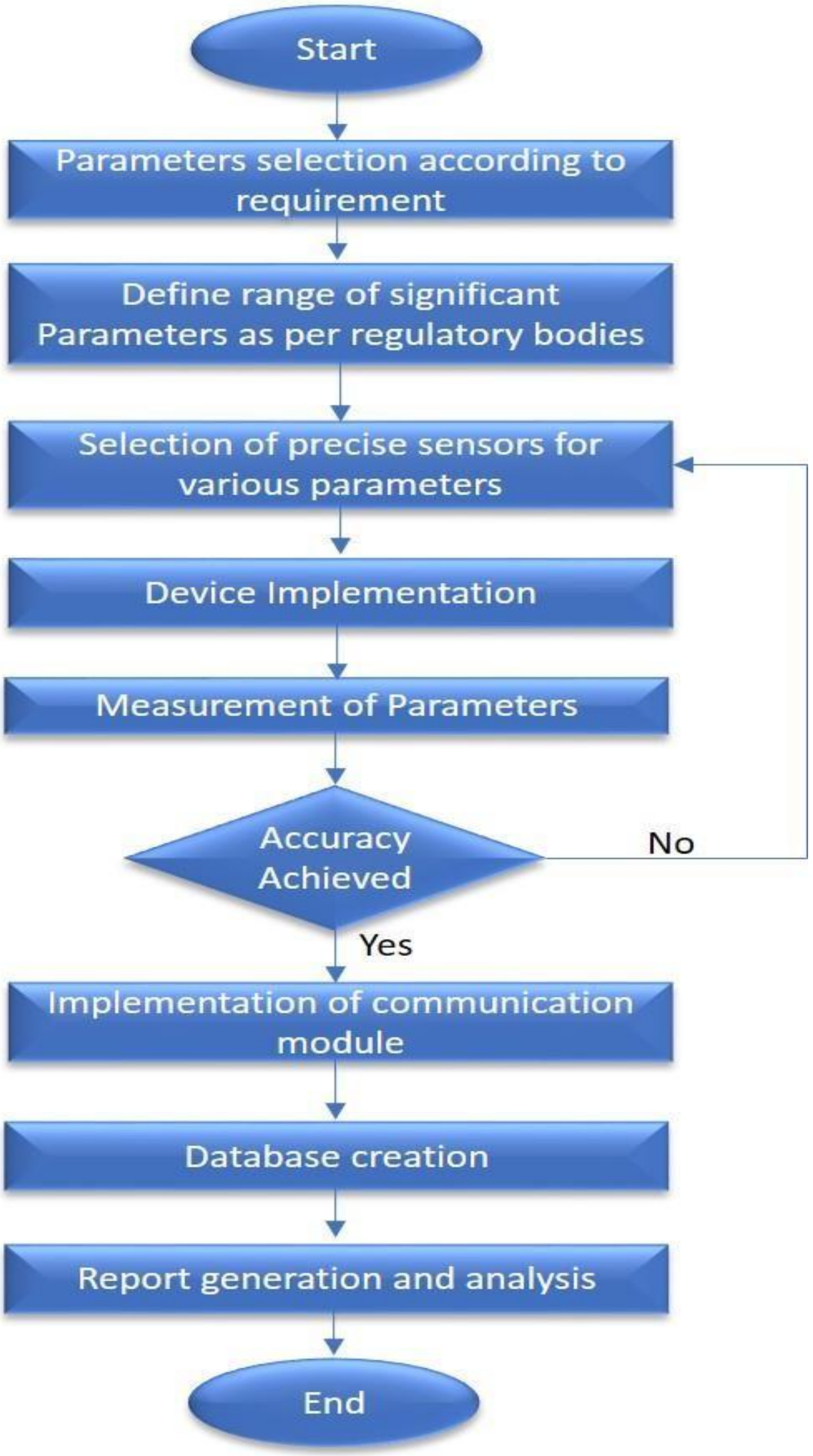

Figure 1.6 : Designed methodology

- **Accuracy check:** In this step, we measure the error rate of predicted and expected output. If this value is higher than threshold then we will make some improvements in the circuit otherwise we will proceed with next step.

- **Database creation:** In the next step, a user interface will be created using some software language that will be helpful in making entries in database for parameter values at different time. Which will further be used in data analysis in order to take multiple decisions related to health hazards or quality of water.
- **Report generation and analysis:** In the last step, reports will be generated and data will be analyzed. Also, it will be helpful in checking the effect of environment on parameter values. It will be helpful in generating alert for filter replacement if required.

1.8 Thesis Organization

The various thesis chapters are structured as follows:

- Chapter 1 provides the introduction to water quality, causes and significance of water pollution, Machine learning and its techniques, water quality parameters and water quality measurement. It primarily describes the research gaps and objectives of the research. This chapter additionally introduces the significant contribution of this research within the society. The methodology adopted for carrying out research is additionally mentioned during this chapter.
- Chapter 2 covers the comprehensive literature survey of all the relevant works done in the realm of water quality measurement, machine learning adopted for water quality estimation and water quality issues and challenges.
- Chapter 3 stress on Artificial Neural Network approach used for calculating reliable WQI. We have presented an algorithm for calculation of final weights using soft max mathematical function.
- Chapter 4 describes the Decision Tree approach used for calculating water quality class.
- Chapter 5 provides information about the designed hardware which will test and validate the proposed methodologies.

- Chapter 6 describes the tools and techniques used in this work. It additionally, focuses on general behavior and working of tools and software employed for this work. It additionally includes the snapshots of mobile application and cloud storage for historical data analysis.
- Chapter 7 concluded this research work while giving directions for future work.

Chapter 2
Literature Survey

Environment is facing a major issue in particular during few last decades. The ozone hole with its consequences (rapid changes in weather and acid rain) and the increasing pollution levels all over the earth alerted the management and the population all over the world in both developed and under development countries about the importance of preserving and recuperating the environment. Generally, quality of life depends on many aspects including environment and that growth must be sustainable growth. Therefore, water quality management and analysis are an important topic for researchers these days also [11]. Researchers started working on this area in 1965 and numerous works have been published since then [12]. Machine learning is a very popular field these days so many researchers have started working on water quality and machine learning collectively. Some of the good literature for both the fields are discussed below.

In 1966, Naeraa and co-authors have given a report showing the effect of pH and CO_2 on human blood vessels [13]. These are the most important parameters of water irrespective of type of water i.e. water present in body or ground water. They have taken various blood samples and estimated the effect of oxygen and partial pressure in different samples. They have also estimated the effect of haemoglobin concentration on these parameters and oxygen dissociation. The effect of blood gas is also estimated in this study which reveals that blood gas does not have any major effect on oxygen dissociation. Similar work on aquatic animals were done by Caldwell also [14].

In 1987, Michalski presents a report on machine learning as an artificial intelligence technique [15]. This report presents number of works published in machine learning. They explained many algorithms of machine learning which can solve different real-life problems. They suggested that cognitive science works on the concept of natural science. This science cannot be explained clearly by machine learning. Although there are many positive examples that can be solved by machine learning but still at that time, the concept was not so clear and it appears to be vague. So, the authors suggested to review more literature in upcoming years to figure out the relation between cognitive science and artificial intelligence. Some authors uses neural network in other applications also [16].

For the first time, in 1987, a methodology was proposed that calculated the water quality index for estimation of water pollution [17]. They have used Delphi equation designed by many scientists to predict the behaviour of pollutants on water quality index. They have mentioned all the pollutants in the index. The index was a multiplicative index that includes all the major factors into one system. They have used a rating scale to estimate the rate of change while quality of water changes. The index developed by them is so useful and simple that it can be understood by non-technical people that can estimate their water pollution level properly. This quality of water is further disseminated to administrators in proper format.

In 1988, a complete thesis has been developed on machine learning and its learning methods [18]. They started with introduction to machine learning and its importance in real-life problems. They discussed the purpose and methodology of learning of machines. The complete strategy is related to our brain neuron system which started learning things when a child is born. In the same way, machine is learnt from existing data and generated a trained model to test any kind of data. If the data set is of large amount and consists of variety of data then trained model will be equally good and finally accuracy will be good, but if dataset is not maintained and created properly then its accuracy will not be good. Additionally, they have described many learning methods according to different applications.

In 1991, a study has been presented that finds the relation between disasters and psychopathology. 52 studies and case analysis are taken to prove this assumption. Numerous variables are taken to analyse the case studies and finally a relation is drawn between victim, disaster and psychopathology. It was also concluded that psychopathology is greater if victims are female. It is also affected by cause of disaster. Additionally, it was also seen that the effect lowers down with time. This study opens up many avenues for researchers to find new methodologies and strategies for evaluating this relation with respect to other parameters. Studies are also undergoing to find variables that minimizes the effect of disaster on psychopathology. The author also finds some disaster factors and type of victims that actually affects this relationship.

Langley and co-authors gave a review about numerous methodologies of machine learning and its application in various fields [19] [20]. They mainly focused on rule induction and its applications. They have used many real-life applications which were successfully solved by rule induction. The complete process is divided into various stages: the first step will be formulation of problem as if problem is not understood clearly the finding a reliable solution is not possible. Therefore, it is the major step; Next step is to represent the data in a proper format. It will include number of parameters, their format and representation; Third step is collection of huge amounts of data in the format decided in the previous step as this data will be used to training purpose; Fourth step is to evaluate the trained model with testing data set. This step ensures that the trained model is producing the accurate output; last step is to generate a knowledge base expert system that will be convenient to use and take decisions.

Initially, neural network and genetic algorithm was first discussed by Chen in 1995 [21]. He used these strategies for information processing and retrieval from a stored database. A new methodology at that time i.e. connectionist modelling is used for this purpose. Additionally, algorithm and data representation of these popular techniques is also discussed in this chapter. Information retrieval is required in many fields e.g. Hopfield net, ID3, genetic algorithms etc. He clarified that there are a few learning techniques for information retrieval systems. He further clarifies that this field requires a strong and full text information collection system. Hence machine learning and genetic algorithm suits best for this kind of study.

Postolache [22] have developed a multi sensor that can evaluate water quality parameters. Additionally, this sensor has the capability to maintain online database of information. They have considered pH, temperature and turbidity of water. This communication enabled sensor is advantageous in avoiding manual intervention as well as enhance real time water quality tracking.

Breijo [23] have designed a system to measure temperature, pH, conductivity, dissolved oxygen, turbidity, Oxidation-Reduction Potential (ORP) and diverse ions. Basically, they have focused on disadvantages of lab methods that are too slow as well as continues measurement is not possible in laboratory process. They have removed these limitations. Adriana [24] have propounded a set of water parameters to be considered in healthcare units. They have also proposed some methods to measure these parameters. Additionally, they have developed computerized solutions for analyses of various water samples used in health sector.

In 2007 some authors have developed a "Smartcoast" sensor to evaluate various parameters i.e. phosphate, temperature, turbidity, pH, dissolved oxygen, conductivity and water level [25]. They have employed wireless sensor network to incorporate plug and play feature in the sensor system. Jiang [26] have designed a sensor to automate water quality monitoring process. The designed system can measure pH (1-14) and temperature (0-80 degree Celsius) of water. Regan [27] have designed a low power solution to measurement of various water quality parameters i.e. temperature, pH, turbidity, conductivity and dissolved oxygen.

In 2010, Tie-Zhu have developed a system to first measure the parameters i.e. temperature, pH and dissolved oxygen and then information is sent to monitoring centre using GPRS technology [28]. Finally, proper analysis is performed at monitoring centre. Chung have developed a hand-held device to measure water parameters based on sensor of YSI6600 [29]. The whole system is battery powered using MSP430F149. Basic advantage of this sensor was its speed, precision and small error.

Liu [30] have used Principal Component Analysis technique to find out important water parameters that may cause variances in the quality. They have started with 13 parameters which comes out to be 3 at the end of the process. Some authors employed economic power solution to evaluate water quality [31]. They have used green power source i.e. Solar energy to optimize power consumption and measured four parameters i.e. pH, turbidity, temperature and dissolved oxygen. Deqing provided low cost solution for water quality analysis i.e. they have employed DS18B20 to prepare their own sensor for temperature, pH and turbidity [32]. Chou [33] proposed a solution to enhance life time of pH and chlorine sensor. Vaddadi [34] have focused on quality of fish pond water. They have considered temperature, pH, conductivity, pressure and dissolved oxygen.

Damian [35] have discussed various parameters and their importance in various fields. They have suggested that salinity is the important parameter to differentiate between ocean water and fresh water. Turbidity is the parameter to specify changes in water quality and finally conductivity is inverse of resistance and it is also an important water quality parameter. Miskam [36] have analysed water quality in paddy fields because it may affect level of production. They have considered temperature and dissolved oxygen.

Water quality monitoring and analysis is always been a topic of discussion for researchers. At present, drinking water safety evaluation research mainly emphasizes on the influence of environment pollutants (biological, chemical and physical) on water quality and risk evaluation of human health effect. WHO, EU and USEPA and other organizations have formulated Guidelines for Drinking Water Quality, Instructions on Drinking Water Quality and standards on Drinking Water Quality [5], [6], [8]–[10], [37], [38]

Water quality is basically monitored with the help of water quality parameters. Although, numerous parameters were decided by standard organizations, various authors have considered different parameters for measuring water quality [13], [39], [40]. Once the parameters were decided, next issue is the source of data collection. Since numerous authors have worked on different sources of data collection. They have targeted a particular source and the complete analysis has been done on that source [1], [7], [41], [42]. They have analysed the quality of water in these sources and proposed different solutions for maintaining the balance of quality.

Water quality is represented in terms of Water Quality Index WQI. Numerous authors have worked on different definitions and representations of WQI [43]–[46]. These indices are very helpful in analysing quality of water. However, there is no standard formula for deciding the prerequisites for WQI. Therefore, we have worked on neural network approach for automatic generation of WQI. Numerous authors have worked on ANN techniques for different applications of water quality monitoring [47]–[49]. However, WQI measurement still needs attention these days.

In 1992, various authors have presented the idea about Back propagation algorithm [50]. This is the most popular algorithm of neural network among all approaches. It consists of layered architecture with input layer which takes input of parameters of dataset, output layer that provides final output label and some hidden layers that processes dataset. Number of hidden layers varies for different applications. This approach sues an activation function and utilized weight values for all the connection between layers. If the output is not accurate enough then these weights are modified at each iteration and the process continues till error is minimized. This is the error between expected output and predicted output.

In 1996, various authors designed a methodology for water sampling, testing and water quality estimation [51]. They have suggested some necessary actions on the basis of estimated quality. First, they have decided some strategies and techniques required for each phase of monitoring process. Then important but application specific water quality parameters were decided which needs to be evaluated to estimate quality of water. In the next step, some categories of pollutants were selected that will be deposited into water for checking the water quality variation. Finally, some sampling locations were selected which were analysed for making a comparative analysis.

In year 2000, a book was published that reviews the use of disinfectant for the treatment of polluted water [52]. They have reviewed the papers from 24 different scientists and gives a clarification that there is an impact of disinfectant on the quality of water. It is an advanced version of WHO guidelines [2] and it provides the guidelines for other 21 organic and inorganic chemical substances. In the next year Cude gives Oregon Water Quality Index for estimation of water quality on the basis of 8 parameters using a single formulation [53]. The main focus was to provide simple formulation to estimate use of Oregon stream for different purposes i.e. swimming or fishing. This formulation is an advanced and improved version of index designed in 1970. They have considered spatial and temporal water quality information also. Additionally, they have mentioned the general issues in water management and its precautions and preventions from pollution.

An article deploying a methodology to remove arsenic from drinking water was proposed in 2002 [54]. They have used various chemicals to remove arsenic e.g. manganese greensand, iron oxide coated sand and ion exchange resin. Studies reveals that iron exchange resin was best among all to remove arsenic as it has higher arsenic absorption capacity. This strategy is best suited for developed countries with urban water supply while other two are best suited for developing countries with low water supply because these approaches are simpler and easier to apply.

