Design and Development of Drinking Water Quality Monitoring System Based on Multi-Sensor Array

THESIS

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

PUNIT KHATRI

(2017PHXF0009P)

Under the Supervision of

PROF. KARUNESH KUMAR GUPTA

And Under the Co-Supervision of

PROF. RAJ KUMAR GUPTA

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2022

Dedication

To my parents & wife for their unconditional support

&

The never give up hope during the journey of Ph.D.

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI – 333031 (RAJASTHAN), INDIA

CERTIFICATE

This is to certify that the thesis entitled **"Design and Development of Drinking Water Quality Monitoring System Based on Multi-Sensor Array"** submitted by **Punit Khatri,** ID No. **2017PHXF0009P,** for the award of Ph.D. of the institute embodies original work done by him under our supervision.

Signature of the Supervisor	Signature of Co-Supervisor
Name: Prof. Karunesh Kumar Gupta	Name: Prof. Raj Kumar Gupta
Designation: Associate Professor	Designation: Professor
Date:	Date:

Acknowledgements

"The Power of God is with you all the time; through the activities as mind, senses, breathing, and emotions; and is constantly doing all the work using you as a mere instrument."

----Shri Mad Bhagavad Gita

First, I would like to thank Almighty for the support of whom I have reached this stage of my Ph.D. I would like to thank the following people for their support and affection during the entire duration of my Ph.D., professionally or personally, without whom I would not have been able to complete my research. I may have forgotten to acknowledge all the people, but such mistakes may be ignored.

I would like to take this opportunity to convey my gratitude and appreciation to my supervisor Prof. Karunesh Kumar Gupta, Associate Professor, Department of Electrical and Electronics Engineering, BITS-Pilani, Pilani campus, for his interest, support, guidance, and encouragement during my research tenure. I am greatly honored to be his student for supporting me either personally or professionally. I wish to express my deepest gratitude to my co-supervisor Prof. Raj Kumar Gupta, Professor, Department of Physics, BITS-Pilani, Pilani campus, for his guidance, suggestions, moral support throughout the research period.

I express my appreciation to the people of BITS-Pilani for providing all the necessary facilities to conduct the research. My special thanks to Prof. Souvik Bhattacharyya, Vice-Chancellor of the university, Prof. S. K. Barai, Director, BITS-Pilani, Pilani campus, for allowing me to pursue my research work successfully. I would like to thank Prof. M. B. Srinivas, Dean, and Prof. Jitendra Panwar, Associate Dean, Academic–Graduate Studies & Research Division, for valuable support throughout the program. I would like to express my gratitude to Prof. V. Manjula Devi (Professor), Department of Physics and Prof. Pawan K. Ajmera (Assistant Professor), Department of Electrical and Electronics Engineering, Members of Doctoral Advisory Committee (DAC), for their motivation and moral support technically as well as personally. I would like to thank Prof. Hitesh Dutt Mathur, Head of Department, Electrical and Electronics Engineering, for his motivation during the entire period of Ph.D. I also would like to thank Ex-HODs (Prof. Navneet Gupta and Prof. V. K. Chaubey) for their support and motivation.

I am also thankful to all my professors for their support and motivation during my Ph.D. at BITS-Pilani, Pilani campus. I am also thankful to all the laboratory and staff members, especially Mr. V. N. Jangid and Mr. Rajkumar, for their support. I would like to thank all my friends Dr. Uttam Goswami, Mr. Anil Kumar Savita, Mr. Devendra Prajapati, Mr. Sandeep Kaydan, Mr. Srinath K, Mr. Kailash Nayak, Mr. Surendra Singh Patel, and Mr. Amit Singh. I also thank Mr. Navjot Kumar, Dr. Amit Sharma, Dr. Hemant Shankar Dutta, Mr. Dharmendra Pal, Mr. Shivanand Pal, Mr. Deepak Panwar, Mr. Ramakant Sharma, Mr. Krishna Veer Singh, Mr. Rahul Sharma and all other whose name is not here and have helped me during my Ph.D. Thanks to all faculty members of BITS-Pilani, Pilani Campus, for helping me out at various times.

I would like to thank the Department of Science and Technology (DST), Govt. of India and the Council of Scientific and Industrial Research-Human Resource Development Group (CSIR-HRDG) for providing financial support as a fellowship for serving as Junior Research Fellow (JRF) and Senior Research Fellow (SRF).

I am also thankful to Dr. P. C. Panchariya, Director, CSIR-Central Electronics Engineering Research Institute (CSIR-CEERI), Pilani, for his motivation, encouragement, and unconditional support, professionally and personally. Finally, I would like to thank my parents-Shri Mohan Lal Khatri & Smt. Manju Khatri for their sacrifice to make me for what I am today, my wife-Ms. Deepika Khatri and my sister-Ms. Madhuri Khatri, for their care, affection, and support during my research journey. I was also blessed with my little princess Poorvaa Khatri during my Ph.D. journey, which was an unforgettable moment for my wife and me.

Punit Khatri

Abstract

Water is one of the natural resources for living, which is being used carelessly by human beings. The population growth has increased water consumption, but the resources are limited. Among all forms of the water available on earth, only 3% is freshwater, out of which glaciers and the icecaps cover 69.7%, 30% is groundwater, and 0.3% is surface water. With swift growth in urbanization, the water consumption has been increased and the contamination has also been increased in proportion due to industrialization, domestic and industrial waste discharge into water resources. Hence, water quality monitoring is always a matter of concern before consumption. The contaminated water causes severe diseases, such as cholera, diarrhea, and dysentery. Traditional water quality monitoring involves the sample collection and subsequent laboratory testing of the samples, resulting in high labor costs and time consumption. Also, the measurement is not in a real-time environment and involves analytical instruments in experimentation. So, there is a need for real-time water quality monitoring. The development of a Multi-Sensor System (MSS) for the same will be an excellent solution, enabling accurate real-time and online water quality monitoring with minimum use of chemicals.

Different regions have different geological conditions, so the parameters responsible for water quality will differ. Hence, selecting water quality parameters will be a critical step in the development, as overall water quality will be decided based on the selected parameters. The sensor selection and system development are also essential. So, the proposed work deals with selecting water quality parameters and designing and developing a sustainable water quality monitoring system for real-time as well as online measurement. The water quality analysis based on statistical modeling and soft computing techniques has been carried out in this thesis work. The proposed water quality analysis techniques are simple, accurate, and easily implementable on the hardware platform.

This thesis work is the continuation of project granted by the Department of Science and Technology (DST), Govt. of India, New Delhi (Grant No. DST/TM/WTI/2K16/103). The work discussed above was the requirement of the

funding agency, such as the selection of water quality parameters and the design & development of a water quality monitoring system. The following work has been extended as a further part of the thesis work.

Although sensor technology has achieved the manufacturing of low-cost and portable water quality sensors, the calibration is also a concern as the sensor faces the drift problem sooner or later after embedding in the system. This sensor drift will demolish the calibration model of any instrument. The sensor drift compensation has also been carried out in this thesis work employing a soft computing technique. Conventional water distribution systems always face leakage, failure, delay in maintenance, which results in high wastage of water. The online water quality monitoring in the water distribution network is also presented in this thesis work.

Table of Contents

List of Figures

List of Tables

List of Selected Abbreviations

ANN	Artificial Neural Networks
BIS	Bureau of Indian Standards
BOD	Biochemical Oxygen Demand
CCMEWQI	Canadian Council of Ministers of the Environment Water Quality Index
CPCB	Central Pollution Control Board
CPS	Cyber-Physical Systems
DO	Dissolved Oxygen
E. Coli.	Escherichia coli
EC	Electrical Conductivity
FF-ANN	Feed Forward-Artificial Neural Network
GA	Genetic Algorithm
GUI	Graphical User Interface
HMI	Human Machine Interface
ICT	Information and Communication Technologies
IDE	Integrated Development Environment
IoT	Internet of Things
IWT	Indian Water Tool
MF	Membership Function
ML	Machine Learning
MLR	Multiple Linear Regression
MSC	Multiplicative Scattering Correction
NSFWQI	National Sanitation Foundation Water Quality Index
NTU	Nephelometric Turbidity Unit
ORP	Oxidation Reduction Potential
OWQI	Oregon Water Quality Index
PCA	Principal Component Analysis
PCR	Principal Component Regression
pH	Potential of Hydrogen
PLSR	Partial Least Squares Regression

RMSE	Root Mean Square Error
RMSECV	Root Mean Square Error for Cross Validation
SCC	Signal Conditioning Circuits
SCT	Sensing and Communication Technologies
SMPS	Switch Mode Power Supply
TDS	Total Dissolved Solids
TSS	Total Suspended Solids
USEPA	US Environment Protection Agency
WAWQI	Weighted Arithmetic Water Quality Index
WRI	Water Resource Institute
WBCSD	World Business Council for Sustainable Development
WHO	World Health Organization
WSN	Wireless Sensor Network
WQI	Water Quality Index
WQR	Water Quality Rating

Chapter 1

Introduction

Water is essential for all known forms of life and covers 70% of the earth's surface. Out of which, 96.5% is found in oceans and seas, 3.4% in ice caps and glaciers, and 0.0001% in clouds and vapors. Among all forms of the water available on earth, only 3% is freshwater, out of which glaciers and the icecaps cover 69.7%, 30% is groundwater, and 0.3% is surface water. The fresh surface water is distributed in lakes, swamps, and rivers in the ratio of 87:11:2. The distribution of available water is shown in Figure 1.1 [1], [2].

Water is an essential resource of survival for a human being, whether for drinking purposes, industrial purposes, irrigation, or any other application. Globally, we utilize 70% of our water assets for agriculture and irrigation purpose and 10% for residential uses [3]. It is consumed carelessly by a human being day by day. The population is increasing rapidly, and thus, the consumption of water is increasing in proportion, but the resources (either natural or artificial) are limited. The freshwater is regularly decreasing and may create a worse situation in the future. In developing countries, eighty percent of sicknesses are connected to poor water quality and sanitation conditions [4]. India is one of the most water-challenged developing countries in the world. In India, 5% of death in an individual is due to diarrhea, which is among the top ten death causes in India. Death statistics in India are shown in Figure 1.2.

Water quality is defined by various physical, chemical, and biological parameters. Monitoring of these parameters will be helpful in deciding the suitability of water usage for different applications (domestic, industrial, or irrigation). The chemical parameters are mostly invoked for characterizing the water quality, such as pH, dissolved oxygen (DO), nitrogen, organic and inorganic compounds, nitrate, fluoride, and toxicants. Below are some parameters responsible for the overall water quality indicator.

- Physical parameters → temperature, turbidity, odor, color, salinity, dissolved solids, and suspended solids.
- Chemical parameters → pH, dissolved oxygen (DO), nitrogen, organic and inorganic compounds, nitrate, fluoride, and toxicants.
- Biological parameters → algae and bacteria

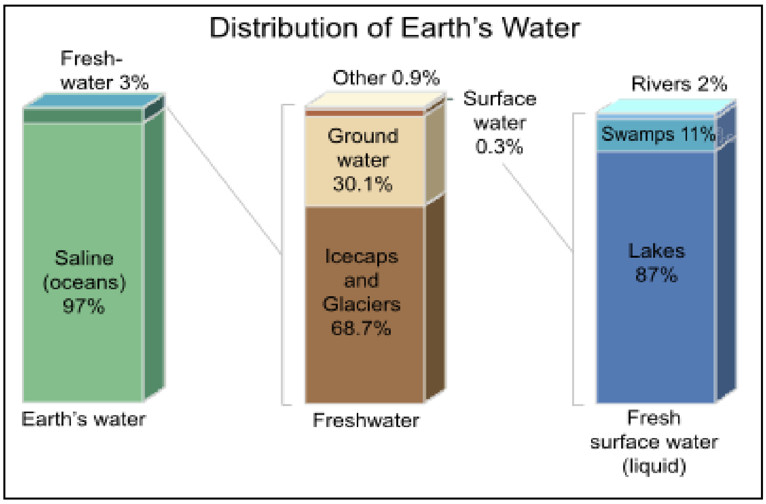


Figure 1.1 Distribution of earth's water systems

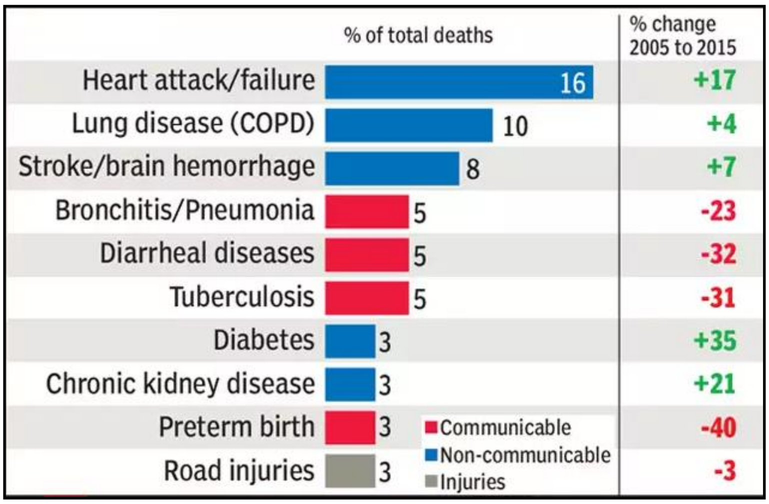


Figure 1.2 Top ten causes of death in India
(Source: World Health Organization-India: WHO statistical profile)

Drinking water quality monitoring is essential before consumption in daily life. It affects human health directly or indirectly [5]. Most of the available water is contaminated and can cause severe diseases like cholera, dysentery, diarrhea, and skin problems. The leading cause of contamination or pollution is the discharge of domestic and industrial wastes into natural water resources. According to recent reports from the Delhi government officials, the water quality fails 19 water quality standards [6]. India water tool (IWT-2.1) is a publicly available web platform created by a group of companies, research organizations, and industry associations, including Water Resources Institute (WRI), and coordinated by the World Business Council for Sustainable Development (WBCSD) [7]. The IWT-2.1 estimates water quality as per the Indian government standard called the Bureau of Indian Standards (BIS). When the contamination exceeds prescribed BIS limits, drinking water is considered unsafe. The groundwater quality was measured by ITW 2.1 tool in 632 districts in India. The contamination was found more than prescribed BIS limits in 59 out of 632 districts. The yellow and red zones in Figure 1.3 indicate places where chlorine, fluoride, iron, arsenic, nitrate, and electrical conductivity surpass national standards. Safe drinking water will be a dream for humans in the future. All the reasons mentioned earlier lead us to the monitoring of water quality before consumption. We have confined our research to groundwater quality as the study area has groundwater as a primary resource of consumption.

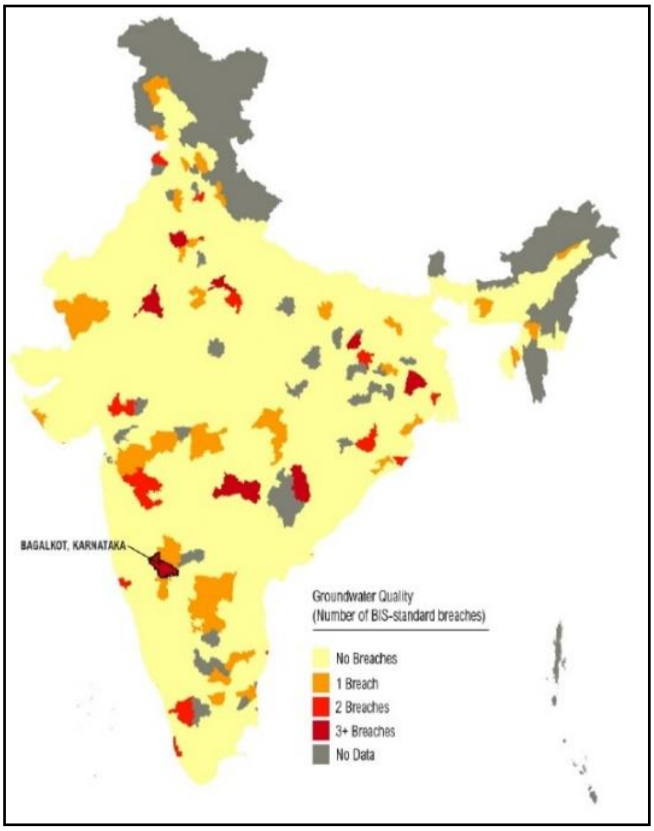


Figure 1.3 Groundwater quality (number of BIS standard breaches)

1.1.Research Motivation

There is always a motivation behind conducting any research work. Some of the facts/issues which motivated us to carry out the research work are as follows:

- Approximately 0.2 million people lack access to safe drinking water [8]. According to a survey from the World Health Organization (WHO), by the end of 2025, half of the world population will be living in water stresses areas.
- The freshwater resources are depleting, resulting in groundwater abstraction for domestic, industrial, and agricultural uses.
- The groundwater is contaminated by agrochemicals (pesticides and fertilizers). As a result, most of the population does not have access to safe drinking water and depends on contaminated water for their daily needs.
- Water quality monitoring before consumption can reduce the risk of illness in an individual.