Sundaray and co-authors presented a multi variate analysis to estimate spatial and temporal water quality variation in Mahanadi [55]. They have studied 31 stations for this river for evaluation of nutrient in relation to physiochemical features. They estimated that there is a large variation of nutrients in unpolluted station as compared to polluted station in monsoon season. It was also calculated that BOD value is higher near Paradip port. Some fertilizer plants are there near this port which produces acidic substances and lower down its pH value. Salinity is increased near Devi region while phosphorous and nitrogen is higher in concentration near Atharbanki station. They have employed R factor analysis to predict effect of anthropogenic contribution in pollution of water.

In 2007, an article was published that describes that water parameters [5], [56] are the good indicators of system contamination [57]. The authors analyses that how changes in contamination are sensed by real time sensors. They targeted drinking water and considered ten parameters for evaluation related to different contaminates. These sensors continuously monitor water samples and some contaminates were also deposited at the same time to influence sensors readings. It was concluded that the sensors responded for chlorine, TOC, ORP, conductance and chloride. This design can be helpful in public health surveillance data to find level of contamination in publically available drinking water.

Lambrou proposed a real time water quality monitoring system for detection of contamination [58]. They targeted drinking water for in pipe monitoring and assessment of quality of water. Their major focus was simple, light weight implementation with reliable long-time operation. This technique is meant for large scale development. Initially, a market research is done to select low cost sensor. It is proceeded with monitoring, processing of analog signals, logging and representation of data. Various algorithms are also developed to estimate water contamination risk. The designed system is capable enough to detect low concentration of these contaminates. The complete system is low power and provides all processing facilities within the system only.

Yeung and co-authors proposed a methodology to estimate compounds present in surface water in other water resources [59]. They investigated the concentration of perfluorochemicals in water resources. The studies reveal that this concentration is lesser in number and different in pattern in India as compared to other countries. Some authors studies the effect of hypoxia on striped bass and food consumption [60].

They also estimated the effect of DO and temperature on this species by conducting experiments at different temperature and DO levels. It was also concluded that temperature do not have much effect on this while DO significantly affect this concentration.

In 2009, a research is done to estimate risk of chemical and radioactive contaminate on water quality [61]. The study estimated that chemical affects quality of water more as compared to other components which further increases the risk of health risk. They further estimated that the risk of chromium is greater than arsenic which is greater than cadmium while that of phosphorous is greater than fluoride which is greater than cyanide and effect of benzene is greater than nitrogen. They mentioned that it is difficult to remove all organic compounds and hence the proposed system can prioritize the compounds such that few can be removed with higher risk values.

In 2010, Doria investigated some factors that can influence drinking water quality [62]. They also suggested that water quality is influenced by organoleptic properties itself. There are other factors also which may influence water quality i.e. chemicals, risk perception etc. Authors suggested that administrators should consider all these factors to avoid potential problems and enforce policies properly. Previous history is also considered while enforcing new policies and standards. Communication methodologies were also used to disseminate risk information to customers. Various surveys can also be done to provide specific improvements. It was also suggested that everyone should be given proper education about freshwater use.

Fengyun proposed an approach to disseminate water quality information through remote sensing and Geographical Information System [63]. Author designed an integrated system for the proposed methodology also. Numerous principles of using a remote sensing technique was also designed in this work. On the basis of this study, author suggested some of the future work that needs to be done to ensure water quality management. Storey and co-authors developed a on line monitoring and warning system for water quality management [64]. They have reviewed some current literature in this field also. They have analysed that in spite of lot many technologies, there is a need of generalize system for water quality monitoring that can work on any water resource because there is always a trade-off between cost of design and ease of use of design.

Bach and co-authors analysed various chemical components and toxic components presents in polyethylene terephthalate bottles [65]. It was widely spread that plastic is not suitable for food material. Therefore, this analysis is done to find food stuff that can be placed safely in plastic containers. They also reveal that some intentionally added substances (To maintain balance between time and quality of food in plastic) can react to plastic bottle and create toxic materials. They have also concluded that some extraction and food treatment method may also have positive and negative effects on food or water. They have surveyed many toxic substances that negatively reacts with bottled water to create toxic material. A limit should be applied to ensure safe bottled water.

In 2013, detailed discussion on water quality index was done by Brown and co-authors [66]. Quality of water is always been an issue since last few decades. Therefore, it is a necessary requirement to analyse water quality continuously to ensure safest drinking water. To this end, water quality index is always been a promising approach. This index can rate any kind of water suitable for any intended use irrespective of types of contaminates present in the water. The index proposed by the authors is so sensitive to parameters values or contamination variation.

In 2014, new methodology for water quality monitoring is developed using wireless sensor network [67]. Greenhouse has emission is a major upcoming problem of pollution for all living being. In such situation, water and air quality need to measure properly which is possible with the help of new technologies and IoT. This design includes low cost nodes and probes for quality monitoring. They considered three parameters i.e. temperature, conductivity and turbidity in detail for this study. Additionally, they employed ZigBee sensor network and Wheatstone bridge amplifier for this purpose. The major drawback of this approach was periodical maintenance of the sensors. The results are verified with lab test also and proves to provide significant improvement in this scenario.

Oyem and co-authors studied seven parameters for ground water quality [68]. They said that lower pH belongs to acidic water leads to corrosion problem and high conductivity and TDS concentration leads to increase in ionic solution in water. These solutions can lead to use of water in agriculture. High Chemical Oxygen Demand increase the life of aquatic animals in water. Finally, they concluded that ground water is good for drinking and agriculture but require some pH maintenance.

Sarkar and Pandey developed a design to monitor water quality using artificial neural network approach [69]. They claimed that DO is the important factor for stream water as aquatic animals breathe in water only. They have proposed an approach for estimating DO concentration using ANN approach. They have considered four parameters in major and train their model in four different situations. Finally, the predicted concentration of DO matched with expected output and error is reduced to minimal amount. ANN is best suited in this scenario as different variables have non-linear behaviour and it can be handled easily with ANN. The only limitation is that dataset should be consistent and huge to train the model properly.

Cloete et al. [70] proposed a design consisting of smart sensors for real time monitoring of data. They considered five major parameters for this analysis. Some micro controller-based circuitry is attached to the designed system for processing of information. Zigbee wireless sensor model is used for information dissemination purpose. Numerous tests been run on this design to continuously monitor the quality of water samples. Whenever quality of water goes beyond range then communication module is used to disseminate this information. The designed system is capable enough to monitor, signal, process and transmit the data.

Hsu et al. [71] proposed a methodology to estimate river pollution using decision tree. They monitored water quality and fish ecology monthly to using surveys. Regression analysis and Decision tree with 5.0 algorithm is employed for this purpose. Finally, they claimed that the used algorithm shows prominent accuracy for the said problem.

Ruben proposed a model to estimate Chemical Oxygen Demand in rivers using ANN technique [72]. They have considered seven parameters in seven input layer, ten hidden single layer and one output layer. Once they trained the model, it shows the accuracy with root mean square error of 0.6 % only. Further they analysed the effect of variation in water quality parameters value on concentration of COD. The proposed model can be used in the process of bacterial treatment also. They have taken 110 samples for this study. Finally, the trained model is tested to ensure its accuracy level. Additionally, they found that there are three parameters which have least impact on COD and hence can be removed from input.

Lakshmi and co-authors designed a drinking WQI for estimation water quality in drinking water [73]. They considered 17 parameters for this study. First, they provide random weights to all the parameters and put them all together in one single equation of WQI. They collected 20 samples for this work. This index is geometric index rather than arithmetic index which provide the proper condition of water quality that was not represented in previous studies as claimed by the author. It also indicates the type of problem associated with water quality so that management can take corrective actions.

In recent years, Adimalla proposed an approach for estimation of chemical contaminates in ground water used for drinking and irrigation purpose [74]. As the requirement for ground water increases, its pollution also increases. They have collected 105 samples for 12 parameters. The results reveal that the ground water has maximum percentage of fluoride and nitrate. This is added in the process of rock formation which release fluoride minerals. This makes it unsuitable for drinking purpose. Ground water is generally considered alkaline and hard and not suitable for drinking in Telangana region. Sodium and SAR concentration found out to be suitable for irrigation purpose.

Authors uses modified Decision tree for aquatic environment monitoring [75]. They have installed various nodes in ponds to estimate water quality parameters. They employed distributed decision tree for the classification of water quality. They have selected 14 parameters from 21 ponds to create a dataset. After collection of data from pond, entropy is used for splitting of tree. Finally, bee colony optimization technique is used to send the results to sink nodes. The performance of proposed algorithm is compared with previous works using energy-based evaluation. Their prediction accuracy comes out to be 80%.

Some authors uses partitioning algorithm for water quality monitoring [76]. This technique is capable enough to predict the relation between output and predictors. This can work well in every situation consisting of linear and non-linear or noisy data. But the tree size is not scalable and optimized.

2.1 Crux of Literature

From the existing literature, we came across some of the problems in existing water quality monitoring system. If we consider manual collection of water sample for quality measurement, then it may lead to many problems: Obscure knowledge of operator may exert influence on water sample; too frequent and too less sampling may have its significant impact also. Next problem arises in transportation of sample to various labs for testing which may increase impurities in the pathway. It will also increase delays. Third problem arises when filters installed at the plant gets damaged and management is not well informed on time about it. This may lead to polluted water being passed through various resources which may affect the complete process. Additionally, existing systems were based on some predefined attributes values which may affect the accuracy of system. Finally, these systems are too costly and such a costly product may not be feasible for small scale deployment.

2.2 Problem Definition

This research work will focus on numerous problems that were identified in the existing literature. We will design and develop a cost-effective sensor-based hardware to collect real time water quality data. Secondly, we will design and implement algorithms to automatically monitor the Water Quality Index (WQI) using Machine Learning approach. Thirdly, we will develop a web interface for data representation from various locations. Additionally, we will develop an automatic alert system for common public and administrative departments. Finally, we will create a database to store all sensed data for analysis and further can be processed/used for determining health hazards.

Chapter 3
Water Quality Assessment using Artificial Neural Network

Increasing rate of water pollution and consequently waterborne diseases are the engrossing evidence towards danger to public health and all living organisms. It becomes a great challenge these days to preserve our flora and fauna by controlling various unexpected pollution activities. In this chapter, an Artificial Neural Network (ANN) based method is presented for calculating the Water Quality Index (WQI) to estimate water pollution. WQI is a single indicator representing overall summary of various water test results. However, deciding the weight values of water quality parameters for calculating WQI is a tedious task. Therefore, ANN approach is found to be useful in this study for deciding weight values and calculating WQI in an efficient way. This work is novel because we have proposed a methodology using a mathematical function to calculate weight values of parameter irrespective of missing values which were taken randomly in previous works. The results calculated using proposed model showed prominent accuracy over traditional methods. Accuracy rate of calculated WQI is also increased to 98.3%. Additionally, we have also designed a web interface and mobile app to alert about contamination status to the concerned authorities.

3.1 Introduction

Water is the predominant substance on earth as it used for agriculture, industry, commerce sector, livestock, generation of hydropower, drinking and household needs etc. The variation in the water quality levels also affects the climatic patterns. So, the availability and the quality of water lead to the immense development of the countries. But by 2025, it is anticipated that the water withdrawals to be increased by 50 percent in the developing countries and 18 percent in the developed countries [56] [37]. Out of 2.5 lakh gram panchayats in India only 28,000 have achieved the Nirmal Gram status [12]. World Bank estimates that 21 percent of contagious diseases in India are bound to polluted water and the absence of hygiene practices [38]. Due to practice of, sewage and garbage dumping into water, industrial waste, oil pollution, acid rains, global warming, eutrophication etc., river Ganga and river Yamuna in

India and now the first and the third most polluted rivers in the world [59]. Water pollution causes many types of diarrheal diseases, cholera, guinea worm diseases, filarial diseases, dysentery, viral gastroenteritis and amebiasis. Therefore, the quality of water determines the quality of mankind and in short, the survival of mankind and any life form is impossible without water.

But there is always a problem associated with the supply and demand of the pure water. To this end, we need to know on which degree the quality of our water is standing. So, there is a need to integrate environment engineering aspects to measure the water pollution to decide the water quality. There are many parameters of water that comprehend its quality like Turbidity, BOD, dust, volcanic gases, suspended solids, phytoplankton, algae, temperature, conductivity, solar radiation, etc. As per guidelines by WHO, five of the major parameters necessary for deciding water quality are Dissolved Oxygen (DO), Turbidity, Temperature, Power of Hydrogen (pH), Total Dissolved Solids (TDS) [38] [77]. Therefore, we have considered these five parameters for this study. A large dataset containing 700 samples was collected from authenticated resources corresponding to these five parameters. This dataset is further used for ANN training. Finally, the trained ANN can be used for water quality calculation for any sample. The quality of water calculated from the proposed model can further determines the uses of it i.e. whether it can be used as drinking water or for irrigation or for domestic purposes [4], [49], [74] [69]. The novelty proposed in this work is by using artificial neural network to evaluate water pollution level by detecting WQI with minimal possible error. We have also calculated weight values for parameter. Additionally, the designed system is completely automatic which can detect unwanted situations e.g. failure of filters which leads to deterioration in quality of water and the system can alert the authorities for the same.

3.2 Water Quality Index

WQI is a single indicator showing entire quality status of water. It is measured over a scale from 0-100. Water quality is defined by several parameters which are divided into three sub categories namely physical, chemical, and biological [42]. Each sub category is again divided into insignificant categories e.g. physical parameters include pH, temperature, conductivity etc; chemical parameters include hardness, calcium, magnesium, chloride levels etc; and biological properties are divided into Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), phenols etc.

So, water quality can be estimated by measuring large number of parameters having some specified range defined by reputed organizations like World Health Organization [38] etc. However, the limits specified for all parameters may vary as per intended application.

It is somewhat easier to measure individual parameter value and decide that this parameter is within defined range or not. Numerous researches is being done in India as well as United Nations and they finally introduced a term called as Water Quality Index which simplifies the problem to measure efficacy of water quality in standard terms [40]. Water Quality Index can be estimated by using equation 1 of chapter 1. Its range lies between 0-100 where 0-25 defines Poor quality, 26-50 below average, 51-70 average, 71-90 good and 91-100 defines excellent water quality [46].

Many authorities like World Health Organization (WHO) and Bureau of Indian Standards (BIS) have set water quality standards for drinking water and other intended uses [5], [38]. These organizations set the permissible limits of various parameters for the consumer usage to reduce consumer health risk and to make sure clean water throughout world. For developing our system, after intensive research we have targeted on some important physiochemical parameters like pH, Temperature and Turbidity for WQM purpose. It can be observed from equation 1 of chapter 1, there are two major factors required for calculating WQI:

- Quality Factor (Q-Factor).
- Weighting Factor

3.2.1 Q-Factor

Q-Factor indicates the quality of water for individual parameter [45]. Where 100 indicate the excellent water quality and 0 indicates very poor water quality. The value of each parameter is given in standard and different terms e.g. pH is unit less, temperature is measured in degree Celsius, TDS unit is ppm, and Unit of turbidity is NTU etc. Although these units are standard, we cannot combine different unit data in single formula of WQI. Therefore, we need a common term for all parameters to put them in single formula. To this end, each parameter value is converted into q factor. We have designed equations 2 – 4 to calculate q factor from standard terms for different parameters. Q-factor for different water quality parameters can be calculated by using following equations:

$$pH_Q = ((-0.776289 \times pH^3) + (11.6597 \times pH^2) - (37.9749 \times pH) + 32.1271) \quad (2)$$

$$Temp_Q = ((0.00723232 \times t^3) - (0.261429 \times t^2) - (0.844661 \times t) + 80.6558) \quad (3)$$

$$TDS_Q = ((1.37374 \times 0.000001 \times tds^3) - (0.00108052 \times tds^2) + (0.128167 \times tds) + 80.0909) \quad (4)$$

For Turbidity, the conversion from NTU to Q-factor is done by mapping using Table 3.1. For Dissolved Oxygen, the conversion from PPM to Q-factor is done by first converting it to %saturation and then to Q factor by mapping the values given in Table 3.2.