These facts mentioned above motivated us to carry out our research work in water quality monitoring. When there were no techniques available for real-time water quality parameter measurement, the traditional methods were used in which samples were collected on-site and sent to quality testing labs for measurement. This process is time-consuming and results in delays in test reports along with labor and cost consumption. A real-time water quality monitoring system is much needed to reduce the time and effort involved in traditional measurement techniques. So, our research work will focus on the design and development of a drinking water quality monitoring system that can further help us determine whether or not the government regulations are being complied with. The real-time water quality monitoring will also help reduce the risk of illness in an individual before consumption.

1.2. Organization of Thesis

Chapter 1 discusses the background, introduction to the research topic, motivation of research work, and the proposed objectives to be carried out in this thesis.

Chapter 2 discusses an extensive literature survey covering the existing water quality monitoring methods (statistical modeling and soft-computing techniques), drift compensation methods, and water quality monitoring in distribution networks.

Chapter 3 presents the proposed methodology, including water parameter selection criteria, prototype development based on commercial off-the-shelf (COTS) modules, software framework development, and experimental procedure.

Chapter 4 presents the water quality analysis based on statistical modeling and soft computing technique.

Chapter 5 discusses the problem encountered pertaining to drift in commercial water quality sensors and the compensation for the same using Artificial Neural Network (ANN).

Chapter 6 presents the proposed water quality monitoring in water distribution systems and implementing a centralized network employing the Internet of Things (IoT).

Chapter 7 discusses the conclusion of the thesis work, specific contributions, and possible future recommendations of the research work.

Chapter 2

Literature Review

Preamble

There are various methods for water quality parameter measurement, including traditional laboratory-based and In-situ techniques. The traditional techniques involve on-site sample collection followed by laboratory testing, which is time and cost-consuming. In comparison, in-situ measurements are fast, accurate, time-saving and cost-saving as well. The overall water quality is defined by a single term or a numeric value, obtained by different techniques employing statistical and soft computing methods. This chapter describes various techniques for water quality parameter measurement (traditional and in-situ). This chapter also provides an in-depth review of work done by researchers and scientists for water quality analysis using the statistical method and soft computing techniques. The drift analysis and compensation techniques have also been discussed in the chapter. Finally, the water quality monitoring in distribution networks has been presented in this chapter. The flow of the chapter is shown in Figure 2.1.

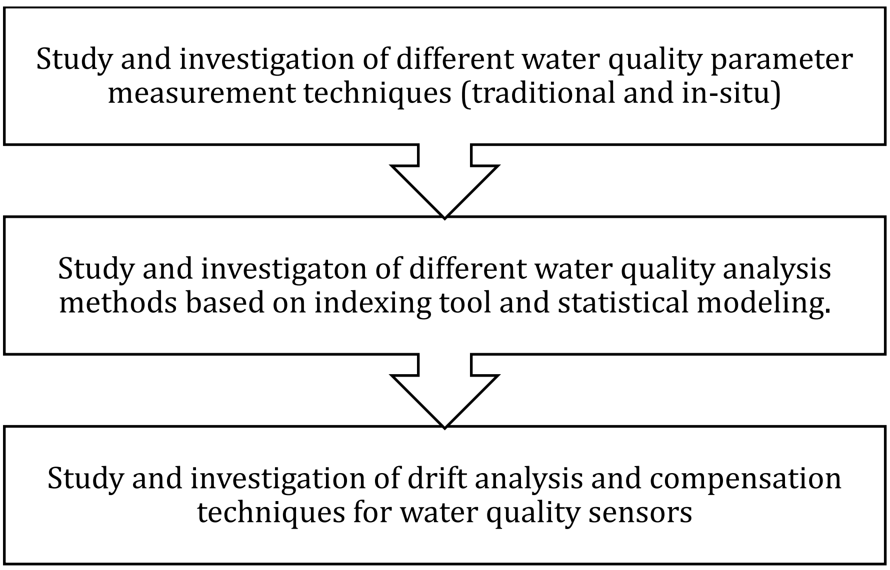


Figure 2.1 Literature review-chapter flow

2.1.Introduction

Water quality deterioration is a global concern as urbanization and industrialization has affected human life and the aquatic ecosystem. The water quality varies by natural as well as human influences. The natural influences may be geological or climatic on which human has no control. In comparison, human influences have worsened the water quality and can still be controlled [9]. Some of the common reasons across the globe for water contamination are as follows [10], [11].

- Direct drainage of human feces into water resources pollutes the water. Such a problem occurs when there is no facility for waste disposal or on-site sanitation facility.
- Almost 75% of the industrial waste and sewage water is disposed into the water resources without any treatment, which pollutes the usable water.
- Contamination is also caused by naturally occurring minerals and chemicals, which is uncontrollable.
- Urbanization and agricultural run-off are also responsible for water pollution.

Water quality is defined by various physical, biological, chemical, and aesthetic (smell and visual) parameters. Additionally, the water quality depends on the geological condition, and according to that, different governing bodies have decided the criteria for defining the water quality. The various governing bodies are US Environment Protection Agency (USEPA) [12], National Water Quality Standard (NWQS), Malaysia [13], Central Pollution and Control Board (CPCB), India [14], etc. In India, CPCB has defined the criteria for water quality for different applications, such as drinking, bathing, irrigation, industrial, and fishing, as shown in Table 2.1.

Water is a good solvent that picks up the impurities easily [15]. Pure water is tasteless, colorless, and odorless, is often called the universal solvent. Dissolved solids refer to any minerals, salts, metals, cations, or anions dissolved in the water. Total Dissolved Solids (TDS) consist of inorganic salts (primarily calcium, magnesium, potassium, sodium, bicarbonates, chlorides, and sulfates) and a small amount of organic matter dissolved in water. TDS in natural water originates from natural sources, sewage, urban run-off, industrial wastewater, and chemicals used in the water treatment process.

When the TDS level exceeds 1000 mg/l, it is considered unfit for human consumption. A high level of TDS should be considered a high priority and should also be examined before drinking. The best water purification system ensures that the filters effectively remove unwanted particles from water [16].

Table 2.1 Central Pollution and Control Board (CPCB) criteria for water quality

Type of Water	Category	Quality Parameter Criteria
Drinking-Water Source without conventional treatment but after disinfection	**'A'**	1. Total Coliforms Organism MPN/100ml shall be 50 or less 2. pH between 6.5 and 8.5 3. Dissolved Oxygen 6mg/l or more 4. Biochemical Oxygen Demand 5 days 20°C 2mg/l or less
Outdoor bathing (Organized)	**'B'**	1. Total Coliforms Organism MPN/100ml shall be 500 or less 2. pH between 6.5 and 8.5 3. Dissolved Oxygen 5mg/l or more 4. Biochemical Oxygen Demand 5 days 20°C 3mg/l or less
Drinking water source after conventional treatment and disinfection	**'C'**	1. Total Coliforms Organism MPN/100ml shall be 5000 or less 2. pH between 6 to 9 3. Dissolved Oxygen 4mg/l or more 4. Biochemical Oxygen Demand 5 days 20°C 3mg/l or less 5. TDS 1000 mg/l
Propagation of Wildlife and Fisheries	**'D'**	1. pH between 6.5 to 8.5 2. Dissolved Oxygen 4mg/l or more 3. Free Ammonia (as N) 1.2 mg/l or less
Irrigation, Industrial Cooling, Controlled Waste disposal	**'E'**	1. pH between 6.0 to 8.5 2. Electrical conductivity at 25°C micromhos/cm Max.2250 3. Sodium absorption Ratio Max. 26 4. Boron Max. 2mg/l

Dissolved Oxygen (DO) is the amount of dissolved gaseous oxygen (O_2) in water. It is essential for the assessment of water quality. Oxygen enters the water by direct absorption from the atmosphere, rapid movement, or as a waste product of plant photosynthesis. Too high or too low levels of dissolved oxygen can harm aquatic life and affect the water quality. Water temperature, moving water, turbulence, and salinity can affect dissolved oxygen levels [17]. Adequate dissolved oxygen is essential for good water quality and necessary for all forms of life. The range for the DO concentration is 5 to 14.5 mg O_2 per liter. A rapid fall in DO indicates higher organic pollution in the river. The optimal value for good water quality is 4 to 6 mg/l of DO, ensuring healthy aquatic life in water. Dissolved oxygen levels below 5.0 mg/l may cause stress to aquatic life.

Turbidity measures any liquid's cloudiness caused by particles. Turbidity can be caused by silt, sand, mud, chemical residues, bacteria, and other germs. It is essential to measure the turbidity in water supplies because it can cause filter blocking and stop the treatment plant's effective working. In high turbidity conditions, the mud and silt can cause blocking of water supply pipes and damage the taps and valves. The low turbidity can prevent chlorine from killing the bacteria in the water [18].

A pH is not the primary determinant of adverse effect because pH in the stomach has a range between 1.0 and 3.5 with a mean of approximately 2.0, and there are certain foods like vinegar, lemon juice that have pH less than 3.0. So, there is no direct relationship between the health and pH of drinking water, and it is impossible to ascertain [19]. However, pH can indirectly affect health because pH can affect the degree of corrosion of pipes, and there will be ingestion of heavy metals in the body from plumbing and pipes.

2.2. Methods for Water Quality Parameter Measurement

There are different techniques available for the water quality parameter measurement, such as lab-based chemical measurement technique, electrochemical sensor-based measurement, optical sensor-based measurement technique, etc. These techniques are discussed in this section.

2.2.1. *Conventional Approach for Water Quality Parameter Measurement*

In the early days, there were no techniques available for real-time water quality parameter measurement; the only solution was to collect the sample and subsequently testing it in the laboratory. The water quality parameters; turbidity, TDS, pH, chlorine, DO, and BOD are commonly measured to define water quality.

Turbidity measurement was based on the Jackson method of candle, which is the oldest method of measurement. In this method, a glass platform is used. A cylindrical glass container is placed above the platform, and a lighted candle is placed beneath. Then the water is poured into the cylindrical container until the candle ceases to be seen from above. Finally, the water level is measured, and turbidity is analyzed. The turbidity is inversely proportional to the height of the water column. The unit of the turbidity is Jackson Turbidity Unit (JTU). This method is slow and consumes more time in measurement [20]. The TDS measurement was based on the gravimetric method. In this method, the water is boiled at 103°C, and the evaporated particles are trapped in the mineral matrix. The dried evaporated particles are weighted on analytical balance having an accuracy of 0.0001 gm. This method is slow but accurate. The only disadvantage is that low boiling chemicals may also evaporate with the water. The chlorine measurement was based on the membrane electrode-based colorimetric method. The method is time-consuming, laborious and its output is affected due to the presence of other heavy metals in water.

The pH measurement technique was based on titration with sodium hydroxide and litmus paper. The limitation of this method is less accurate. DO measurement was based on the Winkler method and polarographic Clark electrode measurement. This method consumes oxygen in the process. The BOD measurement technique was based on the oxidation process by hydrogen peroxide with ultraviolet light. These methods are time-consuming and have a limited range [21]. A summary of the measurement method has been presented in Table 2.2.

All these conventional or traditional methods discussed are chemical based, use of toxic reagents, high cost of measurement and operation. Also, the disposal of toxic reagents produced in the measurement process was a problem [22]. These problems lead the researchers and scientists to move toward the real-time water quality parameter measurement approach described in the next section.

Table 2.2 Traditional methods for water quality parameter measurement

Water Quality Parameter	**Traditional lab-based method**	**Limitation**
pH	Titration with sodium hydroxide and litmus paper	Less Accurate
Dissolved Oxygen (DO)	Winkler method and polarographic Clark electrode measurement	Consumes Oxygen
Turbidity	Nephelometric Method	Less Sensitive
Chlorine	Membrane electrode-based colorimetric method	Time-Consuming and other heavy metals affect the output
Biochemical Oxygen Demand (BOD)	Oxidation process by hydrogen peroxide with ultraviolet light	Limited Measurement range

2.2.2. *Real-Time Approach for Water Quality Parameter Measurement*

The real-time analysis allows the user to acquire accurate water quality information without delay. Different sensor technologies are available for real-time parameter measurement based on electrochemical sensors, ISFET based sensors, MEMs based sensors, and Optical sensors (refer to Figure 2.2).

Electrochemical sensor systems provide quick, precise, selective, sensitive, easy-to-use analytical tools for examining environmental samples. Since the electrochemical analysis requires a relatively small amount of the sample, the sensor systems are effective and excellent for detecting and monitoring contaminants. Heavy metals, pesticides, herbicides, dye compounds, and medicinal chemicals may be detected using electrochemical sensor systems based on amperometry and voltammetry, which are inexpensive, quick, sensitive, reliable, and user-friendly [23]. The ISFET based sensors are used for pH measurement in a real-time environment but always face drift and hysteresis problems. The ISFET based sensor strongly requires a drift compensation [24], [25].

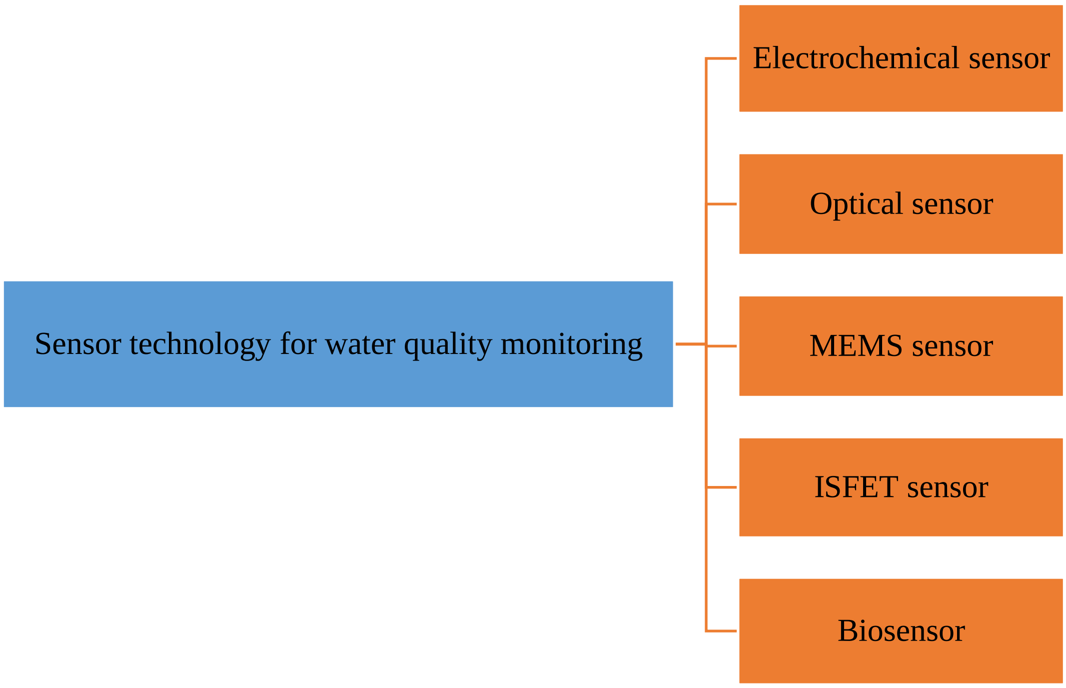


Figure 2.2 Sensor technology for water quality monitoring

Micro-Electro-Mechanical Systems (MEMS) are a collection of micro-and nano-sensors that can perceive their surroundings and react to changes in that environment using a microcircuit. It is based on nanotechnology. Due to the massive scale production of MEMS-based sensors, there was a substantial cost reduction. Furthermore, low power consumption eliminates repetitive calibration [26], [27]. However, this technique is still under development and will require many changes to be suitable for detecting temperature, pH, and heavy metals [28]. The optical sensors measure the visible properties of the water that respond with the change of light intensity and have high accuracy [29]. The optical sensors can measure turbidity, DO, pH, and chlorine in the water.

The biosensor converts a biological signal into measurable quantities. It consists of a transducer and an interface. The microbial cells or tissue slices or cell membrane or enzymes can be used as biological indicators. These indicators interact with a specific parameter. The signal (potential or current) generated by the transducer is proportionate to the object and signifies a physical or chemical reaction change. The biosensors are portable and easy to use. On the other hand, bio-sensors are expensive non-linear system, and their application is problematic due to population expansion, biological indicator genetic evolution, and a complicated fabrication procedure [30]–[33]. The real-time water quality parameter measurement approach is far better and competitive with the traditional method described earlier.

2.3. Methods for Water Quality Analysis

Water quality analysis is necessary to ensure the safety of the water supply and to prevent health risks associated with contamination. There are different methods for water quality analysis based on indexing tool, statistical modeling, and soft-computing techniques. In the previous section, traditional and in-situ water quality parameter measurement techniques were studied. In this section, different methods for water quality analysis have been discussed in detail.

2.3.1. Indexing Tool for Water Quality Analysis

Various methods are available for water quality analysis in literature. Water Quality Index (WQI) tool is one of them. This tool defines water quality by a unique rating to determine water quality based on a single value. A considerable number of water quality indices have been formulated around the world. Several national and international organizations have formulated the Weighted Arithmetic Water Quality Index (WAWQI), National Sanitation Foundation Water Quality Index (NSFWQI), Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI), Oregon Water Quality Index (OWQI). Most of these indices are based on the WQI developed by the US National Sanitation Foundation (NSF) in 1970, and it is commonly used in the world, including the Indian continent [34].

Horton [35] proposed the WQI term for the first time and introduced the aggregate function to calculate the WQI. It was based on ten water quality parameters, including pH, conductance, alkalinity, chloride, and Dissolved Oxygen (DO). The aggregate function multiplied the arithmetic weight of individual parameters with temperature. Brown et al. [36] applied a similar method but without multiplying the weights. The other researchers used the concept with slight modification in terms of different aggregate functions such as arithmetic aggregate mean, multiplicative aggregate function, geometric mean, harmonic mean, and minimum operator to calculate WQI. Harkins et al. [37] developed a new method by varying the number of water quality variables. The standardized distance for each observation vector and rank of variance was calculated for individual water quality parameters. The proposed model was more sensitive but inconvenient to use.

McClelland [38] developed a new geometric mean approach to calculate the water quality index. The earlier developed arithmetic mean method for WQI was insensitive to low-value parameters, a trait later dubbed "eclipsing". Walski & Parker [39] used the geometric mean method to calculate the quality of individual water quality parameters and expressed the quality in the range of 0 to 1. The proposed function was called the "sensitivity function". Various curves and formulas were suggested to determine sensitivity functions from the measured water quality parameters.

Dinius [40] introduced a water quality index employing multiplicative aggregation on a decreasing scale. The indexing values were expressed as the percentage of water quality correspondence to 100%. The relationship between each pollutant and the water quality was obtained using the proposed method. Bhargava [41] developed a simplified water quality indexing model for the classification of water quality of river Ganga for its different uses, such as bathing, swimming, public water supply, irrigation, industry, fish culture, wildlife, and boating. The proposed model was the modification of the model proposed by Harkins et al. [37] and Walski & Parker [39].

Smith et al. [42], [43] developed a new water quality indexing method based on the minimum operator for different water uses. It was based on two indexing methods: water quality standards and expert opinion. The water quality parameter selection, subindices calculation, and weight assignments were calculated using the Delphi method [44].

Dojildo et al. [45] suggested and developed the harmonic mean method to define water quality indexing. The calculated mean does not use the weights for individual water quality parameters. The method is more sensitive as compared to geometric or arithmetic mean. The harmonic mean-based methods have been adopted by many countries, such as the British Columbia Water Quality Index (BCWQI) and the Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI).

Dwivedi et al. [46] used an integrated water quality index to assess the changes in physiological and chemical properties of river Ganga over the years 1985-86, 1986-87, and 1987-88. The pollutant load of the Ganga River was also assessed over a 13-kilometer stretch in India's Varanasi district. Jonnalagadda and Mhere [47] characterized the nature, source, and extent of pollution. Various water quality parameters, namely, pH, temperature, conductance, total and dissolved solids and nitrate, were considered for the water quality assessment of the Odzi River. Bordalo et al. [48] studied the effect of different agroindustry effluents in the Salado River in Buenos Aires Province (Argentina). The indexing was done using a minimum water quality index (WQImin), measuring two water quality parameters, electrical conductivity and dissolved oxygen. The efficacy of WQImin was also studied for river management.

Avvannavar and Srihari [49] applied the Bhargava WQI method for water quality monitoring in the Netravathi river, Mangalore, India. Different water quality parameters were used to calculate WQI, such as Most Probable Number (MPN), DO, TDS, turbidity, and pH. Chaturvedi and Bassin [50] have assessed water quality in different borewells and a treatment plant in the Delhi province of India. The samples were collected over nine months, and NSFWQI was used to calculate the water quality index. Total five water quality parameters, namely, temperature, pH, nitrate, turbidity, and TDS, were used for indexing. Singh and Kamal [51] proposed water quality assessment in different talukas of Goa (India), employing the weighted arithmetic mean method. Around thirty-two samples were collected from various places during pre- and post-monsoon season, and indexing was calculated. The WQI ranges from 34 to 107.

Bhutiani et al. [52] proposed the WQI calculation in the Ganga River Ecosystem at Haridwar, Uttarakhand, India. The WQI was calculated by analyzing 16 physio-chemical parameters over a period of eleven years. The weighted arithmetic mean method was applied to the acquired data for WQI calculation. Shah et al. [53] used the weighted arithmetic mean method to calculate the WQI in the Sabarmati river located in the Gujarat province of India. Shah used six variables, pH, DO, Biochemical Oxygen Demand (BOD), nitrate, nitrogen, total coliform (TC), and electrical conductivity (EC), to calculate WQI. Chaurasia et al. [54] calculated the WQI of the groundwater samples collected from the southern part of the Varanasi district of Uttar Pradesh Province (India). Chaurasia selected eight important water quality parameters out of twenty-four for the WQI calculation employing the weighted arithmetic mean method. Ameen [55] proposed the drinking water quality assessment of samples collected from ten villages during wet and dry seasons in Barwari Bala, Duhok, Kurdistan Region, Iraq. Different physio-chemical water quality parameters were used for indexing employing the weighted arithmetic mean method. Tyagi et al. [56] studied and compared various water quality index methods adopted in different countries. As there is no globally accepted water quality index method, Tyagi has studied different indexing methods used in various continents.

2.3.2. *Statistical Modeling for Water Quality Analysis*

Regression is a powerful tool for statistical modeling, which is widely used in weather forecasting and prediction. The least-squares approach was the first type of regression, which was developed by Legendre in 1805 [57] and Gauss in 1809 [58] for astronomical measurements. They determine the orbits of bodies around the Sun. In 1821, Gauss published a continuation of the theory of least squares, which included a variant of the Gauss–Markov theorem. There are mainly two types of regression: linear regression and nonlinear regression. Linear regression analysis can further be divided into univariate regression and multivariate regression analysis. Nonlinear regression analysis may be further categorized based on function such as gaussian function, logarithmic function, exponential function, power function, and trigonometric function. Various methods are available in the literature for water quality monitoring, characterization, and parameter prediction based on different regression algorithms. Some of the relevant papers are studied in this section.

Table 2.3 Summary of indexing tool for water quality monitoring in Indian continent

Reference	WQI Calculation Method	Application	Parameters used for Modeling
Dwivedi et al. [46]	Multiple aggregate functions were used	River water quality	pH, TDS, TC, nitrate, BOD, DO and more
Jonnalagadda and Mhere [47]	Weighted arithmetic mean	-do-	pH, temperature, conductance, total and dissolved solids and nitrate
Bordalo et al. [48]	Minimum Operator	-do-	pH, COD, TDS, N-KTN, DO, and PT
Avvannavar and Srihari [49]	Harmonic mean method	-do-	Turbidity, DO, MPN, TDS, pH and BOD
Chaturvedi and Bassin [50]	NSFWQI method	Borewell and treatment plant	Temperature, pH, nitrate, turbidity and TDS
Singh and Kamal [51]	Weighted arithmetic mean	Talukas water quality	pH, DO, TDS, TH, BOD, TSS, Nitrate, Sulphate Magnesium, Calcium and Chloride
Bhutiani et al. [52]	-do-	River water quality	EC, Temp, DO, COD, BOD, pH, turbidity, TDS, TS, free CO2, chlorides, alkalinity, phosphates, nitrates and velocity
Shah et al. [53]	-do-	-do-	pH, DO, EC, TC, nitrate, BOD and DO
Chaurasia et al. [54]	-do-	Groundwater quality	22 water quality parameters including pH, EC, TDS and TH

Christensen et al. [59] analyzed the bacteria and nutrient concentrations based on the least-squares regression analysis in the Kansa streams, USA. Selected water quality parameters were used for the modeling, such as pH, DO, turbidity, temperature and chlorophyll. The predicted parameters were Chemical Oxygen Demand (COD), filtered COD, Total Suspended Solids (TSS) and nitrate. The PLSR was used to develop the calibration model to achieve robustness and high correlation quality. Singh et al. [60] used a multiway PLSR (unfold PLS, tri-PLS and N-PLS) for river water quality analysis. A ten-year dataset was used to develop the proposed model for predicting biological oxygen demand (BOD) in river water.

Chenini and Khemiri [61] characterized the ground water quality using hydro-chemical data by Multiple Linear Regression (MLR) and Structural Equation Modeling (SEM) in Maknassy Basin, central Tunisia. Twenty-eight samples were collected from October 2005 to November 2005 in the study region, and the proposed methodology was applied. Koklu et al. [62] used the Principal Component Analysis and Factor Analysis (PCA-FA) and MLR for water quality monitoring in the Melen River, Turkey. The data from five different monitoring stations for 20 years during the period 1995-2006, containing 26 different physical and chemical parameters. The dependency of a water quality parameter with other parameters was also studied in detail.

Shareef et al. [63] proposed a new method to determine the contaminants and water quality parameters. A Gray Level Co-occurrence Matrix (GLCM) was used to predict six water quality parameters (WQP) employing a multi-regression model followed by the fuzzy k-means clustering algorithm. Two different strategies were used in the measurement: one used different fusion levels, and the second was to generate slope-derived spectra by calculating the slope of absorbance.

Shrestha and Basnet [64] used the linear regression analysis for water quality monitoring in Ratuwa River and its tributaries in Nepal. The water quality parameters were obtained using the standard procedure, i.e., ALPHA methods. The correlation of conductivity with other parameters, such as TDS, DO, fluoride, magnesium, total alkalinity (TA), total phosphorous (TP), calcium (Ca) and sodium (Na) was also found out. Dutta and Sarma [65] evaluated different groundwater samples (boring or tube wells) and studied the potability of the water samples in the Nagaon district of Assam, India. Twelve water quality parameters (physical, chemical and biological) were used for the regression modeling, and correlation was found between different physio-chemical water quality parameters. Ahamad et al. [66] developed a regression model for water quality as well as pollution prediction inside two lakes situated in Tezpur university. The modeling was done employing regression analysis and artificial neural network (ANN) and results were compared, and the best suitable technique was suggested.

Table 2.4 Summary of statistical modeling for water quality monitoring

Reference	Regression Method Used	Application	Parameters used for Modeling
Christensen et al. [59]	Least Squares Regression	Stream water	pH, DO, turbidity, temperature and chlorophyll
Singh et al. [60]	PLSR	River water	19 water quality parameters
Chenini and Khemiri [61]	MLR and structural modeling	Groundwater	pH, TDS, EC, NA^{+}, Cl, K^{+} and Mg^{2+}
Koklu et al. [62]	PCA-FA	River water	26 water quality parameters
Shareef et al. [63]	Multi-regression model coupled with fuzzy k-means clustering	-do-	pH, nitrate and phosphate
Shrestha and Basnet [64]	Linear regression	River water	EC, TDS, DO, fluoride, magnesium, TA, TP, Ca and Na
Dutta and Sarma [65]	-do-	Groundwater	Twelve water quality parameters (physical, chemical and biological)
Ahamad et al. [66]	-do-	Lake water	TS, EC and Turbidity

2.4. Drift Analysis of Water Quality Sensors

In recent years, there has been a growing interest in sustainable development for water quality monitoring. The Multi-Sensor Systems (MSS) or E-tongue, which is used for water quality monitoring, mimics the human taste perception [67], [68]. The MSS consists of an array of water quality sensors based on electrochemical, or membrane-based, or optical technology, which depend on application and availability. There are many techniques for water quality monitoring available in the literature. These techniques are real-time monitoring as well as contamination detection in distribution networks [69]–[73] and wireless monitoring [74]–[77]. A typical step in these techniques includes calibration of the sensors before the measurement. The MSS is trained during the calibration with the high-cost known standard concentration solution. For calibration, a considerable amount of labor and time is required. To reduce the calibration frequency, some mathematical models are developed with the help of acquired post-calibration data. Though calibration samples are expensive, the developed model is expected to have a longer lifetime.

This calibration process has been done in a controlled environment chamber or a laboratory environment with a high-cost instrument. But this is not feasible after the field deployment of the instrument, and later, the sensors need to be calibrated regularly based on the field environmental conditions. In such circumstances, poor or incomplete calibration leads to uncertainty in sensor measurement. The limitations of the calibration lifetime are typically associated with the evolution of the sensor material in the measurement process, which leads to the change in the sensor properties, causing drift. Drift is a continuous change in the measuring instrument reading over time due to the change in the metrological properties [78]. It is neither related to the quantity being measured nor to any recognized influence quantity [79]. Various effects like aging, temperature deviation, surface chemical reactions, adsorption of sample components, contamination and/or poisoning of the working electrode (change in catalytic activity) and the reference electrode may lead to sensor deviation, especially in the case of electrochemical sensors [80]–[82]. The electrochemical sensors are the prevalent type of sensors in MSS so far. It is not possible to prevent sensor drift after field deployment, especially when the sensor lifetime can be as long as years [68], [83].

There are some ways to evade such mishappenings: 1) calibrate the sensors before every measurement and 2) by some mathematical drift correction, considering the evolution of sensor readings. The first method is unfavorable since it is not possible to go into the field to calibrate the sensors. In contrast, the second method seems to be more approachable since it reduces frequent calibration of the sensors. The mathematical drift correction can be implemented using statistical methods or machine learning techniques, as shown in Figure 2.3.

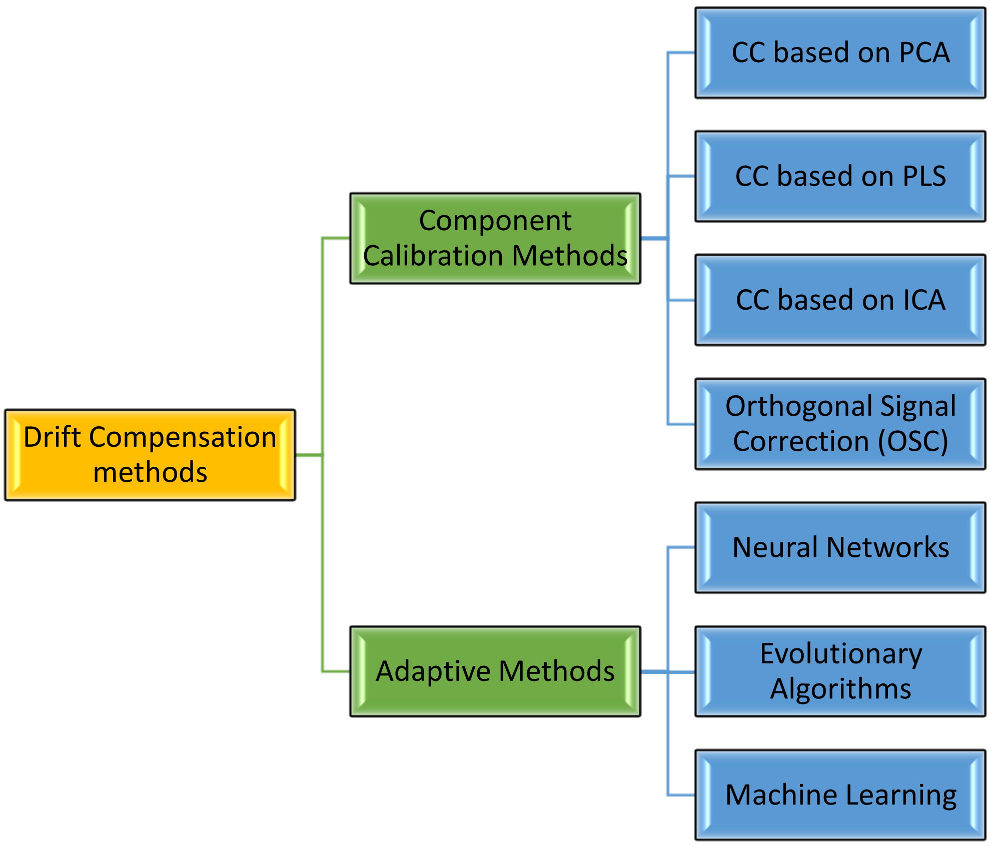


Figure 2.3 Drift compensation methods

Machine learning is becoming part of system design as it has an adaptive nature, which makes it robust for measurement applications. The learning techniques can be supervised, unsupervised, or semi-supervised. In this work, we proposed a method using Artificial Neural Network (ANN), which is a subclass of ML techniques to increase the calibration lifetime of commercial water quality sensors. ANN is inspired by a biological neural network. ANN is a collection of interconnected artificial nodes in which every node can send a signal to another node [84]. It solves the problem in the same way as the human brain does and thus being using in many complex real-life computing applications.