Table 3.1 : Conversion from NTU to Q factor for Turbidity

Transparency (cm)	**Turbidity (NTU)**	
Reading from Tube	Use in database	**Q-Value**
150	0	97
120	5	84
90	10	76
>60 (turb tube)	<15 (turb tube)	70
60	15	68
30	20	62
27.5	25	57
25	30	53
22.5	35	48
20	40	45
15	50	39
12.5	60	34
10	70	28

7.5	80	25
5	90	22
2.5	100	17
<2.5	>100	5

Table 3.2 : Conversion from ppm to Q factor for DO

Temp C	**Solubility (mg/L)**
0	14.6
1	14.2
2	13.8
3	13.5
4	13.1
5	12.8
6	12.5
7	12.2
8	11.9
9	11.6
10	11.3
11	11.1
12	10.9
13	10.6
14	10.4
15	10.2
16	10
17	9.8
18	9.6
19	9.4
20	9.2
21	9
22	8.9

23	8.7
24	8.6
25	8.4
26	8.2
27	8.1
28	7.9
29	7.8
30	7.7

3.2.2 Weight-Factor

Weight factor defines the importance of each parameter in deciding quality of water. Although, tremendous works have been done in this field to decide weights for parameters for calculating WQI, there is no standard formula for assigning weights for parameters. Different researchers have taken different value of weight factor for various parameters [1], [17], [44], [78]. A summary of various works related to water quality weights is shown in Table 3.3.

Table 3.3 shows the weights assigned to each parameter by different researchers. Although, number of parameters were different in each study, mean deviation of weight from value 1 (Sum of weights are 1) should be same for a parameter in every study. The table 3.3 shows the % mean deviation of assigned relative weight. As it can be analysed from Table 3.3 that mean deviation of pH in different studies are 0.02, 0.06 and 0.6 which is not even closer to each other.

Table 3.3: Mean deviation of weights assigned for different parameters

Parameter	Kangabam et al., 2017 [44]		Ewaid et al., 2018 [78]		Debels et al., 2005 [1]		Dinius et al., 1987 [17]	
	Relative Weight	**% Mean Deviation**	**Relative Weight**	**% Mean Deviation**	**Relative Weight**	**% Mean Deviation**	**Relative Weight**	**% Mean Deviation**
pH	0.091174	0.026536364	0.091	0.063636364	0.1	1.7	0.077	0.633333333
TDS	0.100291	0.938236364	0.1	0.963636364	-	-	-	-
Temperature	-	-	-	-	0.1	1.7	0.077	0.633333333

Turbidity	0.087527	0.338163636	0.087	0.336363636	-	-	-	-
Hardness	0.051057	3.985163636	0.051	3.936363636	-	-	0.065	1.833333333
Nitrate	0.109409	1.850036364	0.109	1.863636364	0.07	4.7	0.09	0.666666667
Nitrite	0.093727	0.281836364	0.093	0.263636364	0.07	4.7	-	-
Orthophospate	-	-	-	-	0.12	0.3	-	-
e. coli	-	-	-	-	-	-	0.116	3.266666667
Sodium	0.058351	3.255763636	0.058	3.236363636	-	-	-	-
EC	0.116703	2.579436364	0.116	2.563636364	0.06	5.7	0.079	0.433333333
Alkalinity	-	-	-	-			0.063	2.033333333
Ammonia	-	-	-	-	0.13	1.3	-	-
BOD	0.072939	1.796963636	0.072	1.836363636	0.17	5.3	0.097	1.366666667
Chloride	-	-	-	-	-	-	0.074	0.933333333
COD	0.072939	1.796963636	0.072	1.836363636	0.17	5.3	-	-
Coli	-	-	-	-	-	-	0.09	0.666666667
Color	-	-	-	-	-	-	0.063	2.033333333
DO	0.145878	5.496936364	0.145	5.463636364	0.18	6.3	0.109	2.566666667

The same analysis can be done on other parameters also. Its ultimate leads to probable/ estimated error rate while calculating Water Quality Indices. Hence, we need a standard mechanism for estimating weights and WQI as well. According to the situation and application, Back Propagation method suits best for this purpose.

3.3 Design Methodology

Algorithm 1 gives the complete procedure for carrying out this work. The complete process is divided into following phases:

Algorithm 1: NBP: Normalized Back Propagation

***Input**: Water quality parameters values with WQI for training, Q-factor mapping values, Water sample to be tested.*

***Output**: Þ: Modified accurate weights, Water Quality Index (WQI) for the sample to be tested.*

1. *Calculate the q factor values for the dataset.*
2. *Create a report for the q factor values of water parameters values with their WQI.*
3. *Feed the database into Multi Backpropagation tool to train the model.*
4. *Collect the parameters value using designed hardware consisting of various sensors according to the water quality parameters to be tested.*
5. *Apply the sample values collected at step 4 to the trained model generated at step 3 to find WQI.*
6. *Finally, calculate the reliable parameter weights using soft max function.*

3.3.1 Water quality parameter estimation and mapping

In the initial step, we have used various sensors to estimate the value of five parameters of water i.e. pH, Turbidity, Temperature, TDS and DO. We have used heavy duty sensors for different parameters e.g. SEN0161 for pH, DS18B20 for temperature, SEN0189 for Turbidity, SEN0244 for TDS and SEN0237 for DO. They can be used in critical conditions also. Additionally, their reliability is much greater than any other sensor. In the next step, each parameter value is converted into q value using equations 2 - 4, table 3.1 and table 3.2.

3.3.2 ANN Validation

In this step, 696 samples were collected from various authenticated sources. The WQI value for these parameters was calculated in a lab. Finally, a dataset is prepared consisting of individual parameter value and lab calculated value of WQI for each sample. This dataset is feed to neural network for its training. We have used Back propagation neural network for this purpose. The main goal of back propagation algorithm is to optimize the weights in a way to make neural network more efficient to learn that how to correctly map inputs to outputs. The accuracy of this function is measured in terms of cost function or error/loss function (not to be confused with the Gauss error function), the loss function is a function that maps values of one or more variables onto a real number intuitively representing some "cost" associated with those values [48]. This error is calculated by comparing expected output with output of outer layer and difference is represented in terms of loss function. The Back-propagation method fed back the error and output to hidden layer and this process is repeated until the error is reduced to the minimum possible range. Finally, weight values were calculated that are going from input layer to hidden layer and hidden layer to output layer. Then, soft max function [79] is applied to calculate normalized value for weights for individual parameter as given in equation 5 below.

$$W_i = e^x / \sum e^x \qquad (5)$$

3.3.3 Alert generation and notification

This is the last step of the proposed methodology. In this phase, data (whose WQI needs to be calculated) is fed to the trained ANN network and result in terms of WQI is displayed on LCD screen. Alert generation feature is also introduced in the designed system by setting alarms for any unusual tracing of WQI which represents that any filter is not working so that action can be taken accordingly. Additionally, the data is also stored on cloud and a mobile app has been prepared for monitoring of all the activities and dataset collection. Wi-Fi module is used for storing the data into online database. This database can be further used for analysis of historical data.

3.4 Results and Discussions

In this work, the integration of hardware and back propagation algorithm has been done and there are significant improvements being noticed in the results as of traditional methods. Root Mean Square (RMS) error curve is shown in figure 3.1 for training data water quality information. This error is calculated while training of ANN. This curve is termed as error curve showing root mean square error. It is showing the dropping rate of error with increasing rate of training set. In the above graph the X-axis shows the number of times neural network trained and Y-axis shows root mean square error. It can be analysed from figure 3.1 that error is reduced to negligible amount.

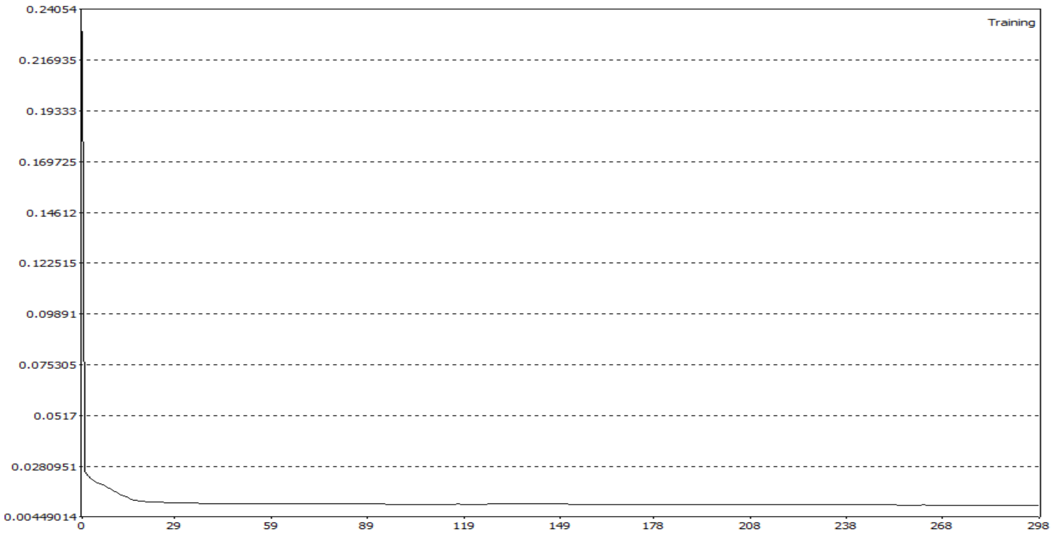

Figure 3.1 : Training of back propagation neural network measured by RMS error

With the applied training set to the software it will undergo into a loop or a cycle and exhibits methods to reduce the error rate. In this training, it went under 298 loops and shows the rapidly reducing error rate. As compared to the first and last iteration, the error rate gradually decreases and becomes very minute at the end of the loop.

Figure 3.2 shows the graph generated between network output and desired output. It can be observed from above stated analysis that the error rate is very much minimal with neural network approach for measuring water quality information. After the analysis and measurement through neural network all the information is also transmitted to mobile app as well as online data storage for further processing.

We have visited nearby places to measure the water quality index of different sources of water. Although, these sources are nearer to each other but there is a huge difference in their WQI as shown in table 3.4. We have tried to find the possible sources of pollution in these water resources. The study reveals that Budha Nallah is dumped with untreated industrial wastage from Ludhiana industries and hence its quality is very poor. This polluted water is being dumped into the Sutlej at 15 km downstream from Phillaur and we have taken the samples 5 km downstream further to this place. Therefore, this source quality is also poor. Sirhind canal and Beas river water is comparatively better than previous resources. It could be used for irrigation purpose but it is not suitable for drinking purpose. Ground water is best among all resources tested so far. We have also shown these locations on geographical map as shown in figure 3.3.

Table 3.5 shows the comparison results of different optimization techniques. We have applied PID3 [80] and Decision tree [75] approach on our data set (696 samples) and calculated the accuracy for all the three techniques. We have shown 5 samples from a large population of 696 samples in table 3.5.

Table 3.6 shows the weight values obtained from Backpropagation method. Final output layer weight was taken for individual parameter and soft max function is applied on these values to get final weight values as shown in table 3.7.

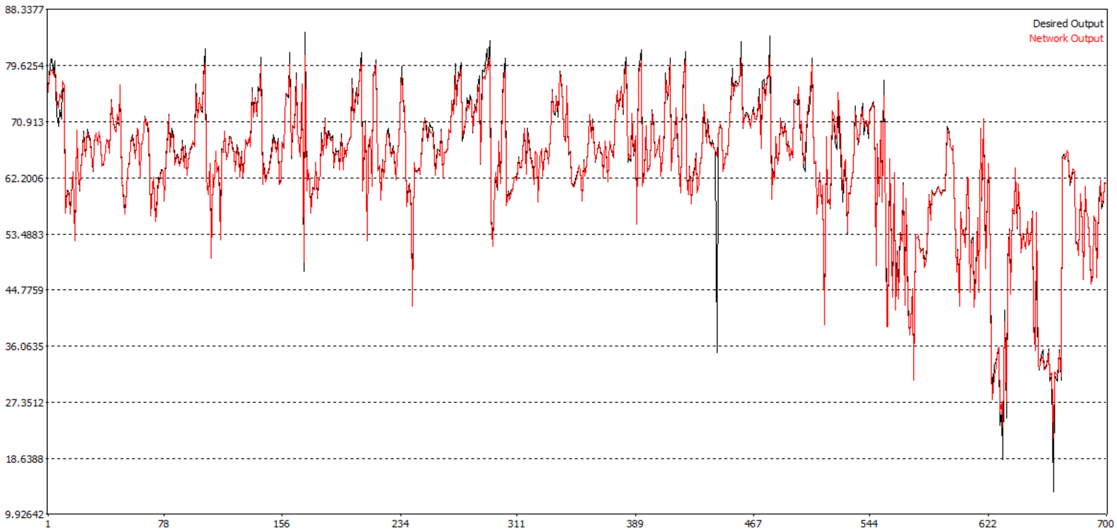

Figure 3.2 : Output vs Desired (Training data)

Table 3.4 : Comparison of WQI from different water resources

S. No.	Sampling Location	Coordinates	Average WQI using ANN
1.	Beas river, Kapurthala	Latitude: 31.368489 Longitude: 75.346263	61.03
2.	Budha Nallah, Ludhiana	Latitude: 30.931188 Longitude: 75.714377	5.66
3.	River Sutlej (20 km downstream from Phillaur)	Latitude: 30.973369 Longitude: 75.623442	17.27
4.	Sirhind Canal, Sirhind	Latitude: 30.586669	64.84

		Longitude: 76.40295	
5.	Ground water	Latitude: 31.372463 Longitude: 75.552464	68.04

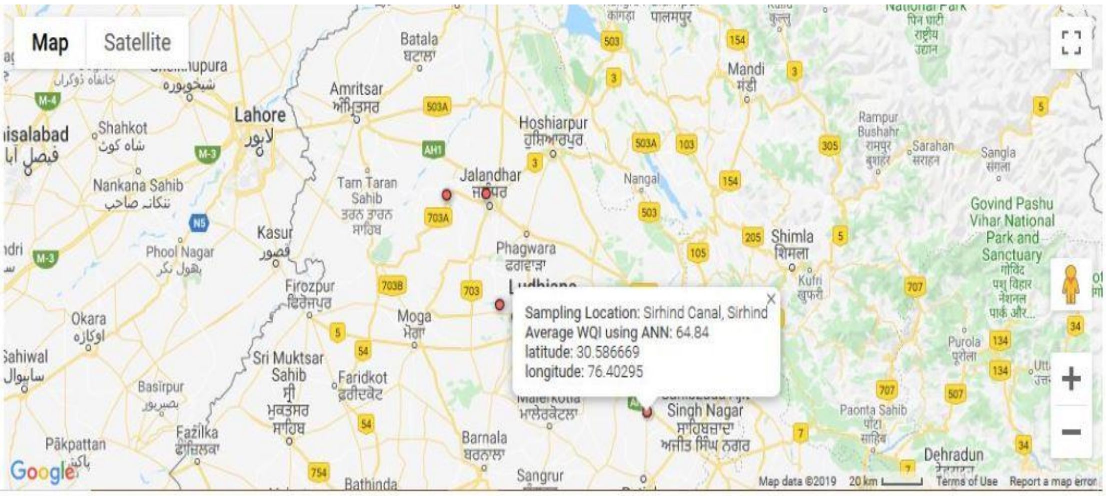

Figure 3.3 : Geographical locations for different water resources

Table 3.5 : Comparison of optimization techniques over same data set

Optimization method	**Water sample**	**WQI/ water quality classification using standard formulae and lab tests**	**Estimated WQI/ water quality classification using different optimization techniques**	**Accuracy (%)**
PID3 [80]	Sample 1	40.4/ Below Average	Below Average	96.23
	Sample 2	26.4/ Below Average	Poor	
	Sample 3	68.4/ Average	Average	
	Sample 4	55.43/ Average	Average	
	Sample 5	62.3/ Average	Average	
Decision Tree [75]	Sample 1	40.4/ Below Average	Below Average	97.4

	Sample 2	26.4/ Below Average	Poor	
	Sample 3	68.4/ Average	Average	
	Sample 4	55.43/ Average	Average	
	Sample 5	62.3/ Average	Average	
NBP (Proposed Approach)	Sample 1	40.4	40.14	98.3
	Sample 2	26.4	26.56	
	Sample 3	68.4	68.23	
	Sample 4	55.43	54.78	
	Sample 5	62.3	61.97	

Table 3.6 : Weight values from input layer to hidden layer and hidden layer to output layer

	From Input Layer				
To Hidden Layer	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron	5th Neuron
1st Neuron	-0.418938	-0.87416	-0.421608	-0.0697913	-0.0965121
2nd Neuron	1.16881	0.147598	0.592252	0.723902	0.949286
3rd Neuron	0.398993	0.634462	-0.606276	0.676783	0.559296
4th Neuron	-0.0597412	-0.328126	-0.219032	1.54461	0.613572
5th Neuron	-0.198823	0.420973	0.760568	0.868954	0.933312
	From Hidden Layer				
To Output Layer	1st Neuron	2nd Neuron	3rd Neuron	4th Neuron	5th Neuron
Output Neuron	-2.06164	2.77701	0.789522	1.46769	1.14347

Table 3.7 : Final weight values for individual parameter

S. No.	Parameter	Value x	e^x	Weight value
1	Temperature	-2.06164	0.127245117	0.00491724
2	pH	2.77701	16.07089706	0.62104172
3	TDS	0.789522	2.202343454	0.08510708
4	DO	1.46769	4.339200003	0.167683498
5	Turbidity	1.14347	3.137637098	0.121250453

3.5 Application of Proposed Approach

We have developed a prototype to showcase the working of proposed approach. Once this prototype is implemented on a large scale, it can be installed on the outlets of the water resource where water treatment plants have been established. We have tested the sample of double distilled water with our hardware and compared the results with lab results as given in table 3.8. The major variation is there in DO only which is due to oxygen level in environment, temperature change and water movement, as DO depends upon these factors. The designed hardware will measure the water quality at regular intervals. This information may be sent to administrators/ management using the mobile app. As soon as water quality goes below the permissible range then management can take the actions properly which were delayed in previous approaches due to intermediate person's carelessness. Hence, the proposed approach is the best possible solution for reducing human errors and lack of knowledge about manual water quality estimation.

Table 3.8: Distilled water sample results

Parameter name	**Lab results**	**Calculated Results**
pH	6.54	6.72
Temperature	25^0C	28.25^0C
Turbidity	0.046 NTU	0.02 NTU
DO	14 mg/L	6.42 mg/L
TDS	20 mg/L	13.53 mg/L

3.6 Conclusions

Different physical and chemical parameters are considered for this work to monitor real time water quality information. Five main parameters including pH, Temperature, TDS, DO and Turbidity are considered to assess water quality. This study reveals that neural network approach is showing significant improvement in tested values over traditional methods. This study focuses on Back Propagation algorithm for training and weight value calculation. Some formulations were given for calculating q value from parameter value automatically. Water Quality Index (WQI) is also calculated during this study. The proposed scheme showed 98% accuracy over traditional methods. We have also calculated weight values for each parameter automatically. This is the first time that a methodology is proposed to calculate reliable weight values instead of assigning random weights to the water quality parameters. Additionally, we have developed web interface to analyse historical data for taking future decisions and a mobile app for alert generation if water quality falls beyond limits.

Chapter 4
Automatic Water Class Estimation using Decision Tree

Increasing rate of water pollution and consequently waterborne diseases are the engrossing evidence towards danger to living organisms. It becomes a great challenge these days to preserve our flora and fauna by controlling various unexpected pollution activities. In this chapter we have presented Machine Learning (ML) approach for calculating the Water Quality Index (WQI) and classification of water quality to estimate water characteristics for usage. For analysis, decision tree method is used to estimate water quality information. The standard value of parameters is selected as per guidelines provided by World Health organization (WHO). Results calculated using ML techniques showed prominent accuracy over traditional methods. Accuracy achieved is also significant i.e. 98%. Likewise, projection of gathered data was done utilizing web interface and web app to alert about contamination to the authorities.

4.1 Introduction

Water being the most essential need in present generation, it is as essential as that to keep it clean. Although the invention of many schemes and programs regarding water purification has done a tremendous job, but still there is something that has been lagging. The main causes of water pollution are industrial dumping of waste directly into water, testing of unwanted weapons like bombs by directly throwing them into water to identify their range, oil pollutant etc. [11]. Major reasons for water pollution include undesirable increase in impurities. These may cause eutrophication of the water body, change in taste, discolouration & odour of water, water borne diseases and increase in water toxic nature etc. For water to be serviceable it should be aesthetically acceptable, chemically safe, absence of organic substances, bacteria free and radioactive elements should be absent. According to statistics WHO reported that approximately 36% of urban and 65% of rural India were without access to safe drinking water [2]. Various studies have been presented those shows that human activities are more responsible for water contamination [81].

The water quality can be measured by checking its various properties like physical, inorganic or chemical, toxic metals, organic nutrient and demand, bacteriological, biology and radioactive elements [51]. Chemical parameters include hardness, calcium, magnesium, chloride, sulphate, nitrate etc. Physical parameters include Temperature, pH, conductivity, turbidity, taste and odour etc. Amongst all these parameters, 5 important parameters have been considered in this work that plays an important role while measuring water quality. These parameters are pH, Temperature, Dissolved oxygen (DO), Turbidity, Total Dissolved Solids (TDS). The novelty proposed in this work is by using Machine Learning (ML) approach to evaluate water quality with minimal error possible. Training of the system with assigned mechanisms and samples collected from authenticated sources to predict the repercussion of water in flourishing years. Present work focuses on water quality monitoring by integrating machine learning approach with sensor-based network. This is ready to lend a hand because it is cost effective and easy to operate.

4.2 Decision Tree

Computers are good at explaining algorithmic and mathematical problems, but frequently the world can't simply be well-defined with a mathematical process. Machine Learning is a thriving field these days, with an increasing credit that ML can play a vital role in variety of applications, such as natural language processing, data mining, image processing, and skilled system [64]. ML is very much helpful in solving problems related to these areas and set to be a scope in near future. ML serves lot of applications including Speech Recognition, Making Predictions and Classifications etc.

Decision tree makes decisions which are more accurate as of regular systems. Decision tree consists of chance nodes, decision node and end nodes. Chance node is represented with circle and commonly used to show probability of certain results. Decision node is represented by square and it is used to make the decisions, and the finally end node represents output [19]. Basically, a decision tree is a tree like structure which is sub divided into branches. These branches carry decision node along with them to take a decision for the end node. As classification of water quality based on the sensed parameters has to be done, so decision tree is best fitted model or algorithm to serve the purpose in proposed work.

4.3 System Architecture

Water Quality Index (WQI) tells us the entire quality status of water after multiple parameter evaluations. It is measured over a scale from 0-100, as discussed in Table 4.1. As water being more efficient in nature along with-it water carries several parameters which are divided into 3 sub categories namely physical, chemical, and biological [53]. Each sub category is again divided into insignificant categories. Like physical parameters include pH, temperature, conductivity etc. Chemical parameters include hardness, calcium, magnesium; chloride levels etc. and biological properties are divided into Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), phenols etc.

Consequently, water quality can be estimated by measuring large number of parameters, those carries some specified range defined by reputed organizations like World Health Organization [2] etc. Also, the limits specified for all parameters may vary from place to place and sample to sample. Various parameters can be accepted by water research experts, but it is not easily understood by the public or by the scheme makers that who are responsible for water improvement schemes. Numerous researches are being done in India as well as United Nations and they finally introduced a term called as Water Quality Index which simplifies the problem to measure efficacy of water quality in standard terms.

WQI factor ranges between 0-100 by observing the graphical curves those are designed for water quality parameters. Table 4.1 shows the WQI classification range [46].

Table 4.1 : WQI classification range with effectiveness of water usability

Range	Class	Rating of water quality
0-25	V	Poor
26-50	IV	Below average
51-70	III	Average
71-90	II	Good
91-100	I	Excellent

Many authorities (WHO, Bureau of Indian Standards) have set water quality standards for drinking water and other intended uses [2], [5]. These organizations set the permissible limits of various parameters for the consumer usage to reduce consumer health risk and to make sure clean water throughout world. For developing our system, after intensive research we have targeted on some important physiochemical parameters like pH, Temperature and Turbidity for WQM purpose.

4.4 Design Methodology

The complete procedure for designing this work is divided into various steps as portrayed in algorithm 1. Figure 4.2 shows the step by step methods followed for designed structure.

4.4.1 Dataset preparation

Calculation of WQI is made possible after calculating Q-value for each parameter. Various researchers have proposed the sample curves for mapping [45]. Some of the relations used for mapping of parameter value with Q-Factor are shown in figure 4.1. Dataset have been prepared for network training. Sample data have been collected from various authenticated resources and used for learning and validation process. Here each parameter is firstly converted into q factor. Equations 2 – 4 of chapter 3 are used for conversion from parameter value to Q-Factor of Temperature (T_Q), pH (pH_Q) and Total Dissolved Solids (TDS_Q).

For Turbidity, the conversion from NTU to q factor is done by mapping using the figure 4.1. Additionally, Dissolved Oxygen q-factor calculation is done by help of percentage saturation value of DO. The conversion from ppm to q factor is done by converting it to percentage saturation first and then to q factor by mapping equivalence as given in figure 4.1.

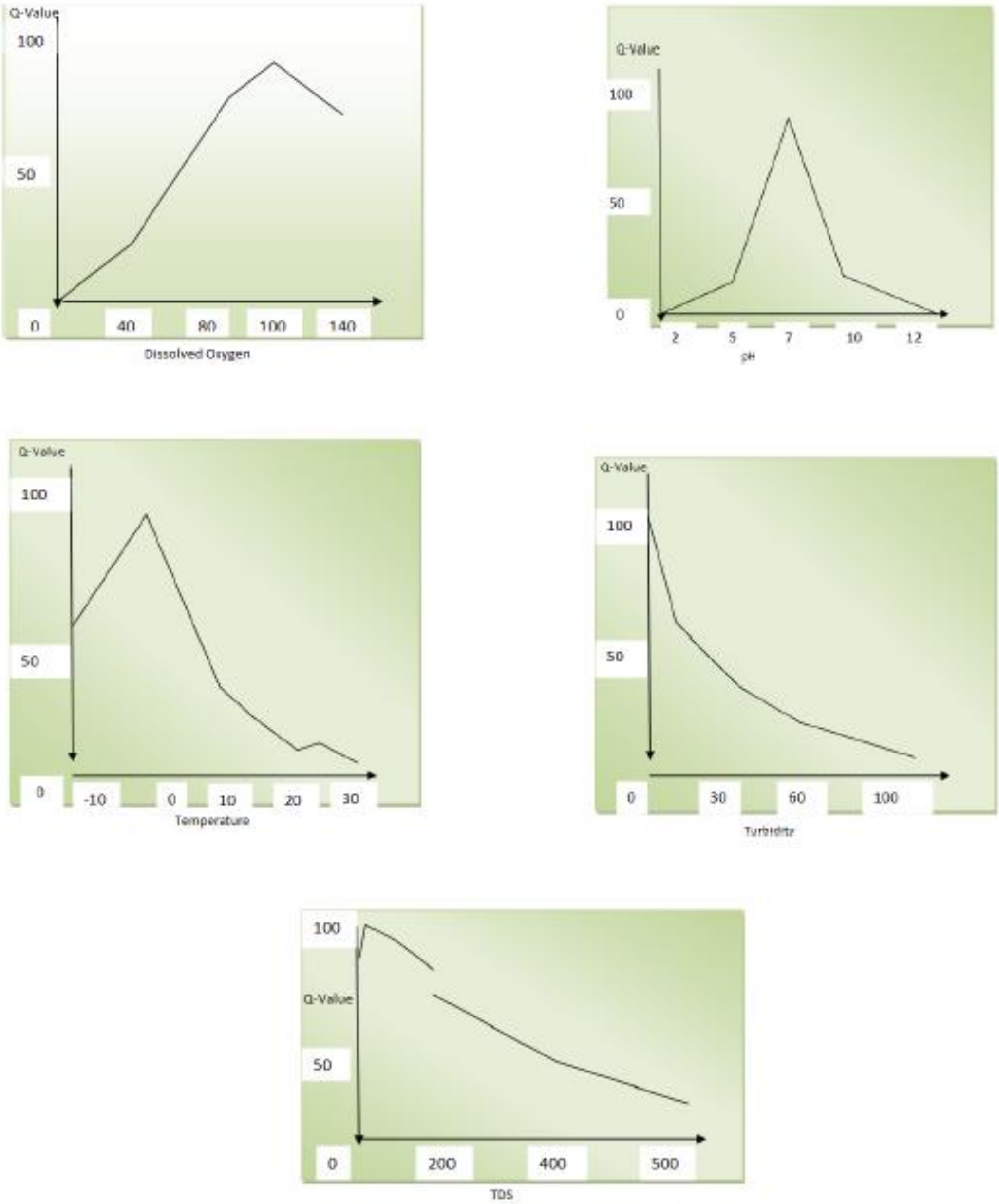

Figure 4.1 : Mapping of parameters with Q value

Algorithm 2

Input*: Water quality parameters values with WQI for training, Q-factor mapping values, Water sample to be tested.*

Output*:* Þ*: Water quality for the sample to be tested.*

---

1. *Calculate the q factor values for the dataset.*
2. *Create a report for the q factor values of water parameters values with their WQI.*

3. *Feed the database into WEKA tool to train and generate the decision tree model.*
4. *Collect the parameters value using designed hardware consisting of various sensors according to the water quality parameters to be tested.*
5. *Apply the sample values collected at step 4 to the trained decision tree model generated at step 3.*
6. *Finally, the quality of water sample can be determined after step 5.*

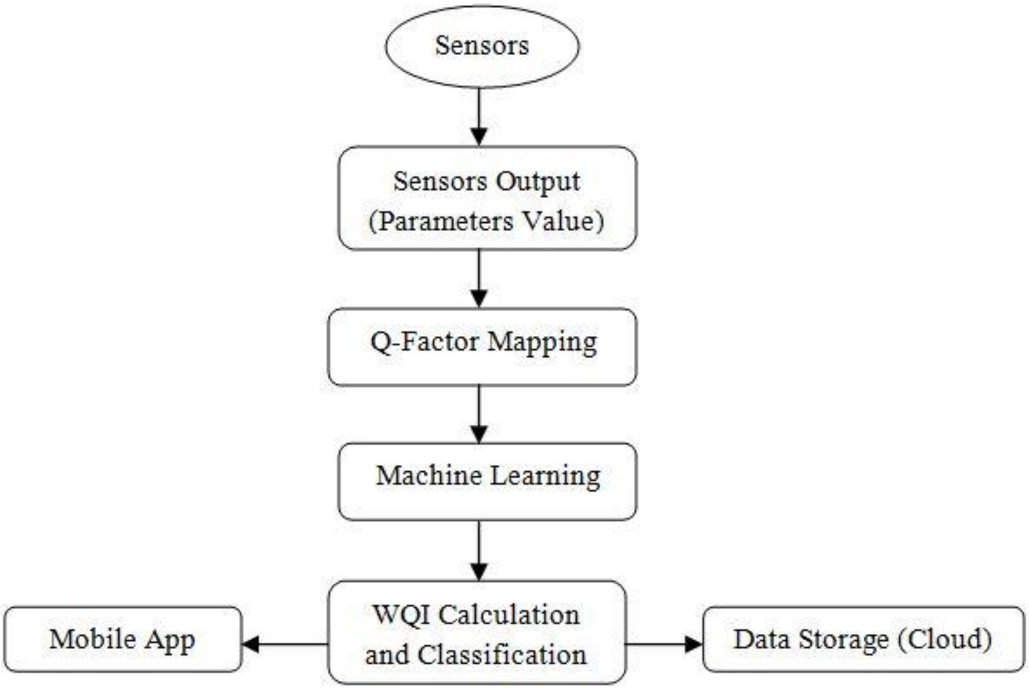

Figure 4.2 : Flow Chart of Designed Structure

4.4.2 WEKA tool

WEKA is open source software that runs in Java platform. It allows you to upload any dataset and apply any algorithm on this dataset to produce required analysis results. We have used decision tree model for training and testing of water samples. WEKA uses an algorithm named J48 for the required purpose. J48 is an improvisation over ID3 algorithm. This algorithm generates a tree based upon the dataset consisting of labelled data. To test any water sample, we need to go through this tree from root node in appropriate direction to reach the leaf node which gives the final water quality estimation.