Many researchers have attempted drift compensation either for a sensor or an instrument using traditional statistical analysis and machine learning techniques. Traditional statistical methods for drift correction include multiplicative drift correction (MDC), orthogonal signal correction (OSC), component correction (CC) based on principal component analysis (PCA), and partial least squares regression (PLSR), independent component analysis (ICA). These traditional methods offer drift prediction and compensation efficiently in case of linear variation in sensor drift. While having nonlinearity and outliers in the dataset, the traditional methods might have incorrect assumptions. To overcome this problem, researchers have tried alternative approaches that can handle nonlinear data as well.

Wold et al. [85] proposed the signal correction of Near Infrared (NIR) reflectance spectroscopy data employing the OSC (a variant of PLSR) without preprocessing the data. Two different datasets were taken from the spectra of modified cellular glucose and the pulp samples. The proposed method showed significant improvements in results as compared to traditional methods, which employ preprocessing techniques. Li et al. [86] studied the PLSR technique correcting the calibration models. The calibration model was developed with the spectra of the primary instrument, and the secondary instrument was calibrated with the developed model, assuming that the spectral difference and the prediction error have a linear relationship. Ziyatdinov et al. [87] used the Common Principal Component Analysis (CPCA) for gas sensor drift compensation. The algorithm's performance is assessed using a classification task, focusing on determining the variance component of drift, which is a fundamental aspect of the approaches. The CPCA technique is unique in that it expresses drift direction as a variance that is common to all odor classes, eliminating the need for a single reference gas. Laref et al. [88] proposed a method for correcting the E-nose signal employing baseline manipulation and OSC. The data was obtained from the gas sensor array with different concentrations of gas vapors. The PLSR model was then used to predict the unknown gas concentration. The proposed method shows the good stability of the results obtained. Yi [89] proposed a novel dimensionality reduction technique to encounter the intelligent system's sensor drift issue. Also, an optimization algorithm is developed to solve the problem encountered in dimensionality reduction. To show the effectiveness of the proposed method, extensive experiments have been performed.

Out of these above-mentioned statistical methods for E-tongue drift correction, Laref got the best results for drift compensation. Similarly, E-tongue drift correction has been attempted by Holmberg et al. [90], Holmin et al. [91], Luo et al. [92], Zhang et al. [93]. More papers for different drift compensation techniques employing statistical methods can be found in [94]–[96].

In the past decade, machine learning techniques have become popular as an alternative to conventional statistical methods. Current applications of machine learning are human activity recognition systems [97], IoT cultural data [98], boredom classification [99], load balancing [100], enhancing the accuracy of data-driven models [101], classification problems [102] and drift correction as well. Pereira et al. [103] proposed a temperature drift correction in a signal conditioning circuit employing ANN. A temperature sensing system containing an AD595 conditioning circuit was considered for the proposed technique. The measurement system was analyzed for pre and post-temperature correction. The system shows significant accuracy after the post drift compensation method.

Kashwan and Bhuyan [104] developed a robust E-Nose system for determining the flavor and aroma of spices and tea employing the drift correction method. The temperature and humidity variation were compensated. The change in temperature and humidity were constantly monitored to determine the net deviation. The E-nose response is quickly and automatically corrected for the net deviation determined from drift calculations by programming. Uthra et al. [105] applied the ANN for drift correction in an induction motor drive control system. The feedback signals fed to the control system face the problem of drift, which were used as input to the proposed ANN model before feeding the control system. The results show good accuracy in controlling the induction motor. Adhikari and Saha [106] combined an ANN and K-Nearest Neighbor (KNN) model for drift compensation in gas sensors. The data was obtained from the UCI machine learning repository. The PCA method was used for classification after applying the compensation methods. The PCA shows significant accuracy post compensation. Sinha et al. [107] applied the drift compensation technique for Ion-Sensitive Field-Effect Transistor (ISFET) based pH sensor employing different machine learning techniques. The drift occurs in the pH sensor due to temperature variation was compensated employing Multi-Layer Perceptron (MLP), Random Forest (RF), Decision Trees (DT), Support Vector Machine (SVM) and Linear Regression (LR).

Among the above-mentioned machine learning techniques, Sinha got the best results for the drift correction using machine learning techniques. So, it can be stated that for drift compensation, the machine learning techniques are better than the statistical ones. Therefore, we can explore the application of ANN for the drift compensation of water quality sensors.

Table 2.5 Summary of drift compensation methods

Reference	**Compensation Method**	**Application**
Wold et al. [102]	OSC	Cellular glucose and pulp samples
Li et al. [103]	PLSR	Calibrating instrument
Ziyatdinov et al. [104]	CPCA	Gas sensor drift compensation
Laref et al. [105]	OSC	-do-
Yi [106]	PCA	Instrument drift correction

Holmberg et al. [107], Holmin et al. [108], Luo et al. [109], Zhang et al. [110]	--	E-tongue
Pereira et al. [120]	ANN	Temperature drift correction
Kashwan and Bhuyan [121]	Deviation subtraction	Temperature and humidity correction
Uthra et al. [122]	ANN	Motor drive control
Adhikari and Saha [123]	ANN and KNN	Gas sensor drift compensation
Sinha et al. [124]	MLP, RF, DT, SVM and LR	ISFET pH sensor drift correction

2.5.Gaps in the Existing Research

Based on the extensive literature survey presented above, the research gaps identified in water quality monitoring are below.

- Real-time monitoring of water quality is a challenge as it is rarely found. There is a need for the development of real-time water quality monitoring with minimum or no sample preparation.
- There are many benchmark instruments available for water quality monitoring. All these sensors used in the instruments have one common problem that is drift. Only a few authors have attempted the drift analysis and compensation of water quality sensors. There are two solutions for the same: one is the sensor's timely calibration, and the second is the auto-calibration of the sensors through some soft computing methods. The first solution is time-consuming and cost-effective as the reference solutions for the calibrations are expensive. The second approach is more approachable using compensation methods.
- There is a scope of water quality monitoring in conventional water distribution networks as it always faces leakage, failure and illegal connection.

2.6. Objectives of the Proposed Work

The focus of the presented work in the thesis has been to devise and develop a real-time water quality monitoring system. The main objectives of the work, based on the research gaps are stated below.

- Study and investigate different methods for water quality analysis techniques based on statistical modeling and soft computing techniques.
- Design and develop a real-time water quality monitoring system based on Multi-Sensor Array (MSA).
- Water quality analysis based on statistical modeling and soft computing techniques.
- Drift analysis and compensation of commercial water quality sensors by soft computing method.
- Centralized water quality monitoring in water distribution networks employing the Internet of Things (IoT).

Chapter 3

Methodology and Experimentation

Preamble

Water quality always plays an essential role in human health as it is one of the most prominent survival resources. The overall water quality depends on available water resources in the specific geological region; hence, it is required to identify the region-specific quality parameters before developing a hardware framework. In addition to that, selecting a core controller and related modules and peripherals is essential for hardware development. A detailed description explaining the methodology for adapting the specific water quality parameters, selection of core controller, associated modules, hardware, software, and experimental methodologies has been presented in this chapter.

3.1. Water Quality Parameter Selection

The selection of water quality parameters for water quality monitoring must be made very carefully as it will be used to determine the overall water quality. Different criteria for different locations decide the water quality parameters as the significant parameters responsible for water quality will vary with the geological conditions. These criteria are determined by water availability and quality surveys conducted by agencies. In India, the water quality criteria have been defined by Central Pollution and Control Board (CPCB), India [14].

The CPCB has categorized the water quality according to their uses and consumption that has already been discussed in chapter 2. The type ‘C’ category has been selected for this study. The parameters that fall in this category are pH, Dissolved Oxygen (DO), and Total Dissolved Solids (TDS), Biochemical Oxygen Demand (BOD) and Total Coliform. The study area is the Shekhawati region of Rajasthan, India, where groundwater is the only source of consumption [108]. We have not added coliform and BOD in our work, as mentioned in the category ‘C’ of CPCB water quality standards. No literature has been identified reporting E. Coli. in Rajasthan province, as the study area has dry weather conditions where the chance of growth of E. Coli. is very less. E. Coli. is present only where the storage container is not appropriately cleaned or old distribution pipeline or the pipeline leakage or bad sanitation condition [109], [110]. Two additional parameters, Oxidation Reduction Potential (ORP) and temperature have been added to the measurement. So, the total parameters, which have been added in the measurement are pH, DO, ORP, EC, and temperature.

3.2.Hardware Modules

The hardware platform plays an essential role in any system development since data acquisition and data processing are done with the help of the hardware platform itself. In this work, commercial off-the-shelf (COTS) modules and devices have been used to develop the experimental setup. These individual modules have been described in detail in the following sections.

3.2.1.Water Quality Sensors

The water quality sensors are selected according to the requirement and cost criteria. The water quality sensors were purchased from the ATLAS scientific [111]. The detail of each water quality sensor, including the development technology, measurement range, and interfacing circuit is described below. The pictorial representation of sensors & their signal conditioning circuits is given in Appendix A.

(a) pH Sensor

The ATLAS Scientific pH sensor has a 0-14 moles/L measurement range with accuracy and resolution of ± 0.002 and ± 0.001, respectively. These specifications make it suitable for various applications, such as water quality monitoring, soil monitoring, hydroponic industries and food industries. The reference electrode used in the pH sensor is made of silver/silver chloride (Ag/Ag-Cl). The sensor body is made of extruded epoxy, which can withstand extremely powerful acids and bases [112].

(b) Electrical Conductivity Sensor

The ATLAS Scientific electrical conductivity sensor has a measurement range of 0.07 − 50,000 μS/cm. The accuracy of the conductivity sensor is ± 2 %, and the response time is 90% in 1 second. Most conductivity sensors have the problem of fringe effect, which means the reading varies when a sensor comes nearby any object. Still, the ATLAS Scientific sensor is designed in a way that it has no fringe effect resulting in stable measurement readings. The applications of the conductivity sensor are water quality testing, soil testing, hydroponics, and fish keeping [113].

(c) Dissolved Oxygen Sensor

The ATLAS scientific DO sensor is a tiny probe that can work in different ambient environments, from water quality monitoring to fish farming. It has a measurement range of 0-100 mg/L with an accuracy of ± 0.05 mg/L. The response time of the sensor is ~0.3 mg/L/sec. The sensor can be used for environmental monitoring, wine testing, fish farming, and hydroponic applications [114].

(d) Oxidation Reduction Potential Sensor

The ATLAS scientific ORP sensor has a measurement range of ± 2000 mV with an accuracy of ±1 mV. The sensor response time is 95% in 1 sec. Because of the probe's chemical inert body, it can be exposed to fluorinated compounds and other potent oxidizers and reducers that would ordinarily damage a lower-quality ORP probe. The Atlas Scientific lab-grade ORP probe provides accurate results in bizarre chemicals also [115].

(e) Temperature Sensor

The ATLAS scientific platinum sensor (PT-1000) has a measurement range from -200ºC to +850ºC. It is a Class-A high purity platinum sensor. The platinum sensor within is quickly heated by the 304 stainless steel tip, resulting in short latency and high precision readings [116].

All the water quality sensors have their dedicated signal conditioning circuits [117]–[121], making them suitable for quick connection and measurement from any controller board which supports UART, SPI, or I2C communication protocol. The pH and ORP sensors are based on electrochemical technology. The DO sensor is membrane-based, and a 2-probe measuring technique is used for the conductivity sensor. The unit and measurement range of water quality sensors is shown in Table 3.1.

Table 3.1 Water quality parameters: unit and measurement range

Parameter	Unit	Measurement Range
Dissolved Oxygen (DO)	mg/l	0 to 100
Temperature	ºC	-200 to 850
Electrical Conductivity (EC)	µS/cm	5 to 200,000
Oxidation Reduction Potential (ORP)	mV	-2000 to 2000
pH	moles/L	0 to 14

3.2.2. Core Controller

The Raspberry Pi is used as a core controller in the experimental test-bed setup for data acquisition and further processing. The proposed system architecture uses Raspberry Pi 3 as a core controller. It is a credit card-sized mini-computer operating on a 5V micro-USB power supply with an ARM cortex A-53 quad-core 1.2 GHz 64-bit CPU. It has many onboard components like four USB ports, 40 GPIO pins, one Wi-Fi and one Bluetooth module, one camera connector, one HDMI port, one display connector, one Micro SD slot, LEDs for the status indicator, and one Ethernet port [122]. The Debian OS is installed in a Micro SD card inserted in the SD card slot. Different applications of the Raspberry Pi can be found in [123]–[127]. Many controller boards are available in the market, such as Arduino, NodeMCU, Raspberry Pi, etc. The main advantage of using Raspberry Pi over other development boards is that it has almost every module on board. Hence, there is no need to interface any module externally. Additionally, the computational capability of Raspberry Pi is much better than other controller boards. A diagram of Raspberry Pi (Model No. 3 B+) with detailed information is shown in Figure 3.1.

3.2.3. Display Screen

A 7-inch touchscreen from Waveshare (firmware 2.1) [128] was used for the interactive human-machine interface (HMI) and displaying the results acquired from measurement and analysis. The LCD screen resolution is automatically set while connecting to windows OS. In the case of Raspberry Pi, the screen resolution needs to be set manually by editing the root file of the Debian OS; otherwise, the LCD screen will not work. This touch screen supports multi-touch up to 10-points. A detailed schematic of the Waveshare touch screen is shown in Figure 3.2.

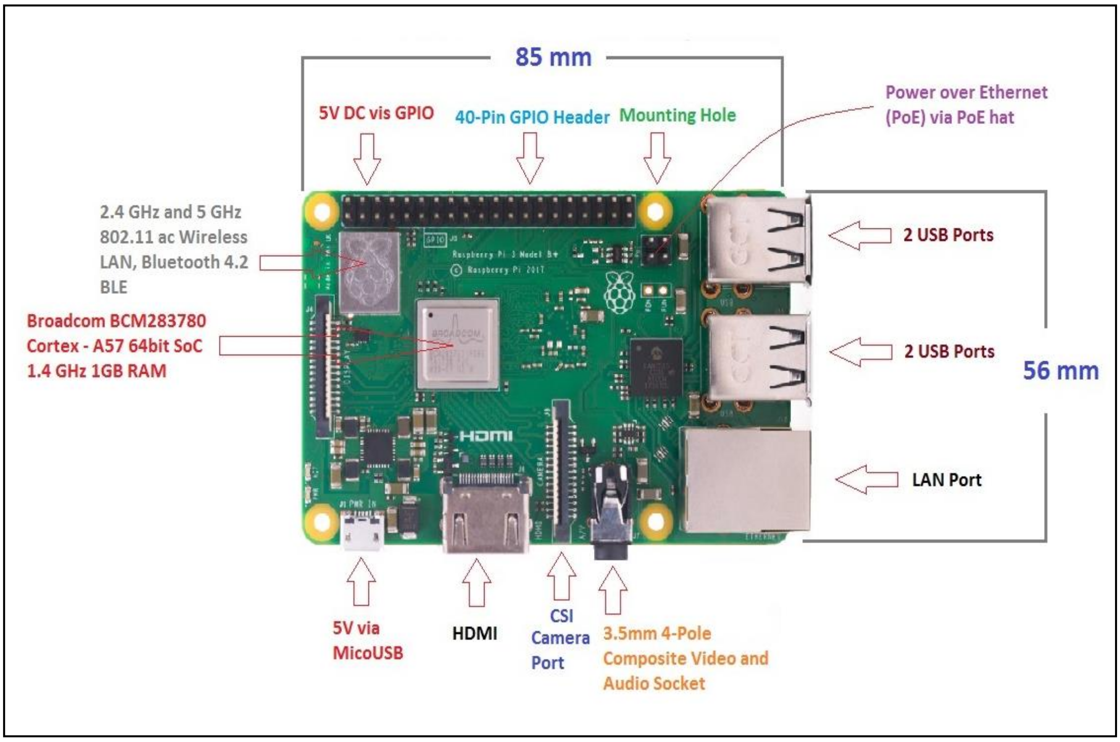
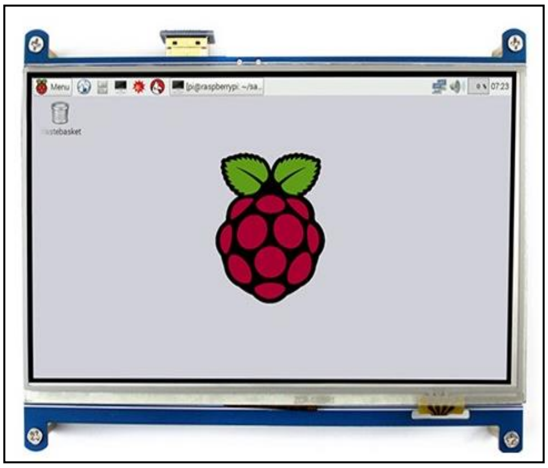


Figure 3.1 Raspberry Pi 3 B+ board detailed schematic



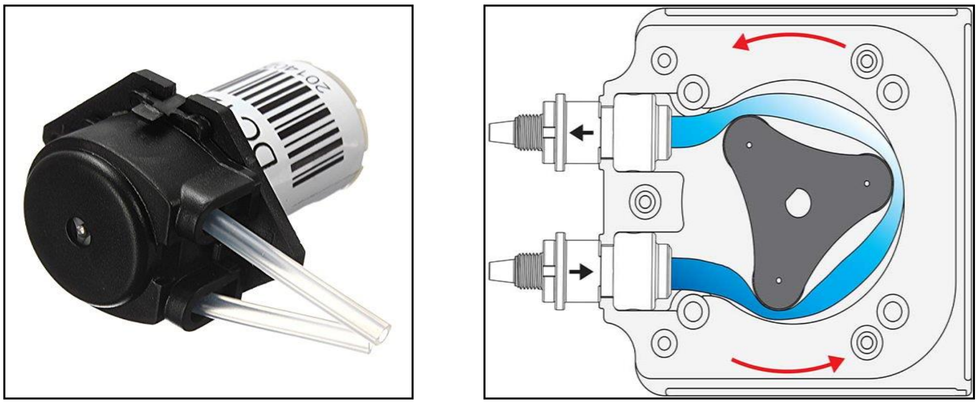
(a) Front View

(b) Back View

Figure 3.2 Waveshare 7-inch touch screen

3.2.4. Pump and driver circuitry

The peristaltic pump is used in this work to automate water intake and throughput from the sample container. It is a type of positive displacement pump in which a flexible tube is inserted within a circular pump casing to confine the sample. Most peristaltic pumps use rotary motion to carry the sample. Several rollers are attached to the rotor's external circumference, compressing the flexible tube as it rotates. A peristaltic pump and schematic of the pump head are shown in Figure 3.3. A 12V power supply operates the DC motor attached to the pump head.