4.4.2.1 Water quality estimation: The values of each parameter are collected through sensors by dipping them into water up to recommended level. The results are fed to the decision tree and finally water quality is estimated.

4.4.2.2 Alert generation and cloud: Alert generation feature is also introduced in the designed system by setting alarms for any unusual tracing of WQI or class identified. So that action can be taken accordingly. Along with that data is also stored on cloud and a mobile app has been prepared for monitoring of all the activities and dataset collection. Wi-Fi module is used for storage of the data into mobile app and an online database.

4.5 Results and Discussions

Decision tree model has been proved very much compatible for identifying water quality in this work. Large database collected from various authenticated resources has also helped in better learning capability of machine and also helped in validation of tested samples. Water quality is classified based upon 5 classes i.e. Class I, Class II, Class III, Class IV and Class V as discussed in Table 4.1.

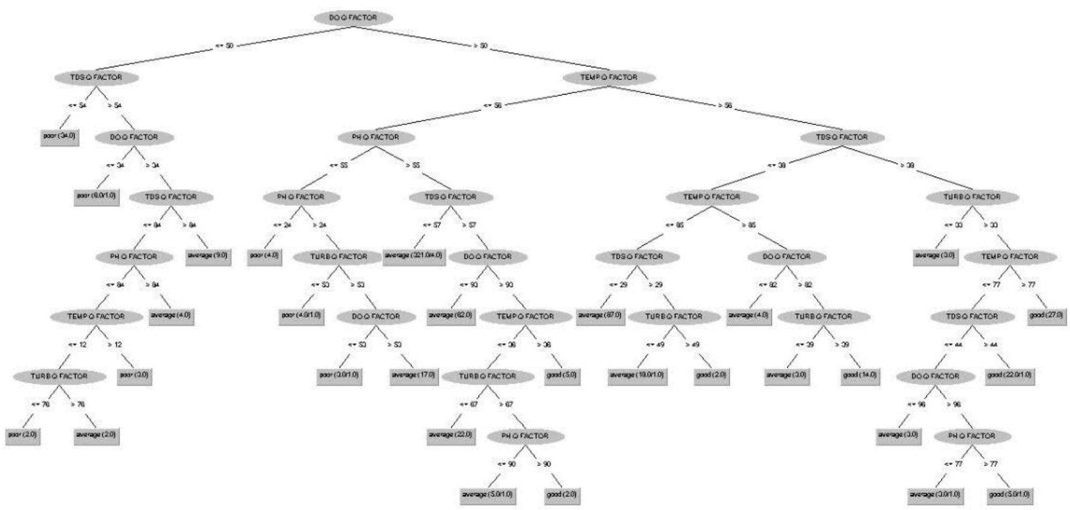

Figure 4.3 : Decision Tree Output Graph for Measuring Water Quality

The above classification is done in decision tree form by using Weka Software after an adequate training with 696 samples and afterwards 200 samples used for result validation. Figure 4.3 portrays the resultant decision tree graph for testing of water quality based on selected parameters.

Confusion Matrix shown in Table 4.2 represents the accuracy achieved through decision tree approach. It shows that overall 684 cases were exactly identified as per training data, whereas only 12 instances were incorrectly identified. It leads to 98.28% accuracy achieved while identifying water quality classification through decision tree method. Along with that it shows mean absolute error as 0.0199 and Root Mean Squared error of 0.0996.

Table 4.2 : Confusion matrix for decision tree output

Class a	**Class b**	**Class c**	**Classified as**
75	3	0	Class a (Good)
2	556	3	Class b (Average)
0	4	53	Class c (Poor)

After the measurement and analysis through decision tree model all the information is also transmitted to mobile app as well as online database storage for further processing. Graphical analysis is also possible from web data and alarming situations can be noticed and informed to concerned authorities.

Table 4.3 portrays the comparison results of NBP approach with other traditional techniques. We have also used the same dataset for calculating WQI using standard formulae provided by various organizations. The estimated water quality is compared with lab results and the accuracy comes out to be 80% approx. Additionally, Qing and co-authors have proposed parallel ID3 approach for water quality classification and the achieved accuracy was 95% approx. [80]. Hence NBP approach provides significant results over the traditional methods.

Table 4.3 : Comparative analysis of proposed approach with other traditional methods

S. No.	**Methodology opted for water quality calculation**	**Accuracy achieved (%)**
1	Standard WQI formulae	80.02
2	Parallel ID3 [80]	95.78
3	NBP (Proposed Approach)	98.3

4.6 Conclusion

Implemented algorithm is easy to understand due to its structure and division. Prediction is the key feature of a decision tree which analyse the present data and further can be used to predict future data. Different physical and chemical parameters are considered for this work to monitor real time water quality information. Five main parameters including pH, Temperature, TDS, DO and Turbidity are considered to assess water quality. This study reveals that machine learning approach is showing significant improvement in tested values over traditional methods. Water Quality is estimated during this study.

Chapter 5
Equipment, Tools and Interfaces

In this chapter we have discussed about the hardware selection criteria and also detailed discussion about the software's used for various analysis purposes during the work. Selection of water quality parameters and hence particular sensors/probes is a tedious task. For parameters selection guidelines provided by WHO has been taken care [1] and for selecting the appropriate sensors various studies and factors has been considered as the designed hardware ought to have properties like robustness, consistency, ability of managing harsh atmosphere etc. The designed hardware is having capability to be used for outlets of Sewage Treatment Plants (STP), river quality monitoring services, agriculture purposes, and drinking water purposes and lot many applications.

An initial approach towards water quality monitoring in real world is the use of quality equipment and technology used for measurement and analysis. Numerous analytical methods have been proposed by various authors to monitor water quality, but most of them have suggested the manual sampling methods to monitor the quality of river water or any other resource. There are lots of drawbacks of using manual sampling as listed below:

- Lack of operator's knowledge
- Inadequate sampling
- Inappropriate delays between sampling and testing
- Effects due to samples transportation

Hence, it is better to have a system capable of measuring parameters value on field that can avoid all these listed problems. The designed system is capable to work in accordance to the requirements discussed above. Furthermore, the machine learning approach makes it more suitable and utmost reliable for water quality measurement applications.

5.1 Topography of the Study Area

Punjab state stretches from 29°32°'-32°32°'N latitude and 73°55°'-76°50°'E longitude, occupy a land of 50,362 sq. kms in the India's north-western part. Roasting summers, torrential monsoons and breezy winters depict the climatic surroundings of the landscape that is shattered by the Ravi, Beas, Sutlej and Ghaggar rivers and their tributaries. Physically, the topography of Punjab can be divided into the upper portion of the sub-Shivalik area and the rest of Punjab is located on the Sutlej - Ghaggar river basin. The Shivalik area at an altitude of 400 to 700 meters higher than sea level is made up of fluvial deposits of conglomerates, clays and silts-all.

5.2 Sampling Locations

We have chosen some important rivers and canals as study area for sampling purpose. Samples have been collected from these resources over the duration of 2 years (i.e. 2016-2018) to measure the various physio-chemical parameters used for the study. Water quality studies were conducted on these samples. Table 5.1 below represents the sampling locations:

Table 5.1 : Sampling locations with coordinates

S. No.	Sampling Location	Coordinates
1.	Beas River, Kapurthala	Latitude: 31.368489 Longitude: 75.346263
2.	Budha Nallah, Ludhiana	Latitude: 30.931188 Longitude: 75.714377
3.	River Sutlej (20 km downstream from Phillaur)	Latitude: 30.973369 Longitude: 75.623442
4.	Sirhind Canal, Sirhind	Latitude: 30.586669 Longitude: 76.40295
5.	Ground water	Latitude: 31.372463 Longitude: 75.552464

5.3 Sensors/Probes Selection

Selection of water quality parameters is as per the guidelines for water quality regulation by WHO and other standard organizations like Bureau of Indian Standards (BIS) [2] etc. These organizations set the acceptable limits for all these water quality parameters for different purposes like domestic purposes, human consumption, agriculture or other needs. Also, they use to authorize which chemical or physical parameter must be considered on priority basis and tested on regular basis to guard the health of the end users and to make sure the water is safe and sound. Mostly physiochemical parameters have been concentrated for monitoring. For this work the five important physio-chemical parameters have been selected for water quality analysis i.e. pH, TDS, DO, Temperature and Turbidity. For taking the direct measurements of these parameters from water resource various sensor probes was used as given in table 5.2.

Table 5.2 : Water parameter sensors

Parameter	**Sensor Used**
pH	SEN0161
TDS	SEN0244
Temperature	DS18B20
Turbidity	SEN0189
Dissolved Oxygen	SEN0237

5.3.1 Temperature Sensor

We have used DS18B20 (Temperature sensor) for water temperature measurement. It measures temperature from -55°C (min) to +125°C(max) and has an accuracy of ±0.5°C, over a range of temperature from -10°C to +85°C. It is water proof temperature sensor. Figure 5.1 shows the temperature sensor used for this work.

Figure 5.1 : DS18B20 Temperature Sensor

The main feature of this sensor is direct-to-digital temperature sensor. It has variable resolution of 9-12 bits corresponding to 0.5^{o}C-0.0625^{o}C increments respectively. It has a great feature of power up in even very low power.

5.3.1.1 Specifications of DS18B20: The sensor technical specifications are:

- Sensor's width is 4mm and length is 90cm.
- Stainless steel probe tube length is 35mm and width is 6mm.
- 3 wires are included called as: 1. Red wire for VCC 2. Black wire for GND 3. Yellow wire for DATA.
- Time taken is less than 750ms (quick conversion).
- It is a single wired sensor and it can be further shared with multiple sensors.
- It is a single wired interface; hence it needs one digital pin.
- Temperature ranges from -55 degree to +125 degree Celsius
- Compatible with 3-5.5V power supply.

5.3.2 pH Sensor

We have used pH electrode/probe SEN0161 to measure the values of pH in this work. Potential difference (or voltage) is being measured of the solution in which it is sunken using Nernst equation, and it gives output as corresponding pH value. Figure

5.2 shows the pH probe (SEN0161) used to take pH value of water resources. This is an expert pH sensor probe with modern anode. It is reliable and long-life capable sensor probe, which makes it very reasonable for long haul web-based checking. It has a power indicator, a BNC connector. It is easy to use and interface. This mechanical pH probe is manufactured of delicate glass layer with small impedance. This probe can be utilized in variety of pH estimations with quick reaction and reliability. It has great reproducibility, is hard to hydrolysis, and can dispose of fundamental soluble base error. From 0-14 pH territory, the yield voltage is direct. The indication framework which comprise of the Ag/AgCl gel electrolyte salt scaffold has a stable half-cell potential.

This sensor is the new type of ascertaining pH because the past sensors were extremely hard to associate with controllers or processor boards. Also, they were over the top expensive. SEN0161 sensor probe is very cost efficient and is exceptionally intended for association with controllers or processor boards. The principle highlight is that it has high precision and ease extend so this makes it to use by most of researchers in venture fields.

Figure 5.2 : SEN0161 pH sensor

5.3.2.1 Specifications of SEN0161: The sensor have following technical specifications:

- It carries power of 5.0V.
- In potentiometer gain is adjustable

- pH2.0 BNC connector
- Quick response <=1min

To use SEN0161 pH sensor, it has to be properly calibrated with the help of standard pH 4 and 7 solutions. While using it, offset errors needs to be properly taken care off while mapping the pH values from standard solutions.

5.3.3 Turbidity Sensor

Turbidity sensor (SEN0189) is embedded into the designed hardware for measuring the turbidity value directly from water resource. It is used to estimate water quality through dimension of turbidity levels. Turbidity is dependent on the measure of Total Suspended Solids (TSS), so it transmits the light and notice the dissipating rate, which further depends on the rate of TSS present in the water. Turbidity level varies if TSS varies in the water. It has simple as well as computerized sign yield modes. One can choose the mode as per the controller as limit is customizable in computerized signal mode. Turbidity sensors can be utilized in estimation of water quality in waterway and streams, wastewater and effluent estimations etc.

It gives us the measure of turbidity range present in water by estimating the measure of suspended solids present those are mindful discouraging light to go into water. The figuring that is done between dissipating worth and light transmittance will furnish us with the data of how much absolute suspended solids are accessible. Total suspended solids and turbidity are contrarily relative to one another, as TSS expands then there is decline in turbidity esteem and when TSS diminishes then the dimension of turbidity is moderately high. This sensor has a double nature as it has both advanced and simple yield modes. The yield mode can be chosen by us which can be either advanced or simple relying upon the smaller scale controller unit that we are utilizing.

5.3.3.1 Specifications of SEN0189: The sensor have following technical specifications:

- Power voltage required approx. 5V DC
- It carries weight of approx. 30g

- Temperature limits from -10 degree till 80 degree Celsius

Figure 5.3 : SEN0189 Turbidity Sensor

- Time taken for response <500ms
- Digital output is high/low level signal
- Analog output is 0-4.5v
- Output method is analog
- Operating current maximum is 40mA
- Insulation resistance is 100m-ohm
- Dimensions are 38mm*28mm*10mm

So, this sensor is best suitable for the designed application and gives reliable and accurate values of turbidity levels present in water. Its design configuration makes it best fit for such applications.

5.3.4 TDS Sensor

SEN0244 (TDS sensor/Meter Kit) is having compatibility with latest controllers used for embedded technology for estimating TDS estimation of any water resource, to mirror the cleanliness of the water. It is dedicatedly designed for testing of any class of water may it be domestic water, industrial water, hydroponics or different applications.

TDS is the measure of solvent solids (in mg) disintegrated in per litre water. It has an inverse relation to water quality, as more solvent solids break up in water, more the TDS levels and lesser the water quality. Water with TDS level above 500mg/L is not recommended as per guidelines by esteemed organizations.

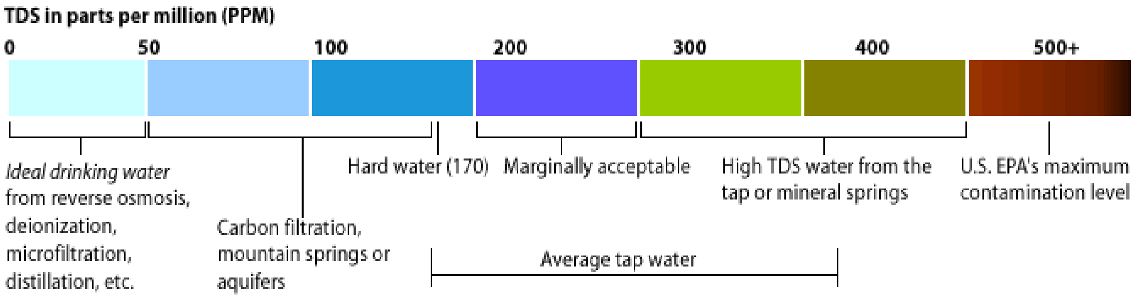

Figure 5.4 : TDS in parts per million (PPM)

TDS probe is a generally utilized hardware to quantify TDS esteem. It is cost effective, and it is simple to use, yet it can't transmit information to the control framework for web based observing to do some water quality investigation. This expert instrument has high precision and can send information to the control framework. It is waterproof in nature; it very well may be drenched in water for long time estimation.

Water higher than 55 degree Celsius is not preferred for this probe. It must stay a bit far away from the container as this will not affect the reading. The connector and signal transmitter board of device is not waterproof.

5.3.4.1 *Specifications of SEN0244:* The sensor technical specifications are:

- It takes voltage of 3.3 to 5.5v.
- It gives output voltage of 0-2.3v.
- It has electrode interface of XH2.54-2P.
- Its module size is about 42*32mm.

Figure 5.5 : SEN0244 TDS Sensor

- It has module interface of PH2.0-3P.
- Accuracy of TDS is +-10%FS (25 degree).
- Working current required is 3-6 mA.
- Length of the wire is 83cm.

5.3.5 Dissolved Oxygen Sensor

Dissolved Oxygen is the measure of oxygen level available in water. It is measured in mg/L or PPM. SEN0237 is the sensor unit for DO measurements directly from groundwater or any other type of water. It is totally compatible with controller board used for further processing. It provides additional circuitry of signal conditioning and processing so that low level signals picked from water can be used for calculation and broadcasting. Level of DO present in the water informs about the quality status of water as a famous saying “Good fish always deserve good water”. It is used in many applications such as aquaculture, common science, observing conditions etc.

Filling arrangement is 0.5 mol/L of Sodium Hydroxide (NaOH) solution. It requires precise calibration before use into actual field. It devours a very little amount of oxygen during the estimation. It would be ideal if tenderly blend the arrangement and let the oxygen to be conveyed uniformly in the water.

5.3.5.1 Specifications of SEN0237: Technical specifications of the sensor are:

- Probe type is Galvanic, no requirement of polarization time
- Membrane cap and Filling solution can be changed and having low maintenance
- 3.3~5.5V wide-range power supply, compatible with various controller boards, Dimension: 42mm * 32mm
- Converts input signal into 0~3.0V approx. output signal, compatible with various controller boards having ADC functions
- Easy to use, plug and play features make it more user friendly
- Response Time: Up to 98% full response, within 90 seconds (25)
- Pressure Range: 0~50PSI, with BNC connector

Figure 5.6 : SEN0237 Dissolved Oxygen Sensor

5.4 Hardware Prototype

In order to sense the real time water quality data, we have designed a prototype by integrating the sensor nodes/probes discussed in previous section. The designed hardware is capable of taking real time data/reading of dedicated water quality parameters from various water resources. The portable and heftiness design makes it appropriate for different water quality monitoring applications like drinking water, agriculture, aquaculture and industrial purposes. The circuit operates on approx. 5V DC supply. Power can be provided through Type-B USB male connector or adapter of 5V dc adapter with input supply of AC 100~240V, 50/60Hz. Figure 5.7 below shows the power ports available on the prototype.

Also, in order to make it user friendly and easy to use, we have provided the five connectors for the five sensors on the front side as represented in figure 5.8, sensors can be easily removed and inserted whenever required. Staring from left very first connector is BNC type connector for pH sensor, second is also of BNC type connector for DO probe, third is for TDS sensor, forth is for temperature sensor and extreme right is for turbidity sensor connectivity.

Figure 5.7 : Power supply connectors

TFT coloured screen is used on the top of the designed hardware to show all the sensed parameter values along with water quality class. It helps the intended user or layman to read the water quality status directly from it without any difficulty. Snapshot of the screen used is shown in figure 5.9 below.

Figure 5.8 : Connectors for water quality sensors

Figure 5.9 : Output at screen for different parameters and water class

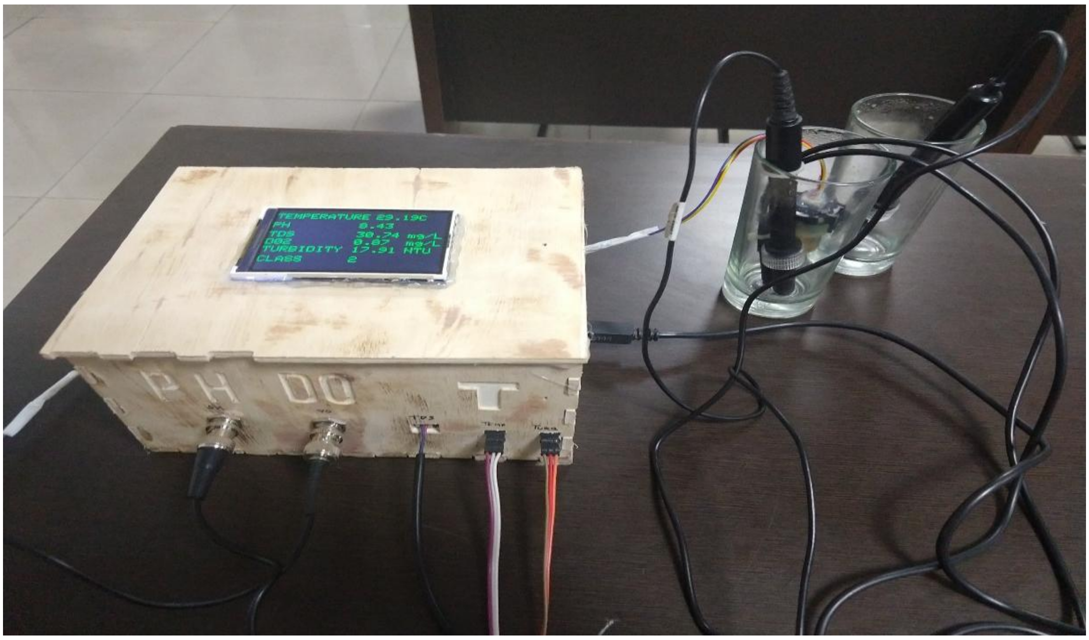

Figure 5.10 : Experimental setup

This LCD screen have six values displaying on it those are temperature, pH, TDS, dissolved oxygen, turbidity and water class respectively.

Table 5.3 : Cost analysis of existing system with designed system

Device Name	Parameters considered	Cost
Multiparameter Probe (HI98196 – HI7698196)	pH, DO	51,200 INR
Multiparameter – HI98194	pH, ORP, EC, TDS, Salinity, DO, Pressure, Temperature	95,100 INR
HI9829 Multiparameter with GPS	pH, ISE, EC, DO, Turbidity	1,94,425 INR
YSI Pro plus Multiparameter water quality meter	DO, Conductivity, Salinity, Resistivity, TDS, pH, ORP, Ammonium, Nitrate, Chloride and Temperature	$1290 + Cable cost for various sensors
Seametrics Multiparameter water quality loggers	Water level, Conductivity, pH, ORP, DO, Turbidity, Temperature	$5360

Manual Sampling	Any parameter as per requirement	200000 INR + 50000 INR per year
Designed System	pH, Temperature, Turbidity, TDS, DO	30000 INR

Figure 5.10 shows the experimental setup for testing of different samples in order to validate the circuit results, for this we taken different samples from different water resources and tested the water quality status from the designed hardware. Table 5.3 shows the cost comparison among different existing devices and designed system.

Chapter 6

Software and Tools

While carrying out the data analysis during the research work various tools and application have been utilized. This section discussed about Weka tool, Multi-Backpropagation tool and Thing Speak application used for data analysis and data demonstration.

6.1 Multi Back-Propagation

Motivation behind backpropagation to train a network with multiple layers neural network such that the high learning efficiency can learn internal representations for mapping of input to output. Multi Back-Propagation is a best suitable tool for this purpose. It is a freeware software available for training of neural networks using Back Propagation algorithms. It has features like easy to use, highly configurable, fast training, provides RMS error graphs during training, provides output graphs compared with desired graphs, allows input sensitivity analysis etc. It has a Graphical User Interface for setting up various parameters like learning rate, conditions on RMS error etc. as shown in figure 6.1 below.

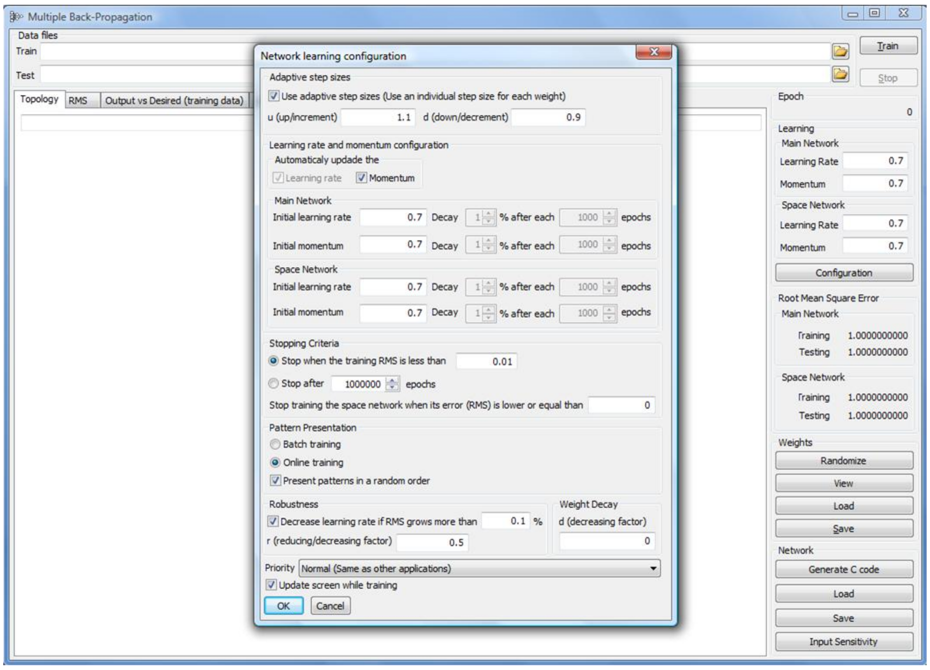

Figure 6.1 : Interface of Multi-Backpropagation tool

The main features of the tool include setting up the network topology by setting up number of input layers, hidden layers and output layers. Interface for the same is shown in figure 6.2 below.

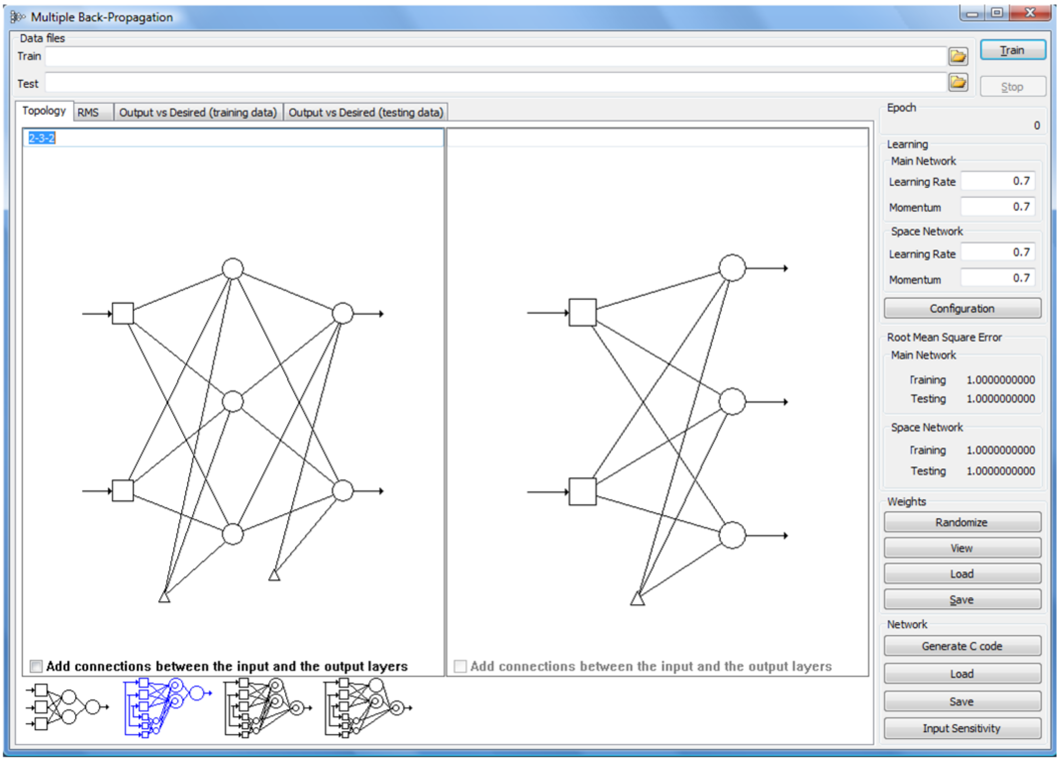

Figure 6.2 : Network Topology Interface

In the designed methodology we have used five input layers, five hidden layers and one output layer in network configuration topology. Initially the water quality parameters weights have been set on very small weights randomly, so as to learn from prepared dataset of 700 samples.

6.2 Weka Tool

In this era of Artificial Intelligence and machine learning there is an imperative need of simple and user-friendly tool to process and visualize the data. We have used aforesaid tool for data mining called Weka. Weka is a GUI based user-friendly tool having a wide range of functioning for data like data processing, data clustering, data visualization and a lot more in the field of data and its science. Unlike other data science tools, Weka is not compatible with python and/or R language. Instead, Weka rely on JAVA programming language. We can also use Weka for Big Data computation as well, because JAVA is also used for Hadoop Framework. Weka also provides SQL database connectivity using JAVA for storing data to the cloud. Weka can be used in three different ways:

- Weka GUI
- Weka CLI
- Weka API

6.2.1 Data analysis using Weka

Weka uses a set of algorithms to perform Data Visualization.

- Linear Machine Learning Algorithms
- Non- Linear Machine Learning Algorithms
- Ensemble Machine Learning Algorithms

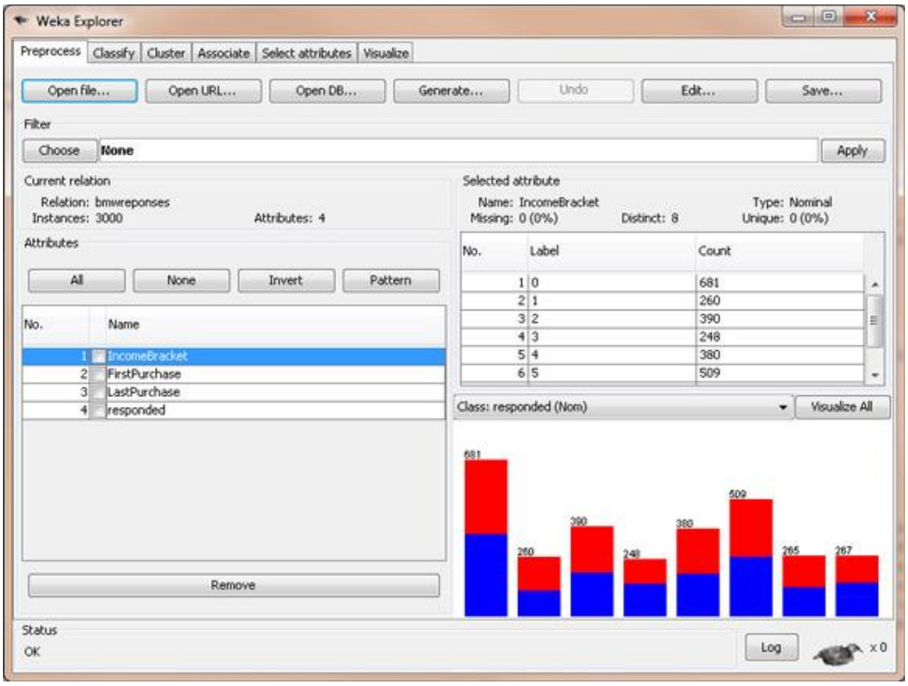

Figure 6.3 : Data Analysis using Weka

6.2.2 Classification of Data using Weka

There are various Data classification algorithm to visualize data, Decision Tree is one of them.