(a) (b)

Figure 3.3 Peristaltic pump and pump head

An L298N driver module was used for driving the pump in both directions, clockwise and anticlockwise. The speed and direction can both be controlled using the module. The direction can be changed using H-bridge, and the speed can be controlled by Pulse Width Modulation (PWM). Figure 3.4 shows the L298N driver module, which was used in this work. Four DC motors or two motors with directions can be controlled with this module. The specifications of the driver module are shown below [129].

- Model No – L298N 2A
- Dual H Bridge
- Maximum Voltage – 46 V
- Minimum Current – 2A
- Logical Current – 0-36 mA
- Maximum Power – 25 Watt
- Heatsink for better performance
- Current Sensor for motor

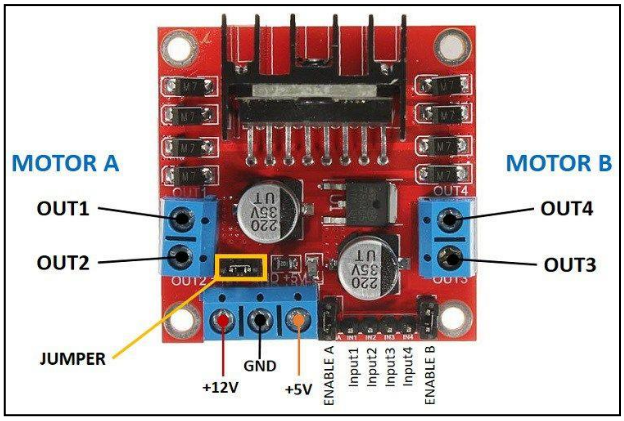


Figure 3.4 L298N motor driver module

3.2.5. Zigbee Module

A Zigbee module is used in this work for wireless data transmission for water quality monitoring in a distribution network. It is a series 2 module (S2C), which provides a low-cost wireless solution for device development. The Zigbee module is based on the wireless communication protocol standard IEEE 802.15.4. It is most widely used in home automation applications. The data transmission rate is 250 Kbps (RF) and up to 1 Mbps (serial). The coverage range of the Zigbee module is 200 feet indoor and 1200 meters outdoor. It has a low transmission power of 3.1 mW (+5 dBm) and supports up to 16 channels [130]. Figure 3.5 depicts a Zigbee module.

Figure 3.5 Zigbee module

3.2.6. Arduino and NodeMCU Development Boards

The Arduino is an open-source prototype used to read the sensor output, activate an actuator, and publish the data online. It consists of an ATmega2560 microcontroller, which can be programmed using embedded C or C++ in the Arduino software

Integrated Development Environment (IDE). The Arduino board has several analog and digital channels along with 6-PWM channels to sense the input and generate the output. The communication between sensors and Arduino is done through the UART protocol [131]. The NodeMCU development board is based on the ESP8266 controller, which combines GPIO, UART, I2C, PWM, and ADC on a single board. It is an open-source, simple, low-cost, programmable, and Wi-Fi-enabled IoT platform [132]. A detailed pinout schematic of the Arduino and NodeMCU development boards is shown in Figures 3.6 and 3.7, respectively.

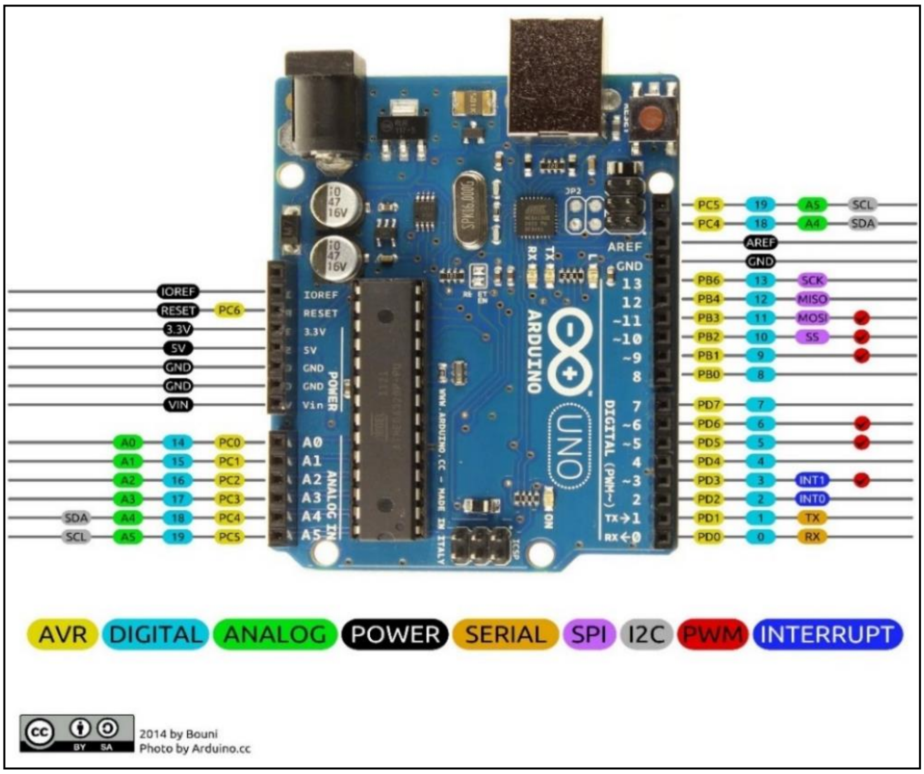


Figure 3.6 Arduino development board schematic

3.2.7. Power Supply and Protective Case

A 5V-2A adaptor operates the Raspberry Pi development. A 3.3V supply to operate the water quality sensors and interfacing circuitry is available on the Raspberry Pi board. Additionally, a 12V Switch Mode Power Supply (SMPS) is used for the motor as the 5V adaptor cannot supply the current and voltage needed for the peristaltic pump. A case made of a Perspex sheet has been used to protect all the components of the system. The dimensions of the case are 30x30x30 cm^3.

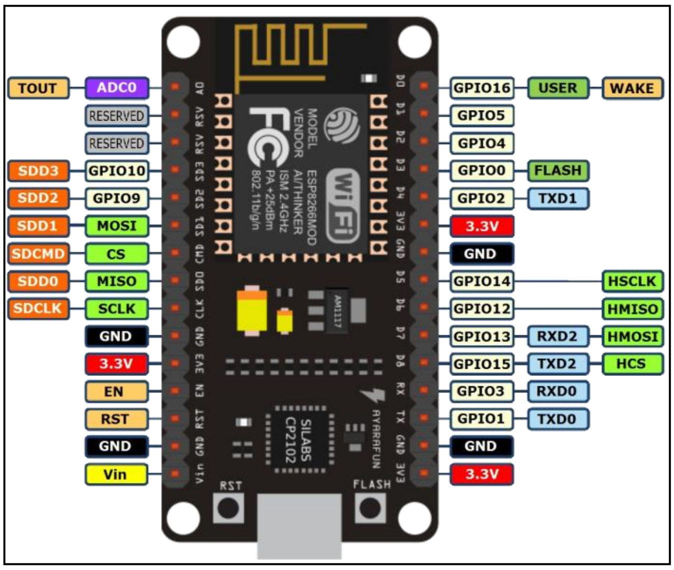


Figure 3.7 NodeMCU development board schematic

3.3.Benchmark Instrument

The reference data used for calibration and evaluation of the developed system was obtained from an Industrial-grade water quality monitoring system. Exo-1 Multiparameter Sonde as shown in Figure 3.8. It is imported from YSI Incorporated, Yellow Springs, Ohio, USA [133]. It is a highly versatile instrument having the ability to monitor various water quality parameters such as conductivity, dissolved oxygen (DO), oxidation reduction potential (ORP), pH, turbidity, total dissolved solids (TDS), total suspended solids (TSS), and salinity. It can be used for marine water, freshwater, surface water, groundwater, and estuarine water. The reference is calibrated regularly to maintain its accuracy. The sampling time of the benchmark is ~30 seconds. The measurement can be done with a wired connection or wireless via Bluetooth in the computer attached. It is a very robust instrument and can be deployed anywhere in the field. There are four universal ports in the reference instrument. The sensor can be connected to any port. It will automatically identify the sensor and measures the water quality parameter accordingly. The sensor specifications are shown in Table 3.2.

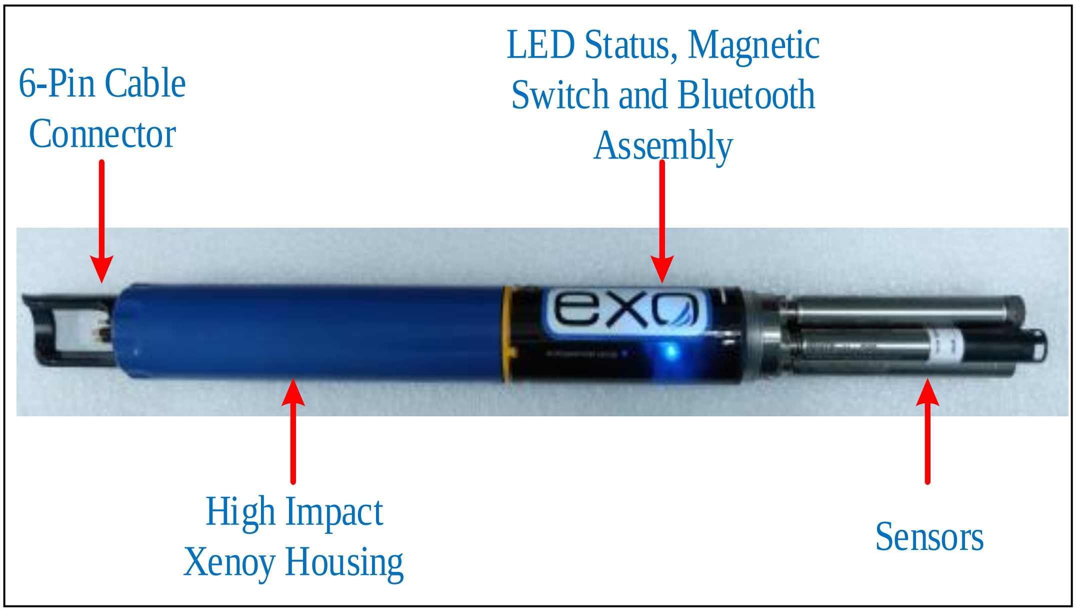


Figure 3.8 Exo-1 multiparameter sonde

Table 3.2 Benchmark instrument sensor specifications

Sensor	**Unit**	**Measurement Range**	**Accuracy**
pH	Moles/L	0-14	±0.1 pH units within ±10°C of calibration temperature; ±0.2 pH units for entire temp range
DO	Mg/L	0-50	0-200%: ±1% reading or 1% air sat., whichever is greater; 200-500%: ±5% reading 0-20 mg/L: ±1% of reading or 0.1 mg/L; 20-50 mg/L: ±5% reading
ORP	mV	-999 to +999	±20 mV in Redox standard solution
EC	mS/cm	0-200	0-100 mS/cm: ±0.5% of reading or 0.001 mS/cm, whichever is greater; 100-200 mS/cm: ±1% of reading
Temperature	ºC	-5 to +50	-5 to 35°C: ±0.01°C 35 to 50°C: ±0.05°C

3.4. Proposed Framework

A major part of the hardware development is the integration of different modules with the core controller. As discussed earlier, five water quality sensors were selected and characterized. The sensors are arranged in an array to make a Multi-Sensor Array (MSA), as shown in Figure 3.9.

The hardware was developed on a step-by-step basis, starting from sensor interfacing to experimental testbed setup to final prototype development. The MSA was connected to a signal conditioning circuit dedicated to each sensor. The signal conditioning circuit can provide raw data (voltage) and the measured water quality parameter. Then the signal conditioning circuits were connected to the core controller (Raspberry Pi) through a multiplexer to extend the UART channels as the Raspberry Pi has a limited number of UART. A keyboard and a mouse were also interfaced with the Raspberry Pi for programming and user interfacing. The interfacing of water quality sensors to Raspberry Pi is shown in Figure 3.10.