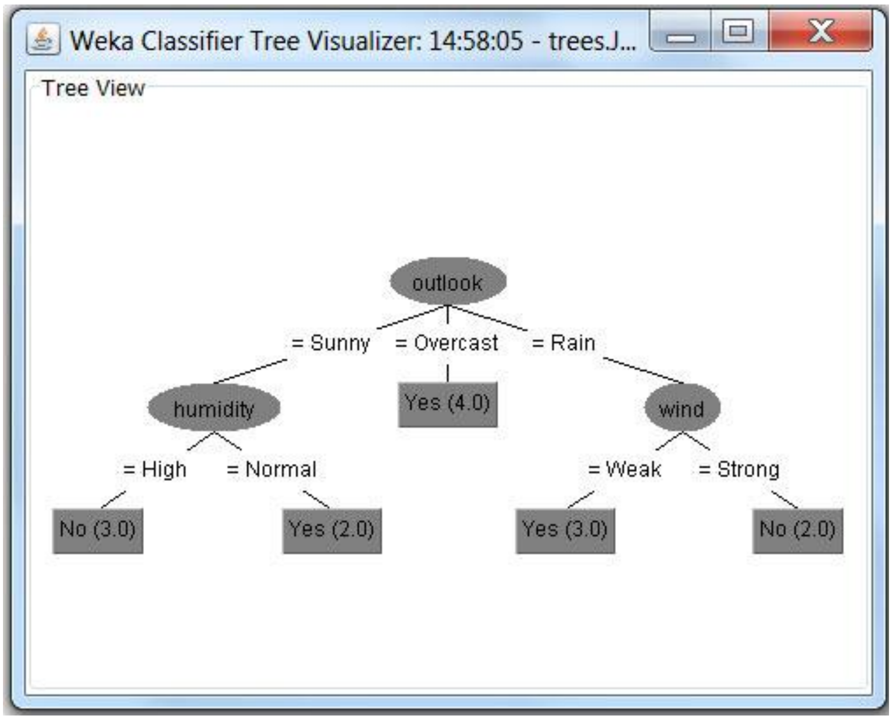

Figure 6.4 : Classification using decision tree

Let's take an example of a data set which stores information about the growth of a flower during different weather conditions. Data set "Flower Data" having data elements sunny, overcast, rain, humidity and wind.

Table 6.1: Example dataset

<table>
<tr><td colspan="4">Sunny</td><td colspan="2">Overcast</td><td colspan="4">Rain</td></tr>
<tr><td colspan="2">Yes</td><td colspan="2">No</td><td>Yes</td><td>No</td><td colspan="2">Yes</td><td colspan="2">No</td></tr>
<tr><td colspan="4">Humidity</td><td colspan="2"></td><td colspan="4">Wind</td></tr>
<tr><td colspan="2">High</td><td colspan="2">Normal</td><td colspan="2"></td><td colspan="2">Weak</td><td colspan="2">Strong</td></tr>
<tr><td>Yes</td><td>No</td><td>Yes</td><td>No</td><td colspan="2"></td><td>Yes</td><td>No</td><td>Yes</td><td>No</td></tr>
</table>

The above Data set for predicting the growth of flower in different climatic conditions is saved in ARFF format which is accepted by Weka for further data visualization.

Weka API is used for supervised and unsupervised learning. It is used for Data Visualization in the form of different charts and graphs for better understanding. Weka API can be used using specific Weka API key in JAVA programming.

6.2.3 Data Depiction

In the designed methodology, after transmitting the results from actual hardware with the help of Wi-Fi module, a mobile app interface is used to demonstrate the data of real time water quality parameters along with water quality class. For this purpose, we have used a freeware application i.e. Thing Speak. It is trending applications for the IoT projects these days. It is an analytical platform amenity that allows the user to envisage, comprehensive and analyse real time data stream using cloud. It provides the three important features for data analysis those are collect, analyse and Act. It provides instant picturing of the data sent from our sensors. These properties of the application make it best suitable for the targeted application in this work. Interface has been prepared for displaying all six parameters used in this work i.e. Temperature, pH, TDS, DO, Turbidity and Water Class using this application and cloud interface. Screenshots for data visualization of various parameters is shown in figure 6.5.

The figure 6.5 represent monitored parameters values received at cloud storage which are sent from actual hardware. Administrators can utilize this data for interpretation and decision making for avoiding any disaster situations related to human health etc. It is utmost important as these can avoid the interference of the middle man from actual reports on water quality and administrators. Because sometimes the action cannot be taken on right time due to lack of operator's interest and knowledge. Also, this continuous monitored data can also be used for future predictions on health hazards etc. It provides the flexibility to retrieve stored data for any analysis in the three given formats as per figure 6.6 below.

Screenshots of data stored on cloud storage is shown in figure 6.7. It provides data with date and time stamp for different parameters.

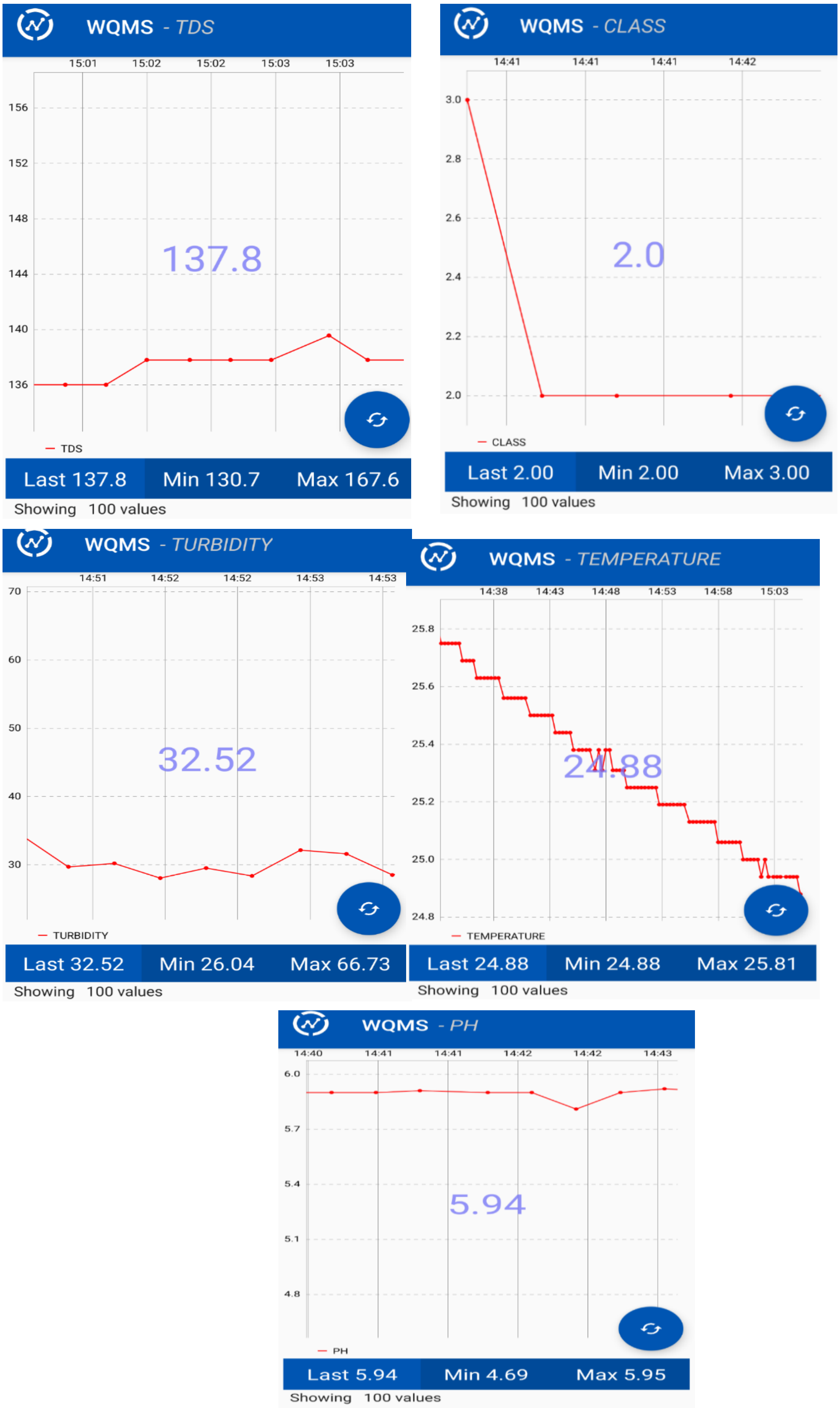

Figure 6.5 : Mobile application screenshot

WQMS Channel Feed:	JSON XML CSV
Field 1 Data: TEMPERATURE	JSON XML CSV
Field 2 Data: PH	JSON XML CSV
Field 3 Data: TDS	JSON XML CSV
Field 4 Data: DO	JSON XML CSV
Field 5 Data: TURBIDITY	JSON XML CSV
Field 6 Data: CLASS	JSON XML CSV

Figure 6.6 : Interface to retrieve stored data on cloud

created_at	entry_	empera	pH	TDS	DO	Turbid	Class
2019-06-27 19:28:34 IST	217	30.38	5.92	28.73	3	0.09	2
2019-06-27 19:28:53 IST	218	30.25	5.81	32.74	3.5	0.08	2
2019-06-27 19:29:11 IST	219	30.06	5.94	34.74	3	0.09	2
2019-07-22 20:09:34 IST	238	29.5	6.13	58.33	3	0.82	1
2019-07-22 20:09:52 IST	239	29.5	6.11	62.19	3.43	0.79	2
2019-07-22 20:10:11 IST	240	29.5	5.93	62.19	4	0.76	2
2019-07-22 20:10:30 IST	241	29.5	6.1	62.19	5	0.79	2
2019-07-22 20:10:49 IST	242	29.44	6.09	64.11	5.3	0.78	2
2019-07-22 20:11:08 IST	243	29.44	6.08	64.11	5	0.81	2
2019-07-22 20:11:26 IST	244	29.44	6.08	64.11	5	0.82	2
2019-07-22 20:11:45 IST	245	29.44	6.07	71.77	5	0.79	2
2019-07-22 20:12:03 IST	246	29.44	6.06	86.86	4.5	0.81	2
2019-07-22 20:12:22 IST	247	29.38	5.89	99.85	5	0.78	2
2019-07-22 20:12:41 IST	248	29.44	6.05	90.59	5	0.79	2
2019-07-22 20:13:00 IST	249	29.44	6.05	75.56	7	0.79	2
2019-07-22 20:13:18 IST	250	29.44	6.05	67.95	7	0.79	2
2019-07-22 20:13:37 IST	251	29.44	6.05	64.11	7.5	0.81	2
2019-07-22 20:13:56 IST	252	29.44	6.05	64.11	7	0.81	2
2019-07-22 20:14:15 IST	253	29.44	6.05	64.11	7	0.79	2
2019-07-22 20:14:34 IST	254	29.44	6.05	66.03	7	0.79	2

Figure 6.7 : Cloud storage screenshot

Chapter 7
Conclusions and Future Work

Water is the essential element for almost every part of living being. As the demand of water increasing, the sources of pollution are also increasing. Although the invention of many schemes and programmes regarding water purification has done a tremendous job, but still there is something that has been lagging. Still in few urban areas or forest areas, people do not receive proper aid for the water. This can be caused either due to individual carelessness or inappropriate resources. We have tried to bridge the gap between problems faced and water quality monitoring.

7.1 Significant Contribution - 1

In this contribution, we have designed a system for reliable water quality estimation. Designed system is an automatic, effective and innovative solution for water quality monitoring in real time. It is capable of measuring the various parameters of water quality using different sensors and communicates the results to the user automatically. The system is simple and reduces human intervention. Different physical and chemical parameters are considered for this work to monitor real time water quality information. Five main parameters including pH, Temperature, TDS, DO and Turbidity are considered to assess water quality. This study reveals that neural network approach is showing significant improvement in tested values over traditional methods. This study focuses on Back Propagation algorithm for training and weight value calculation. Some formulations were given for calculating q value from parameter value automatically. Water Quality Index (WQI) is also calculated during this study. The proposed scheme showed 98.3% accuracy over traditional methods. We have also calculated weight values for each parameter automatically. This is the first time that a methodology is proposed to calculate reliable weight values instead of assigning random weights to the water quality parameters. Additionally, we have developed web interface to analyse historical data for taking future decisions and a mobile app for alert generation if water quality falls beyond limits.

7.2 Significant Contribution - 2

In this contribution, we have tested the first approach using Decision Tree algorithm also. Implemented algorithm is easy to understand due to its structure and division. Prediction is the key feature of a decision tree which analyse the present data and further can be used to predict future data. Different physical and chemical parameters are considered for this work to monitor real time water quality information. Five main parameters including pH, Temperature, TDS, DO and Turbidity are considered to assess water quality. This study reveals that machine learning approach is showing significant improvement in tested values over traditional methods. This approach also proves to be providing similar results as that of Backpropagation algorithm.

7.3 Significant Contribution - 3

We have designed a hardware to verify and validate the complete methodology. This hardware consists of all sensor nodes for detecting individual parameter value. This signal is processed and passed to Arduino where we have implemented proposed algorithms. The data is fed to it and results is displayed on LCD screen. These results consist of parameter values and water quality class type. The information is passed to higher authorities using mobile application. Additionally, the data is also stored in cloud storage system which is further helpful in analysing historical data to take appropriate actions on time.

7.4 Future Work

The future scope in this area is very vast and depending upon received data, actions for quality improvement can be taken automatically. The proposed approach can be validated with a greater number of parameters to find the effect of parameter over water quality. We can apply genetic algorithms also for comparing the results. Furthermore, the complete system can be installed at water treatment plant to find out the reliability, lifetime and effectiveness of hardware. We can also apply other methods of artificial intelligence to further investigate the parameter dependency upon each other as well as water quality. In addition to this, dataset consisting of more than

2000 samples from different sources with all kind of class type can be collected, model can be trained and tested using the proposed approach. This will further improve the results.

List of Publications

1. Bansal, S., & Geetha, G. (2018). A Portable and Low-Cost Multi-sensor for Real Time Remote Sensing of Water Quality in Agriculture. *Pertanika Journal of Science & Technology*, *26*(3). (Scopus)
2. Bansal, S., & Geetha, G. (2019). Advanced Evaluation Methodology for Water Quality Assessment Using Artificial Neural Network Approach. *Water Resources Management*, 1-15. (SCI, IF:2.987)
3. Bansal, S., & Geetha, G. (2019). A Machine Learning Approach Towards Automatic Water Quality Monitoring. *Water Chemistry and Technology* (Accepted). (SCI, IF:0.504)
4. Patent entitled "A Novel Method for Water Quality Parameter Evaluation using Machine Learning" filed with full specification (Application number: 201811024276).

Bibliography

[1] P. Debels, R. Figueroa, R. Urrutia, R. Barra, and X. Niell, “Evaluation of water quality in the Chillán River (Central Chile) using physicochemical parameters and a modified Water Quality Index,” *Environ. Monit. Assess.*, vol. 110, no. 1–3, pp. 301–322, 2005.

[2] H. G. Gorchev and G. Ozolins, “Guidelines for Drinking-water Quality,” *Who*, vol. 1, p. 564, 2011.

[3] P. L. Smedley and D. G. Kinniburgh, “United Nations Synthesis Report on Arsenic in Drinking Water Developed on behalf of the United Nations Administrative Committee on Cooperation Sub-Committee on Water Resources, with active participation of UNICEF, UNIDO, IAEA and the World Bank,” *Br. Geol. Surv. Wallingford*, pp. 1–390, 2001.

[4] Z. Wu, X. Wang, Y. Chen, Y. Cai, and J. Deng, “Assessing river water quality using water quality index in Lake Taihu Basin, China,” *Sci. Total Environ.*, vol. 612, pp. 914–922, 2018.

[5] BIS, “BIS - Drinking Water Specification, 2012,” 2012.

[6] European Union, “European Union Drinking water regulations,” no. 122, pp. 1–34, 2014.

[7] N. Gupta, P. Pandey, and J. Hussain, “Effect of physicochemical and biological parameters on the quality of river water of Narmada, Madhya Pradesh, India,” *Water Sci.*, vol. 31, no. 1, pp. 11–23, 2017.

[8] M. T. Ladan, “Review of NESREA act 2007 and regulations 2009-2011: a new Dawn in environmental compliance and enforcement in Nigeria.” Law Env’t & Dev. J., 2012.

[9] D. Spry and T. Branch, “An Overview of Canadian Water Quality Guidelines,” *USEPA Expert Meet. Washingt. DC*, 2015.

[10] S. Yamamura, “Drinking water guidelines and standards”, In United Nations Synthesis Report on Arsenic in Drinking Water,” p. 18, 2001.

[11] V. Venkatraman, “STUDY ON THE SAFETY AND SECURITY,” *Int. J. Manag.*, vol. 6, no. 10, pp. 69–82, 2015.

[12] S. Barnard *et al.*, “Impact of Indian Total Sanitation Campaign on Latrine

Coverage and Use: A Cross-Sectional Study in Orissa Three Years following Programme Implementation," *PLoS One*, vol. 8, no. 8, 2013.

[13] N. Naeraa *et al.*, "pH and molecular CO2 components of the Bohr effect in human blood.," *Scand. J. Clin. Lab. Invest.*, vol. 18, no. 1, pp. 96–102, 1966.

[14] C. A. Caldwell and J. Hinshaw, "Physiological and haematological responses in rainbow trout subjected to supplemental dissolved oxygen in fish culture," *Aquaculture*, vol. 126, no. 1–2, pp. 183–193, 1994.

[15] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, "Machine Learning: An Artificial Intelligence Approach," *J. Math. Psychol.*, vol. 305, pp. 299–305, 1987.

[16] S. Simon, "Accuracy Driven Artificial Neural Networks in Stock Market Prediction," *Int. J. Soft Comput.*, vol. 3, no. 2, pp. 35–44, 2012.

[17] S. H. Dinius, "Design of an Index of Water Quality," *JAWRA J. Am. Water Resour. Assoc.*, vol. 23, no. 5, pp. 833–843, 1987.

[18] R. S. Michalski J. G. Carbonell T. M. Mitchell, *Machine Learning An Artificial Intelligence Approach*, no. 2. 1988.

[19] P. Langley and H. A. Simon, "Applications of machine learning and rule induction," *ACM Commun.*, vol. 38, no. 11, pp. 54–64, 1995.

[20] P. Langley, "Applications of Machine learning and rule induction." pp. 1–17, 1995.

[21] H. Chen, "Machine learning for information retrieval: neural networks, symbolic learning and genetic algorithms," *J. Am. Soc. Inf. Sci.*, vol. 46, no. 3, pp. 124–216, 1995.

[22] O. Postolache, P. Girao, M. Pereira, and H. Ramos, "An Internet and microcontroller-based remote operation multi-sensor system for water quality monitoring," *Proc. IEEE Sensors*, pp. 1532–1536, 2002.

[23] E. G. Breijo, L. G. Sanchez, J. I. Civera, A. T. Ferrando, and G. P. Boluda, "Thick-film multisensor for determining water quality parameters," *IECON Proc. (Industrial Electron. Conf.*, vol. 4, no. December, pp. 2791–2796, 2002.

[24] A. Manciu, M. Popa, D. Mitrea, and D. Căpăţînă, "Parameters monitoring solutions for the quality control of water used in healthcare units," in *2006 IEEE International Conference on Automation, Quality and Testing, Robotics,*

AQTR, 2006, vol. 2, pp. 0–5.

[25] H. M. B O'Flynn, Rafael Martínez- Català, S. Harte, C. O'Mathuna, John Cleary, C. Slater, F. Regan, D. Diamond, "SmartCoast A Wireless Sensor Network for Water Quality Monitoring," *Proc. - Conf. Local Comput. Networks, LCN*, pp. 919–926, 2007.

[26] P. Jiang, H. Xia, Z. He, and Z. Wang, "Design of a water environment monitoring system based on wireless sensor networks," *Sensors*, vol. 9, no. 8, pp. 6411–6434, 2009.

[27] F. Regan *et al.*, "A demonstration of wireless sensing for long term monitoring of water quality," *2009 IEEE 34th Conf. Local Comput. Networks*, no. October, pp. 819–825, 2009.

[28] T. Z. Qiao and L. Song, "The design of multi-parameter online monitoring system of water quality based on GPRS," *2010 Int. Conf. Multimed. Technol. ICMT 2010*, pp. 3–5, 2010.

[29] W. Y. Chung and J. H. Yoo, "Remote water quality monitoring in wide area," *Sensors Actuators, B Chem.*, vol. 217, pp. 51–57, 2015.

[30] G. Liu *et al.*, "Assessment of surface water quality of the Songhuajiang River basin, China," *2011 Int. Conf. Remote Sensing, Environ. Transp. Eng. RSETE 2011*, pp. 5755–5758, 2011.

[31] M. A. Nasirudin, U. N. Za'bah, and O. Sidek, "Fresh water real-time monitoring system based on Wireless Sensor Network and GSM," *Open Syst. (ICOS), 2011 IEEE Conf.*, pp. 354–357, 2011.

[32] M. Deqing, Z. Ying, and C. Shangsong, "Automatic Measurement and Reporting System of Water Quality Based on GSM," *Intell. Syst. Des. Eng. Appl. (ISDEA), 2012 Second Int. Conf.*, no. 5, pp. 1007–1010, 2012.

[33] J. C. Chou, C. C. Chen, M. W. Su, C. C. Chen, and S. Y. Yang, "Cl - and H + Sensing Devices for Water Quality Monitoring System," *2012 IEEE Int. Symp. Circuits Syst.*, pp. 2043–2046, 2012.

[34] S. K. Vaddadi, "Development of Embedded Wireless Network and Water Quality Measurement Systems for Aquaculture," *Sixth Int. Conf. Sens. Technol. Dev.*, pp. 637–641, 2012.

[35] C. Damian, C. Fosalau, J. M. Dias Pereira, O. Postolache, and P. S. Girao,

"Sensor network for water quality assesment," *EPE 2012 - Proc. 2012 Int. Conf. Expo. Electr. Power Eng.*, no. Epe, pp. 849–852, 2012.

[36] M. A. Miskam, I. A. Rahim, O. Sidek, M. Q. Omar, and M. Z. Ishak, "Deployment of wireless water-quality monitoring system at titi serong paddy crop field, Malaysia," *Proc. - 2013 IEEE 3rd Int. Conf. Syst. Eng. Technol. ICSET 2013*, pp. 57–60, 2013.

[37] UNEP, *The Emissions Gap Report 2012*. 2012.

[38] WHO, *Guidelines for drinking-water quality: first addendum to the fourth edition*. 2017.

[39] M. J. B. Alam, M. R. Islam, Z. Muyen, M. Mamun, and S. Islam, "Water quality parameters along rivers," *J. Environtal Sci. Technol.*, vol. 4, no. 1, pp. 159–167, 2007.

[40] K. Nosrati, "Identification of a water quality indicator for urban roof runoff," *Sustain. Water Qual. Ecol.*, vol. 9–10, no. July, pp. 78–87, 2017.

[41] S. H. Ewaid, S. A. Abed, and S. A. Kadhum, "Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis," *Environ. Technol. Innov.*, vol. 11, pp. 390–398, 2018.

[42] S. Naidoo and A. O. Olaniran, "Treated wastewater effluent as a source of microbial pollution of surface water resources," *Int. J. Environ. Res. Public Health*, vol. 11, no. 1, pp. 249–270, 2013.

[43] S. H. Dinius, "Design of an Index of Water Quality," *JAWRA J. Am. Water Resour. Assoc.*, vol. 23, no. 5, pp. 833–843, 1987.

[44] R. Das Kangabam, S. D. Bhoominathan, S. Kanagaraj, and M. Govindaraju, "Development of a water quality index (WQI) for the Loktak Lake in India," *Appl. Water Sci.*, vol. 7, no. 6, pp. 2907–2918, 2017.

[45] G. Srivastava and P. Kumar, "Water quality index with missing parameters," *Int. J. Res. Eng. Technol.*, vol. 2, no. 4, pp. 609–614, 2013.

[46] A. Said, D. K. Stevens, and G. Sehlke, "An innovative index for evaluating water quality in streams," *Environ. Manage.*, vol. 34, no. 3, pp. 406–414, 2004.

[47] M. S. B. Maind, "Research Paper on Basic of Artificial Neural Network," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 2, no. 1, pp. 96–100, 2014.

[48] Z. Minli and Q. Shanshan, "Research on the Application of Artificial Neural

Networks in Tender Offer for Construction Projects," *Phys. Procedia*, vol. 24, pp. 1781–1788, 2012.

[49] Asce, "ARTIFICIAL NEURAL NETWORKS IN HYDROLOGY," *J. Hydrol. Eng.*, no. April, pp. 124–137, 2000.

[50] N. Benvenuto and F. Piazza, "The backpropagation algorithm," *IEEE Trans. Signal Process*, vol. 40, no. 4, pp. 967–969, 1992.

[51] D. Chapman, "Water Quality Assessments - A guide to use of biota, sediments and water in enviromental monitoring," *Univ. Press. Cambridge ISBN*, vol. 2, no. 1, p. 609, 1996.

[52] M. Fielding and F. M., "Urban Water Book reviews," *Prog. Hum. Geogr.*, vol. 24, no. 1, pp. 154–156, 2000.

[53] C. G. Cude, "OREGON WATER QUALITY INDEX A TOOL FOR EVALUATING WATER QUALITY MANAGEMENT EFFECTIVENESS â€™ quality issues by the public and policy makers . tors and other water resource policy makers via sible on the internet (http://www.deq.state.or.us/ annual McKenzie ," *J. Am. Water Resour. Assoc.*, vol. 37, no. 1, pp. 125–137, 2001.

[54] O. S. Thirunavukkarasu, T. Viraraghavan, K. S. Subramanian, and S. Tanjore, "Organic arsenic removal from drinking water," *Urban Water*, vol. 4, no. 4, pp. 415–421, 2002.

[55] S. K. Sundaray, U. C. Panda, B. B. Nayak, and D. Bhatta, "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of the Mahanadi river-estuarine system (India) - A case study," *Environ. Geochem. Health*, vol. 28, no. 4, pp. 317–330, 2006.

[56] Unep, *The Emissions Gap Report 2012, A United Nations Environment Programme (UNEP) Synthesis Report*. 2012.

[57] B. Y. J. Hall, A. D. Zaffiro, and J. G. Herrmann, "On-line water quality parameters as indicators of distribution system contamination," *AWWA*, no. January, pp. 66–77, 2007.

[58] J. Skadsen *et al.*, "Distribution system on-line monitoring for detecting contamination and water quality changes," *J. / Am. Water Work. Assoc.*, vol. 100, no. 7, pp. 81–94, 2008.

[59] L. W. Y. Yeung *et al.*, “A survey of perfluorinated compounds in surface water and biota including dolphins from the Ganges River and in other waterbodies in India,” *Chemosphere*, vol. 76, no. 1, pp. 55–62, 2009.

[60] S. B. Brandt, M. Gerken, K. J. Hartman, and E. Demers, “Effects of hypoxia on food consumption and growth of juvenile striped bass (Morone saxatilis),” *J. Exp. Mar. Bio. Ecol.*, vol. 381, no. SUPPL., 2009.

[61] X. L. X. Lianqinq, W. C. W. Chunling, G. F. G. Fuming, and W. Y. W. Yifeng, “Water Quality Health and Hazard Risk Assessment on Drinking Water Sources,” *2009 3rd Int. Conf. Bioinforma. Biomed. Eng.*, no. 50609006, pp. 4–7, 2009.

[62] M. de Franc and ¸Doria, “Factors influencing public perception of drinking water quality,” *Water Policy*, vol. 12, no. SUPPL. 1, pp. 1–7, 2010.

[63] M. Fengyun, “Progress in Water Quality Monitoring Based on Remote Sensing and GIS,” *Challenges Environ. Sci. Comput. Eng. (CESCE), 2010 Int. Conf.*, vol. 2, no. 66, pp. 209–212, 2010.

[64] M. V. Storey, B. van der Gaag, and B. P. Burns, “Advances in on-line drinking water quality monitoring and early warning systems,” *Water Res.*, vol. 45, no. 2, pp. 741–747, 2011.

[65] C. Bach, X. Dauchy, M. C. Chagnon, and S. Etienne, “Chemical compounds and toxicological assessments of drinking water stored in polyethylene terephthalate (PET) bottles: A source of controversy reviewed,” *Water Res.*, vol. 46, no. 3, pp. 571–583, 2012.

[66] R. M. BROWN, N. I. McCLELLAND, R. A. DEININGER, and M. F. O’CONNOR, “a Water Quality Index – Crashing the Psychological Barrier,” *Adv. Water Pollut. Res.*, pp. 787–797, 2013.

[67] P. Silva Girão, O. Postolache, and J. D. Pereira, “Wireless sensor network-based solution for environmental monitoring: water quality assessment case study,” *IET Sci. Meas. Technol.*, vol. 8, no. 6, pp. 610–616, 2014.

[68] H. Oyem, O. M., and D. Ezeweali, “Temperature, pH, Electrical conductivity, TDS and Chemical Oxygen demand of groundwater in Boji-BojiAgbor/Owa Area and Immediate Suburbs,” *Res. J. Environ. Sci.*, vol. 8, no. 8, pp. 444–450, 2014.

[69] A. Sarkar and P. Pandey, "River Water Quality Modelling Using Artificial Neural Network Technique," *Aquat. Procedia*, vol. 4, no. Icwrcoe, pp. 1070–1077, 2015.

[70] N. A. Cloete, R. Malekian, and L. Nair, "Design of Smart Sensors for Real-Time Water Quality Monitoring," *IEEE Access*, vol. 4, pp. 3975–3990, 2016.

[71] C. Y. Hsu, S. J. Ou, and W. F. Hsieh, "Predicting Fish Ecological As Indicator of River Pollution Using Decision Tree Technique," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 164, no. 1, 2018.

[72] G. B. Ruben, K. Zhang, H. Bao, and X. Ma, "Application and Sensitivity Analysis of Artificial Neural Network for Prediction of Chemical Oxygen Demand," *Water Resour. Manag.*, vol. 32, no. 1, pp. 273–283, 2018.

[73] S. Ponsadailakshmi, S. G. Sankari, S. M. Prasanna, and G. Madhurambal, "Evaluation of water quality suitability for drinking using drinking water quality index in Nagapattinam district, Tamil Nadu in Southern India," *Groundw. Sustain. Dev.*, vol. 6, pp. 43–49, 2018.

[74] N. Adimalla, P. Li, and S. Venkatayogi, "Hydrogeochemical Evaluation of Groundwater Quality for Drinking and Irrigation Purposes and Integrated Interpretation with Water Quality Index Studies," *Environ. Process.*, vol. 5, no. 2, pp. 363–383, 2018.

[75] S. B. Chandanapalli, E. S. Reddy, and D. R. Lakshmi, "DFTDT: distributed functional tangent decision tree for aqua status prediction in wireless sensor networks," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 9, pp. 1419–1434, 2018.

[76] M. Geetha Jenifel and R. Jemila Rose, "Recursive partitioning algorithm in water quality prediction," *Int. J. Environ. Sci. Technol.*, no. 123456789, 2019.

[77] J. A. Cotruvo, *2017 WHO Guidelines for Drinking Water Quality: First Addendum to the Fourth Edition*, vol. 109. 2017.

[78] S. H. Ewaid, S. A. Abed, and S. A. Kadhum, "Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis," *Environ. Technol. Innov.*, vol. 11, pp. 390–398, 2018.

[79] M. T. H. de Frahan, S. Yellapantula, R. King, M. S. Day, and R. W. Grout, "Deep learning for presumed probability density function models," 2019.

[80] Q. He, Z. Dong, F. Zhuang, T. Shang, and Z. Shi, "Parallel decision tree with

application to water quality data analysis.," *Adv. Neural Netw.*, vol. 7951, no. July, 2013.

[81] A. V. Rubonis and L. Bickman, "Psychological impairment in the wake of disaster: The disaster-psychopathology relationship," *Psychol. Bull.*, vol. 109, no. 3, pp. 384–399, 1991.