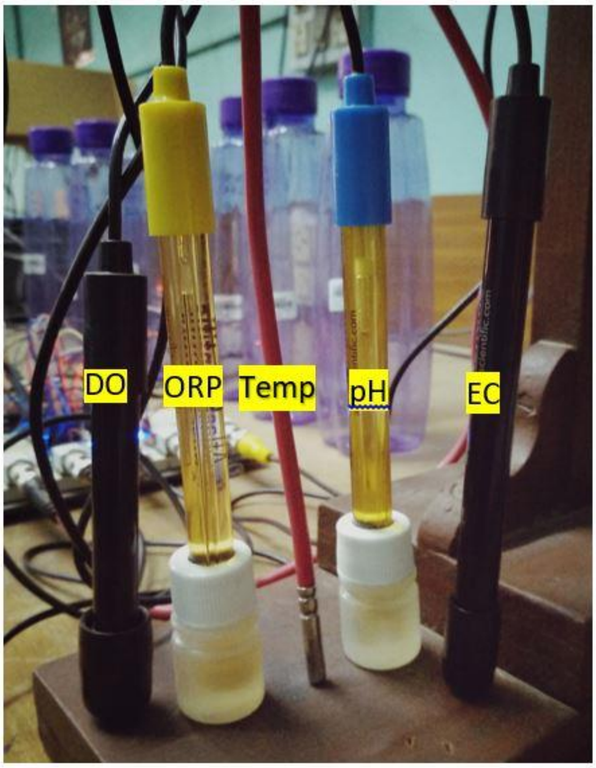


Figure 3.9 Multi-Sensor Array (MSA)

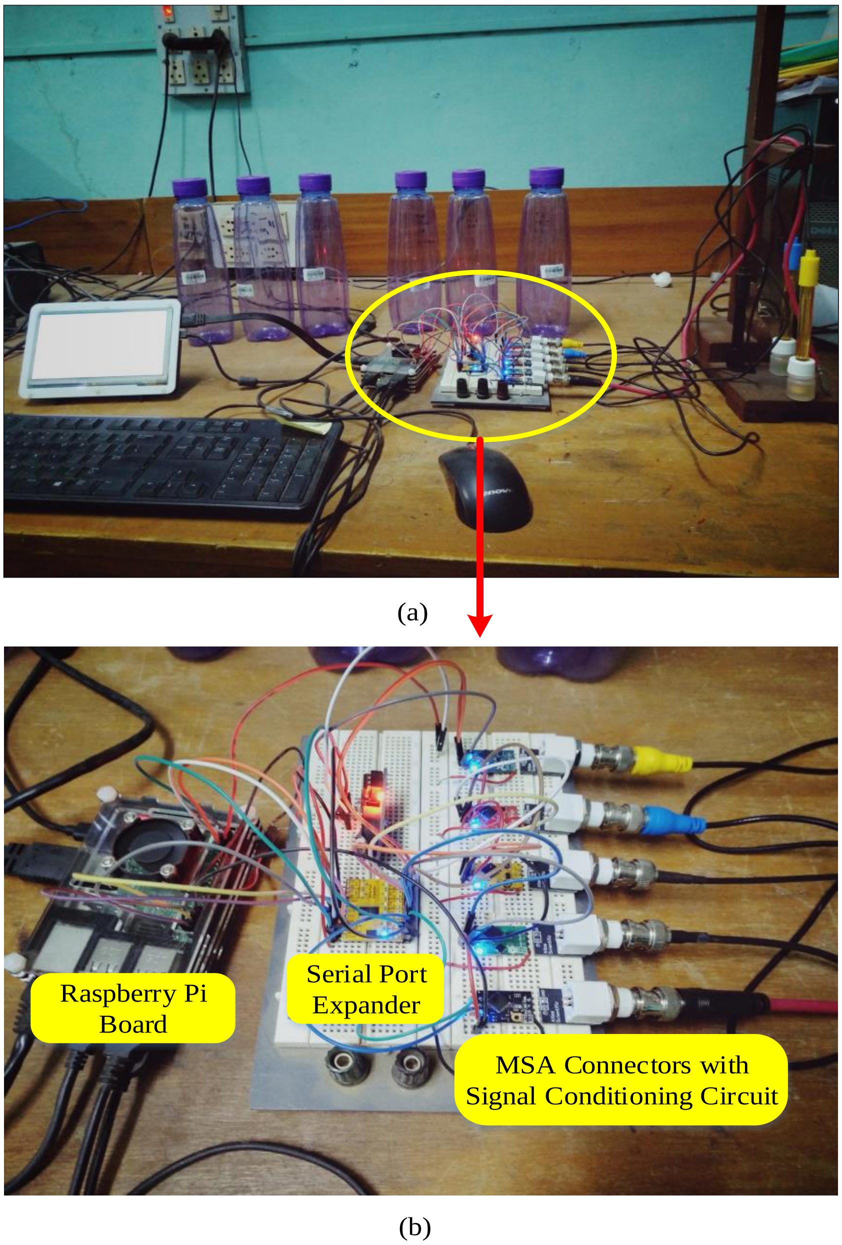


Figure 3.10 Experimental testbed setup

Initial experiments were performed on this setup by manually pouring the sample into the container and placing the sensor array in the container. The next step in the final hardware development was the automation of measurement. The peristaltic pump was used for auto intake and throughput. The SMPS was used to supply sufficient voltage and current to the peristaltic pump. The proposed schematic and final developed prototype are shown in Figures 3.11 and 3.12, respectively. The specifications of the developed prototype are given in Appendix B.

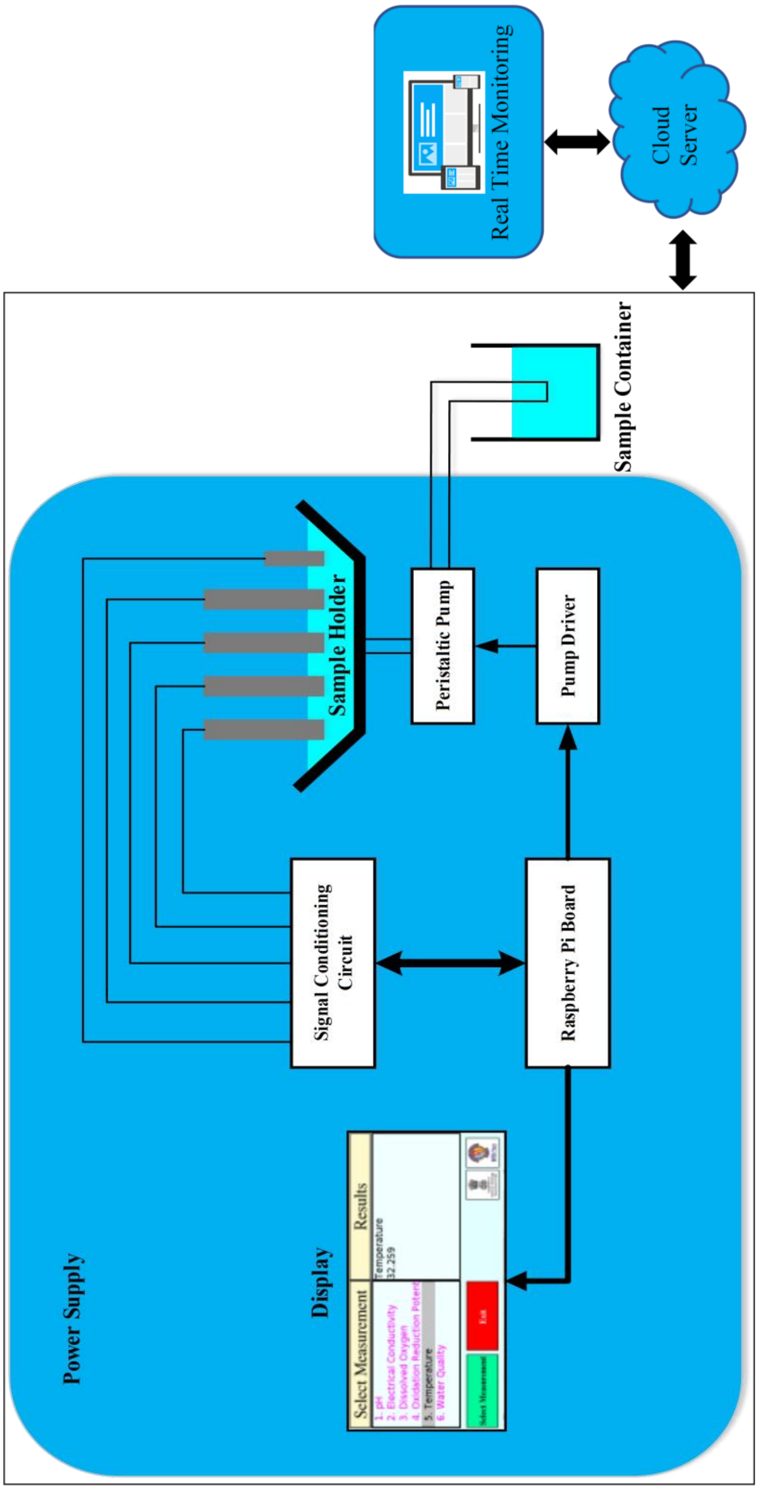
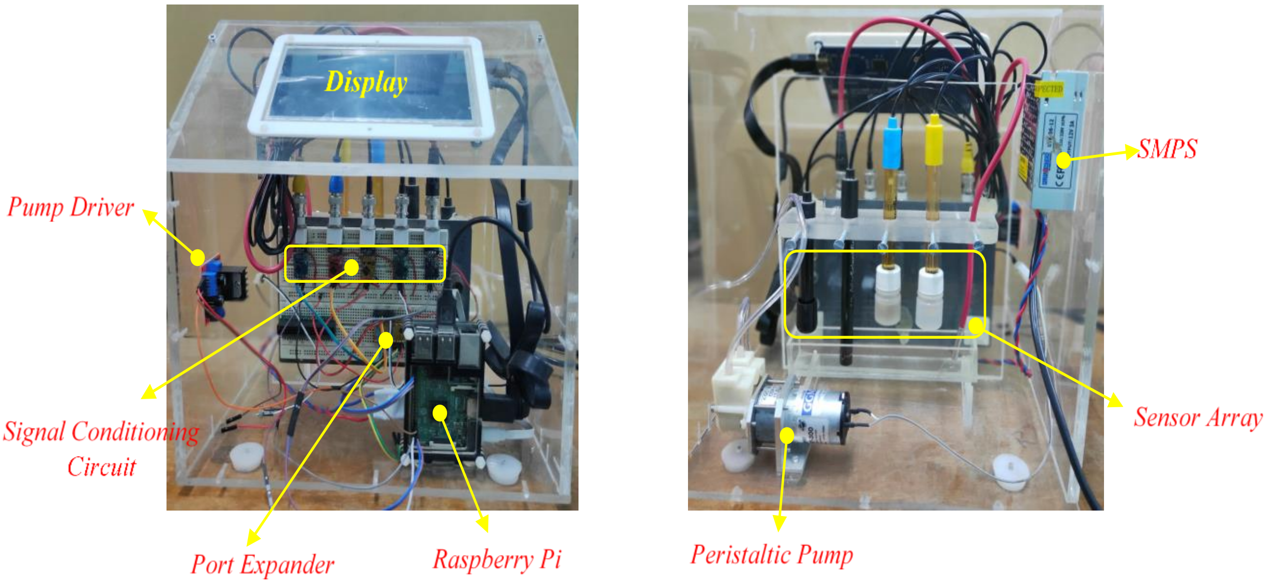


Figure 3.11 Schematic of water quality monitoring system



(a) Front view (b) Back view

Figure 3.12 Prototype system

3.5. Software Framework Development

3.5.1. Python-Programming Software

Python is an open-source software platform, which is freely distributable or usable for educational and commercial use. It was suggested by the DST project review committee. Python was coined by Guido Van Rossum in the late 1980s, and its first version was released in December 1989. Afterward, many versions were released. The latest version is Python 3.9.7, which was released on 31st August 2021. The Python software can be downloaded from www.python.org. Different sample scripts, modules, tools, or documentation can also be downloaded from the Python site. It supports object-oriented programming and graphical programming as well. It can be used for data analysis, exploration, and visualization by embedding different libraries [134]. The programming for the experimental procedure was written in Python V3.7, which was installed in the Debian OS on the Raspberry Pi board. Many scientific packages, e.g., NumPy, SciPy, scikit-learn and Matplotlib have been embedded in the programming for the proposed system.

3.5.2. Interactive User Interface

The graphical representation was provided for real-time data obtained from various sensors with the help of the GUI platform for the interactive HMI. An interactive GUI has been designed in the Python framework with a touch interface. Initially, the user login function is provided to enter the GUI, as shown in Figure 3.13. The touch interface is provided for ease of operation for the operator. In the GUI, the operator can select the measurement from the menu, whether it is an individual parameter or the overall water quality, with a single touch (refer to Figure 3.14). The acquired data were kept for future use by saving them in a memory drive provided with the Raspberry board. The live plotting of data is shown in Figure 3.15. The x-axis represents time, and the y-axis represents the sensor node reading.

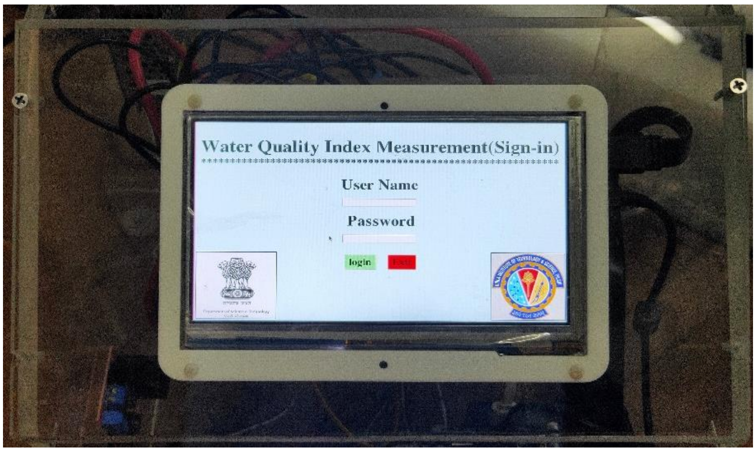


Figure 3.13 Login panel to access GUI

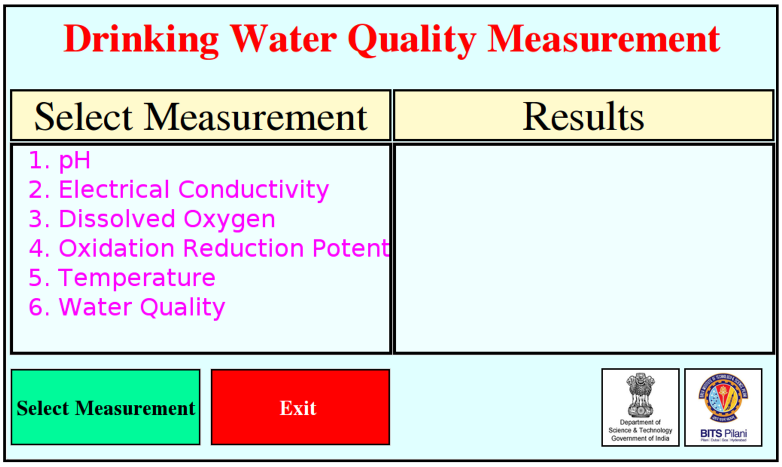


Figure 3.14 Interactive user interface

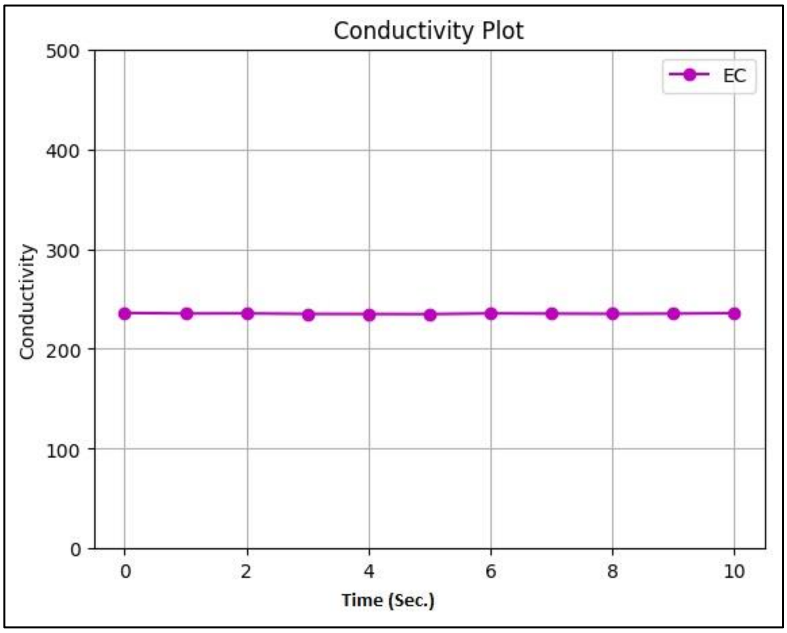


Figure 3.15 Conductivity plot in GUI

3.6. Experimental Procedure

3.6.1. Study Area and Sample Collection

The sample collection area in this study is BITS-Pilani, Pilani campus, which is located at the Pilani town in Rajasthan province of India, which comprises an area of 1320 acres (5.3 km^2). This area is located at 28° 21′ 49.96″ N and 75° 35′ 13.26″ E. Six different locations have been considered for the sample collection. The only source of water on the campus is groundwater. There are different wells located on the campus, which have been targeted for sample collection.

3.6.2. Calibration of the sensors

The calibration should be done before the measurement to get accurate readings from the sensors. We have calibrated the water quality sensors using reference solutions at ambient temperature. The reference solutions for the pH sensor are 4 pH, 7 pH, and 10 pH. The conductivity sensor was calibrated using a 1000 µS/cm reference solution. The DO sensor was calibrated using 0 mg/l solution, and the ORP sensor using a 225-mV reference solution (refer to Appendix B). A 3-point calibration was performed for the pH sensor, whereas the conductivity, DO, and ORP sensors were calibrated using 1-point calibration [133]. This calibration procedure was followed for benchmark equipment. For the developed system, the calibration procedure was the same for pH, DO, and ORP sensors. For the conductivity sensor, 2-point calibration was performed to cover a wide range of accuracy [135]. All the reference solutions were of analytical grade and non-toxic.

Figure 3.16 Location of sample collection

3.6.3. Sample Collection Procedure

The samples were collected from different wells located in the campus area mentioned above. The sample collection and subsequent testing were done on experimental test-bed setup (as shown in Figure 3.10) as well as the benchmark instrument. The sensors were calibrated with the reference solutions before the measurement procedure to avoid any uncertainty in the measurement.

For the developed prototype, approx. 50 ml of the water sample for each sample measurement is injected from the sample container into the sample holder by the peristaltic pump. After the injection, the signals of the electrodes are acquired by the Raspberry Pi. After each measurement, the sample is throughput into the sample container. This measurement sequence was controlled by an algorithm written in Python installed on the Raspberry Pi board. The block diagram shown in Figure 3.17 represents the overall process of the sample measurement.

Figure 3.17 Block diagram of sample measurement

3.7. Summary

Water quality monitoring before consumption is essential to reduce the risk of illness in an individual. The traditional chemical-based water quality approach consumes more time for sample collection and conveyance to the laboratory for testing. These conventional methods are now obsoleting due to the development of a real-time water quality monitoring system. The design and development of a water quality monitoring system are presented in this chapter. The proposed water quality monitoring system can provide the measurement of various water quality parameters. The measured parameters are pH, EC, DO, ORP, and temperature. Based on the acquired water quality parameters, the overall water quality can be defined based on the statistical method or soft computing technique, which has been presented in the next chapter.

Chapter 4

Data Analysis and Results

Preamble

After the data acquisition from the developed setup, the next essential step is water quality analysis. The water quality can be defined by either statistical method or soft computing method. Water Quality Index (WQI) is a unique rating for water quality monitoring. It represents overall water quality in a single term and helps decision-makers evaluate the water quality and its possible usage. As discussed in the literature review chapter, people collect the measurement data and later perform the analysis for water quality monitoring, which is not real-time. This chapter deals with the data analysis employing different methods for water quality monitoring on the developed setup in a real-time environment right after the data acquisition. The classification of water quality employing fuzzy modeling is presented in this chapter. This chapter also deals with the WQI calculation by statistical method and Artificial Neural Network (ANN). The validation of measured water quality parameters has also been presented in this chapter.

4.1. Introduction

Data analysis is a process of visualizing, cleansing, smoothing, modeling, and analyzing the data, which helps in extracting information from the data and making decisions based on the analysis. We can relate data analysis with our day-to-day lives, analyzing our daily observations and making future decisions. The same happens with the data analysis. There are multiple ways for data analysis based on statistical modeling and soft computing techniques. The statistical modeling establishes the mathematical relationship between dependent and independent variables, which results in regression coefficients that can be used for the analysis. Various statistical modeling methods are Principal Component Analysis (PCA), Principal Component Regression (PCR), Multiple Linear Regression (MLR), and Partial Least Squares Regression

(PLSR). The soft computing methods are based on machine learning methods, such as Artificial Intelligence (AI) algorithms, fuzzy computing, and Genetic Algorithm (GA). Whilst talking about water quality analysis, three types of analysis are available based on indexing tool, statistical modeling and ANN, which have been discussed in the following sections.

4.2. Indexing Tool for Water Quality Analysis

Water Quality Index (WQI) is a specific and clear metric for policymakers to assess water quality and its potential applications. The WQI provides overall quality by integrating information from water quality sensors and converging information into a single value. Different aggregate functions include arithmetic aggregate mean, multiplicative aggregate function, harmonic mean, and geometric mean to calculate WQI [53]. The calculation of the WQI includes three necessary steps [136].

(a) Obtaining individual water quality parameters,
(b) Transforming water quality parameters into subindices to represent them on the same scale
(c) Applying an aggregate function to measure the WQI.

The researchers developed several water quality indices based on the aggregate functions used, but there is no specific globally accepted method. Even though multiple aggregation algorithms have been devised, with various changes suggested, arithmetic mean, and multiplicative geometric mean functions remain the most widely used [137]. Choosing the best aggregation methodology is a never-ending task, given that each method has its own set of benefits and drawbacks. As a result, it is up to the water quality index (WQI) developers to use their expertise and knowledge to choose the most appropriate method, preferably with the minimum drawbacks. Based on the abovementioned facts, we have adopted the arithmetic mean aggregation method for WQI calculation. The water quality has been divided into five categories based on their uses, as shown in Table 4.1. The aggregate function for WQI calculation is given by Eq. (4.1).

$$\boldsymbol{WQI} = \frac{\sum \boldsymbol{W_n Q_n}}{\sum \boldsymbol{W_n}} \tag{4.1}$$

Where W_n is the unit weight and Q_n is the rating for *nth* parameter. Q_n is given by Eq. (4.8).

$$\boldsymbol{Q_n = \frac{D_n - D_i}{D_s - D_i} * 100} \tag{4.2}$$

Table 4.1 WQI: range, category, and possible application

Water Quality Index	**0-25**	**26-50**	**51-75**	**76-100**	**>100**
Category	Excellent	Good	Poor	Very Poor	Not suitable for any application
Possible applications	Industrial, irrigation, and drinking	Industrial, irrigation, and drinking	Industrial and irrigation purpose	Irrigation purpose only	Treatment is necessary before use

Where D_i is the optimal value for the parameter which is '0' (with the exception of DO (14.6 mg/L) and pH (7.0)), D_n is the value of *nth* parameter, and D_s is the standard value. W_n is determined by $W_n = k/D_s$. Here k is the proportionality constant and given by Eq. (4.3).

$$\boldsymbol{k = \frac{1}{\sum 1/D_s}} \tag{4.3}$$

Where $s = 1,2, \dots, n$. The WQI range, it's category, and possible application can be found in Table 4.1 [36]. Table 4.2 shows the standard value and calculated weight for each parameter. The constant of proportionality is calculated using Eq. (4.3) and is equal to 2.875.

Table 4.2 Standard values and calculated unit weights for all the parameters

Parameters	**Standard permitted limit**	**Calculated weight**
Temperature	35	0.088
pH	6.5-8.5	0.365
Dissolved Oxygen	5	0.62
Oxidation Reduction Potential	600	0.0051
Electrical Conductivity	1,000	0.0031

After the calculation of unit weight, the WQI is calculated using Eq. (4.1). The measured water parameters and their calculated water quality index is shown in Table 4.3. From Table 4.3, it can be found that two samples are of good quality, which can be used for industrial, irrigation and drinking purposes. The remaining samples are of poor quality and can only be used for irrigation purposes and agriculture.

Table 4.3 Measured water quality parameters and their calculated WQI

Location	pH	EC	DO	ORP	Temp	WQI	Category
1	7.45	385	8.20	213	22.5	49.7051	Good
2	7.62	435	7.90	212	21.5	55.4431	Poor
3	6.95	510	9.50	185	19.6	30.7766	Good
4	8.1	390	9.23	171.1	18.9	55.9950	Poor
5	7.8	445	8.05	191	23.8	58.9636	Poor
6	8.2	1582	7.5	170.5	29.2	73.5747	Poor

4.3. Statistical Modeling for Water Quality Analysis

Different regression methods are available, such as multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression (PLSR). The MLR is the basic of the regression analysis, but the disadvantage of this regression modeling is that it cannot handle missing data and collinearity. This can be handled by PCR analysis. The principal components of X are used as independent variables in PCR to predict the response variable Y. This approach focuses solely on variables that describe X. Unlike PCR, PLSR identifies X components that are significant for Y also. In this study, the PLS regression (PLSR) method is used for the water quality indexing of different water samples. PLSR projects the input and output matrix to the direction of maximum covariance and is utilized to model the relationship between the score matrix of two data sets: X-data and Y-data. Here X is the input data matrix (measured water quality parameters), and Y (WQI) is the output or response data matrix given by Eqs. (4.4) and (4.5).

$$X = \begin{bmatrix} X_{11} & X_{12} & . & X_{1n} \\ X_{21} & X_{22} & . & X_{2n} \\ . & . & . & . \\ X_{m1} & X_{m2} & . & X_{mn} \end{bmatrix} \tag{4.4}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} \tag{4.5}$$

After centralization X and Y are decomposed into the form [138] given by Eqs. (4.6) and (4.7).

$$X = m_1 o_1^T + m_2 o_2^T + \cdots + m_n o_n^T = MO^T + R \tag{4.6}$$

$$Y = n_1 p_1^T + n_2 p_2^T + \cdots + n_n p_n^T = NP^T + S \tag{4.7}$$

Where M and N are score matrices, O and P are loading matrices, R and S are residuals of input and output data matrix, respectively. The relation between M and N is known as inner relation [139], and coefficient B of this inner relation ($N = MB$) is called regression coefficient. Now, the final response of matrix Y can be expressed as Eq. (4.8). And the residuals can be found out using Eq. (4.9) and (4.10).

$$Y = MBP^T \tag{4.8}$$

$$R = X - MO^T \tag{4.9}$$

$$S = Y - NP^T \tag{4.10}$$

The regression coefficients are used to determine the relationship between the input data matrix X and the output data matrix Y. A PLS flow diagram is shown in Figure 4.1, and a detailed description of the PLS algorithm is explained in Annexure D.

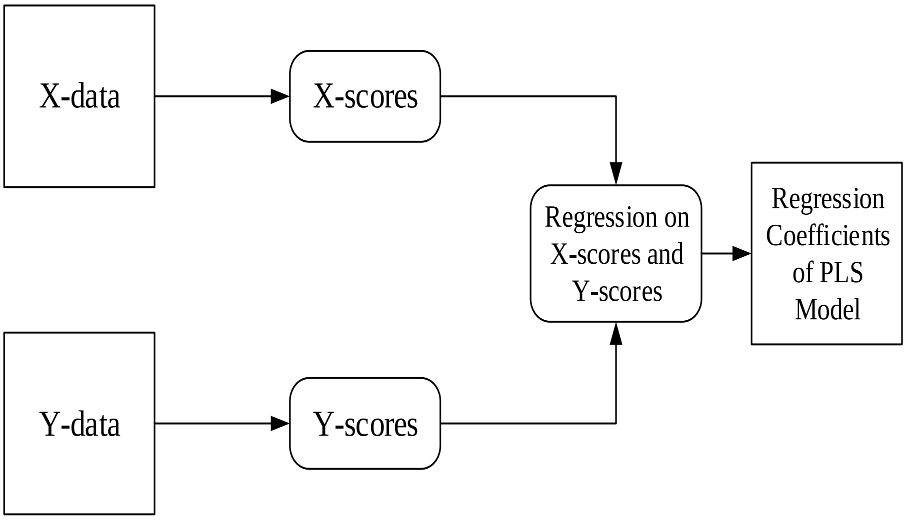


Figure 4.1 PLS flow diagram

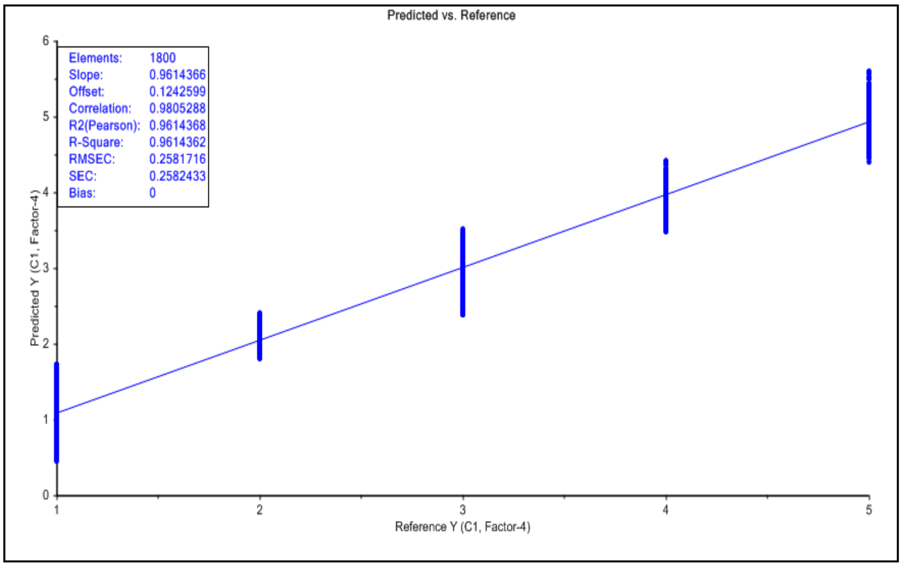
The water quality is defined using an index rating (water quality index (WQI)) developed by the PLSR technique. The dataset was generated using uniformly distributed random numbers, and different categories were assigned to the water quality parameters according to their range, as seen in Table 4.4.

Table 4.4 Different water quality parameters, their range, water quality, and defined rating

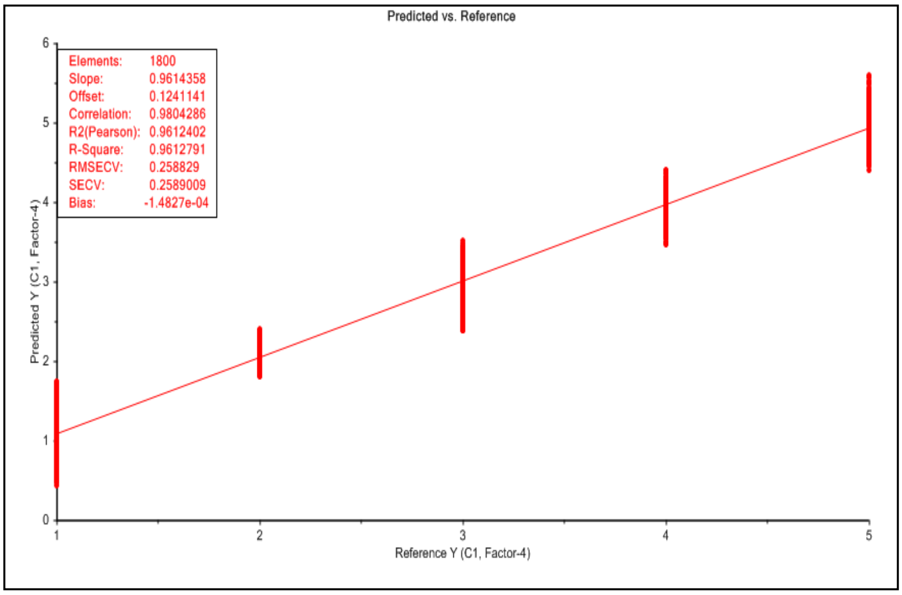
Water Quality Parameters					Water Quality	Rating
pH	**EC**	**DO**	**ORP**	**Temp.**		
(6.5-7.5)	(0-250)	(10-15)	(150-400)	(20-25)	Excellent	Class I
(7.5-8.5)	(250-500)	(8-10)	(400-600)	(15-20) & (25-30)	Good	Class II
(6-6.5) & (8.5-9)	(500-1000)	(4-8)	(600-800)	(10-15) & (30-35)	Satisfactory	Class III
(4-6) & (9-11)	(1000-2000)	(2-4)	(0-150) & (800-1000)	(5-10) & (35-50)	Poor	Class IV
(0-4) & (11-14)	(>2000)	(0-2)	(>1000)	(>50) & (<5)	Bad	Class V

To assign different categories for water quality parameters, reference [140] has been referred. A total of 1800 samples were used in the modeling. The parameters used in modeling are pH, EC, DO, ORP, and temperature. The PLSR model was constructed using a scikit-learn python library. The water quality index was calculated for 50 iterations to ensure the model's reproducibility. The plot for model calibration and validation is shown in Figure 4.2. The regression coefficients and PLSR-WQI equation is shown in Eq. (4.11). The same equation was applied to calculate PLSR-WQI for real water samples also.

$$PLSR - WQI = 4.2556 + 0.0024 * pH + 0.0003 * EC - 0.2590 * DO + 0.0001 * ORP - 0.0006 * Temp \quad (4.11)$$



(a) Model Calibration



(b) Model Validation

Figure 4.2 Predicted vs. reference for model calibration and validation

4.4. Soft Computing for Water Quality Analysis

Soft computing can be defined as a collection of computational methods based on artificial intelligence mimicking the human brain. The word "Soft Computing" was first introduced by Zadeh in 1992. Unlike hard computing techniques, soft computing methods are robust, fast, accurate and cost as well as computationally effective. The distinctive advantages of soft computing are as follows.

- Multiple variables can easily be handled.
- These techniques are scalable and have adaptive nature.
- The hybridization of different soft computing techniques can be done, which reduces the traditional packages dependency.
- Local minima can be prevented in optimization problems.

Different techniques, such as fuzzy modeling, Artificial Neural Networks (ANN), and Genetic Algorithm (GA), were used to usher soft computing in the computing world. Later scope of these techniques was extended to encompass the particle swarm optimization (PSO) and bacterial foraging algorithm (BFO) [141]. In this work, fuzzy modeling and ANN have been used for water quality analysis, which are discussed in the following sections.

4.4.1. Fuzzy Modeling for Water Quality Analysis

There are two approaches for obtaining the information required to design, evaluate, and implement real-world engineering systems. One method gathers knowledge through sensors in experimental measurements, while another extracts expert knowledge in linguistic form. The language information is simple to comprehend and apply. Expert systems are systems that are built on expert knowledge. The primary goal behind the development of fuzzy systems is to create a systematic and efficient mechanism to express expert knowledge. Among the numerous artificial intelligence methodologies, fuzzy modeling is one of the most appealing strategies.

The fuzzy inference system (FIS) can include vagueness in decision-making and reasoning. Hence, techniques based on fuzzy logic have proved very useful since they are less mathematically intensive than neural networks, genetic algorithms, etc., and they support approximate reasoning. In FIS, knowledge is presented as linguistic rules. The inputs are converted from a crisp value to a linguistic variable by fuzzification, and these variables are fed to the inference system. This interference system gives a new set of linguistic variables converted to a crisp value with the help of defuzzification [142]. The process to design fuzzy logic thus involves three necessary steps as shown in Figure 4.3: (1) define the membership function for each variable, (2) perform fuzzy inference based on the inference method, and (3) select the defuzzification method to determine water quality.

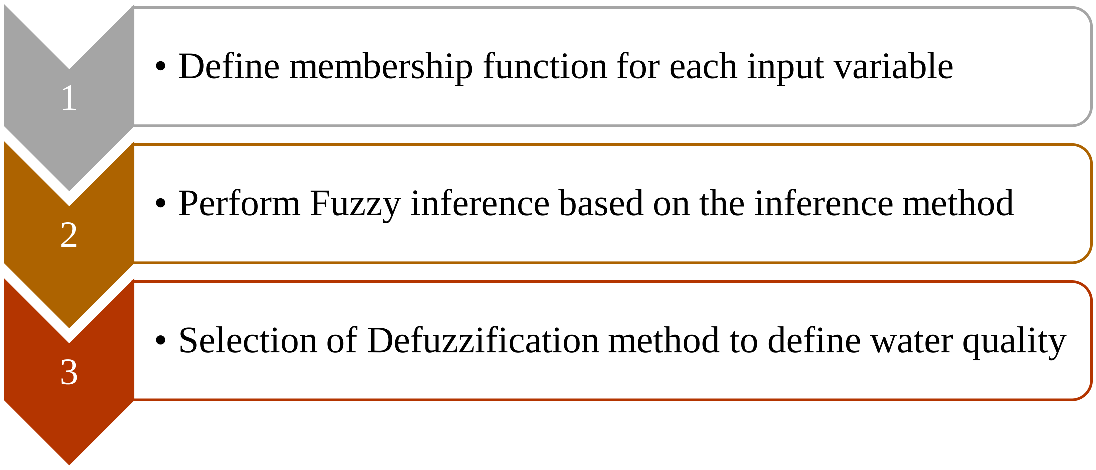


Figure 4.3 Fuzzy logic designing process

The proposed fuzzy logic was implemented in Python with the help of a library developed by the SciKit-Fuzzy development team [143] to define the water quality from groups of five linguistic variables defined as bad, poor, satisfactory, good, and excellent. The fuzzy system uses the Mamdani implication model, which takes five inputs: pH, electrical conductivity (EC), oxidation reduction potential (ORP), dissolved oxygen (DO), and temperature. The Mamdani FIS produces a more accurate response than the Takagi–Sugeno type model since it uses the centroid method of defuzzification. The defuzzified output of the model is water quality, which corresponds to five inputs of the model. In this work, the Mamdani-type FIS model was implemented for the decision support system since it has a spontaneous and rule-based decision-making capability. The modeling was performed based on five input parameters and one output parameter to determine the water quality. The selection of the membership function was made based on the complexity of the system considered for decision-making.

Triangular membership functions (MFs) are the most commonly used membership functions because of their linear nature and ease of implementation [144], [145]. Hence, we have selected a triangular MF to fuzzify the crisp variable into a linguistic one. The triangular membership function depends on three parameters, l, m, and n, given by Eq. (4.12).

$$f(x; l, m, n) = \begin{cases} 0 & \text{for } x < l \\ \dfrac{x-l}{m-l} & \text{for } l \le x \le m \\ \dfrac{n-x}{n-m} & \text{for } m \le x \le n \\ 0 & \text{for } x > n \end{cases} \tag{4.12}$$

The logic operations used in the fuzzy logic are min, max and complement. These operations are defined by the Eqs. (4.13), (4.14), and (4.15), respectively. A and B are two subsets.

$$\mu_{A\cup B}(x) = \max\,[\mu_A(x), \mu_B(x)] \tag{4.13}$$

$$\mu_{A\cap B}(x) = \min\,[\mu_A(x), \mu_B(x)] \tag{4.14}$$

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \tag{4.15}$$

After the logic operations, the “if-then” rule was applied. All the rules were applied in parallel, and the rule which did not affect the output was ignored. The outputs of all rules were then aggregated, and all fuzzy sets that affect the output were combined into a single fuzzy set. Finally, the fuzzy set was converted into a crisp set through defuzzification, in which a single number was generated. There are several methods for defuzzification, such as the centroid, maximum, mean of maxima, height, and modified height method. In this work, the centroid defuzzification method was used, which is the most popular method. The output was calculated by averaging individual centroids, weighted by their heights as given by Eq. (4.16) [146].

$$U_o = \frac{\sum_{i=1}^{n} u_i \mu(u_i)}{\sum_{i=1}^{n} \mu(u_i)} \tag{4.16}$$

where $\mu(u_i)$ is the min and max value of the membership degree of the input values (depends on min and max operator). Here, we have assigned two groups to each water quality parameter: desirable (DES) and undesirable (UNDES), as described in Table 4.5. The triangular membership functions for each water quality parameter are shown in Figure 4.4. If the parameters are in the desirable range, then only fuzzy logic has been applied; otherwise, not. For example, if the pH is in the range of 6.5 and 8.5, the assigned group is DES; similarly, if $\mathrm{pH} < 6.5$ or $\mathrm{pH} > 8.5$, the assigned group is UNDES. In the same way, the group is checked for all five water quality parameters. After checking the desirable group, the individual membership function is assigned to each parameter. Also, we have defined the membership function for water quality on a scale of 0 to 100. After assigning the membership functions, the if-then rule was applied, and overall quality was defined based on the adopted rule-based formulation.

Table 4.5 Groups defined for water quality parameters

Range / Parameters	UNDES	DES	UNDES
pH	< 6.5	6.5-8.5	> 8.5
EC	< 300	300-1000	> 1000
ORP	< 200	200-800	> 800
DO	< 3	3-11	> 11
Temperature	< 2	2-35	> 35

* UNDES = undesirable DES = desirable

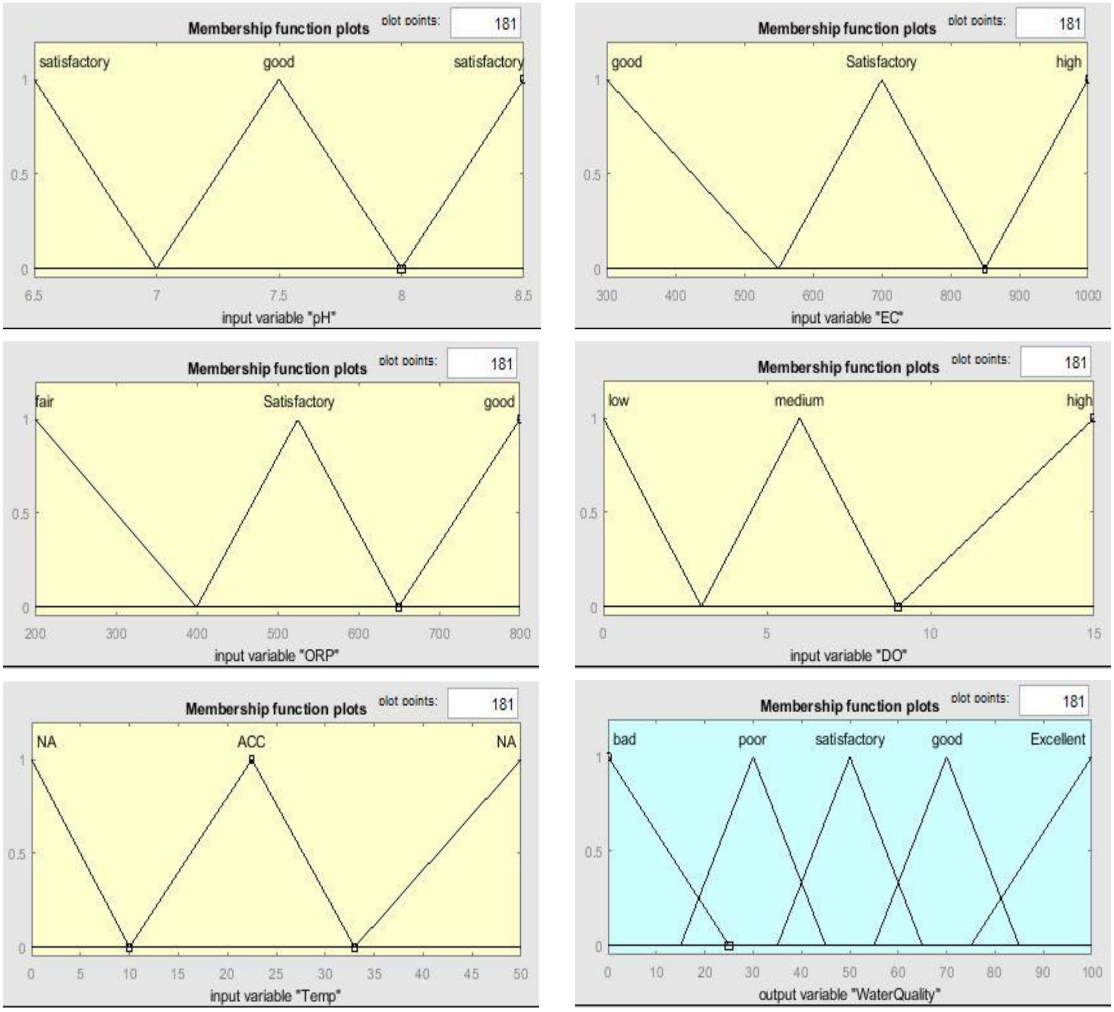


Figure 4.4 Input and output membership functions

4.4.1.1. *Experimental Procedure and Results*

The developed system has been tested for water samples from five different locations. The sensors were calibrated before the measurement to get accurate readings. The calibration has already been discussed in chapter 3. Initially, the measurement iteration was carried out for at least 5 min so that sensor reading became stabilized because the original readings must be recorded only after the sensor attains stability in order to make any conclusive decision out of data. The system was tested for a total duration of 21 hours over seven days. The average values of the experiment for all locations are shown in Table 4.6. Locations 1 to 5 are the drinking-water samples from the distribution networks, and the samples from locations 6 to 8 are the simulated sensor readings. This was done to validate the sensor readings, as the acquired results from the distribution networks are in real-time. Hence, sensor readings do not have much variation within the desirable range. All the targeted parameters were measured through the developed system and benchmark instrument as well. The results obtained from MSA were accumulated through the fuzzy inference system implemented in the Python framework. The water quality defined by the fuzzy modeling is shown in Table 4.7. The water quality is good for locations 1, 2, 4, and 5 in the distribution network and satisfactory for location 3. For locations 6 to 8, the water quality is poor. To measure the accuracy of the developed system and for the authenticity of the results acquired, the proposed system was compared with the benchmark instrument, and the percentage relative error (PRE) was calculated [147]. PRE expresses the percentage error to determine the accuracy given by the following formula.

$$\mathbf{PRE} = 100 * \left(\frac{\mathbf{actual} - \mathbf{observed}}{\mathbf{actual}} \right) \tag{4.17}$$

Table 4.6 Measured water quality parameters and their calculated PRE
(a) pH, DO and EC

Location	**pH**			**DO (mg/l)**			**EC (µS/cm)**		
	MSA	**Benchmark**	**PRE (%)**	**MSA**	**Benchmark**	**PRE (%)**	**MSA**	**Benchmark**	**PRE (%)**
1	7.45	7.51	0.8	8.20	8.16	0.49	385	387	0.51
2	7.62	7.68	0.78	7.90	7.89	0.13	435	426	2.1
3	6.95	6.99	0.57	9.50	9.47	0.32	510	515	0.97
4	8.1	8.2	1.2	9.23	9.21	0.22	390	385	0.13
5	7.8	7.87	0.88	8.05	7.98	0.88	445	455	0.22
6	8.2	8.21	0.12	7.8	7.7	1.2	1582	1576.2	0.36
7	8.1	8.04	0.75	7.5	7.4	1.35	1635	1621.8	0.81
8	8.1	8.01	1.12	7.5	7.4	1.35	1672	1662.1	0.59

(b) ORP and Temperature

Location	ORP (mV)			Temperature (ºC)		
	MSA	**Benchmark**	**PRE (%)**	**MSA**	**Benchmark**	**PRE (%)**
1	213	212	0.47	22.5	22.7	0.88
2	212	210	0.95	21.5	21.3	0.93
3	185	187	1.06	19.6	19.7	0.50
4	206	208	0.96	18.9	19.1	1.04
5	191	194	1.54	23.8	24.1	1.24
6	170.5	168.5	1.18	29.2	28.98	0.76
7	170.9	168.7	1.30	30.1	29.84	0.87
8	171.1	168.4	0.58	30.5	30.21	0.96

The calculated PRE is presented in Table 4.6 for all the samples, and the graphical representation is shown in Figure 4.5. The x-axis shows the different locations, and the y axis displays the corresponding calculated PRE. For all the acquired water quality parameters, the PRE lies in between 0 % and 2 %, except EC, which is in the range of 0 % to 3 %, thus showing the good accuracy of the sensors used. Based on the results of the parameters obtained from MSA, water quality has been defined for all the locations using fuzzy libraries, as shown in Table 4.7.

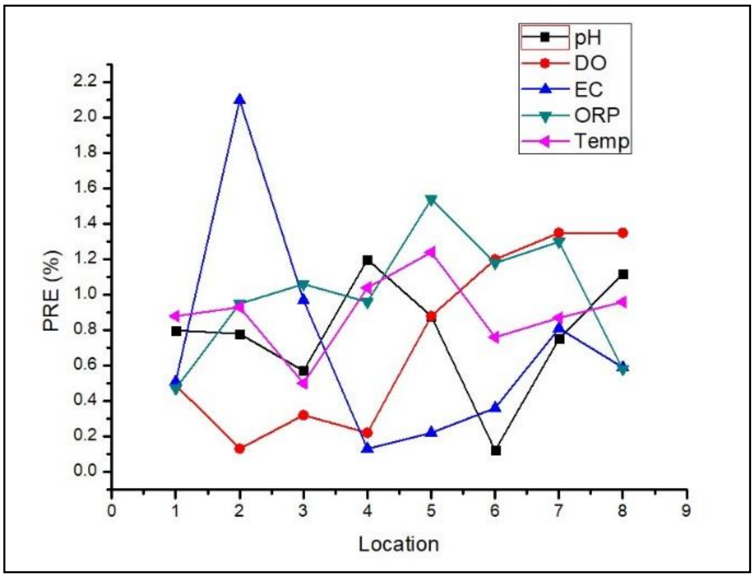


Figure 4.5 PRE calculated for different locations

Table 4.7 Fuzzy water quality for all locations

Location	**Fuzzy Water Quality (FWQ)**
1	Good
2	Good
3	Satisfactory
4	Good
5	Good
6	Poor
7	Poor
8	Poor

4.4.2. Artificial Neural Network (ANN) for Water Quality Analysis

The ANN can mimic the human brain as this has been inspired by a biological neural network. McCulloch first coined the ANN in 1943. The ANN may be seen as a parallel processor, which stores knowledge as well as performs computing. It can learn in either supervised mode or unsupervised mode. Figure 4.6 depicts the structure of an ANN neuron in which the input variable can be represented as a matrix (A) of the dimension $Nx1$. These inputs are fed to the neuron, where they are multiplied by weights (w_i). Then these weighted inputs are summed up, and a bias (b) is added. The output of the neuron can be denoted as x. The final output y is the transfer function of x, represented by Eq. 4.18.

$$y = f(x) = f(wA + b) \tag{4.18}$$

In the ANN network, the neurons are placed in parallel, and this parallel form of the neuron is collectively called a layer. Layers may be arranged one after the other as the complexity of the neural network grows, with the output of one layer acting as the input to the next. Each layer does not have to have the same number of neurons. There are three basic steps to implement the ANN as follows:

1. Data generation: The weights and biases of every neuron in all layers have a significant role in how a neural network works. The input-output linkages in the given problem are studied to determine these values. As a result, a neural network for the problem requires a large quantity of accurate, pre-determined input and output. This information can be gathered by experimentation or simulation.
2. Training: By modifying its weights and biases, the neural network 'learns' to produce the desired output. i.e., a neuron learns a certain rule, alters its weight and bias, and trains itself to produce a specific set of outputs. Perceptrons—computer models that can imitate the brain's ability to discriminate—are used throughout the process.
3. Testing: The neural network is tested for specified test vectors, and the outputs are inspected and validated for accuracy after obtaining the optimal weights and biases of the neurons. The values for the weights and biases can be regarded as finalized if the desired results are obtained.

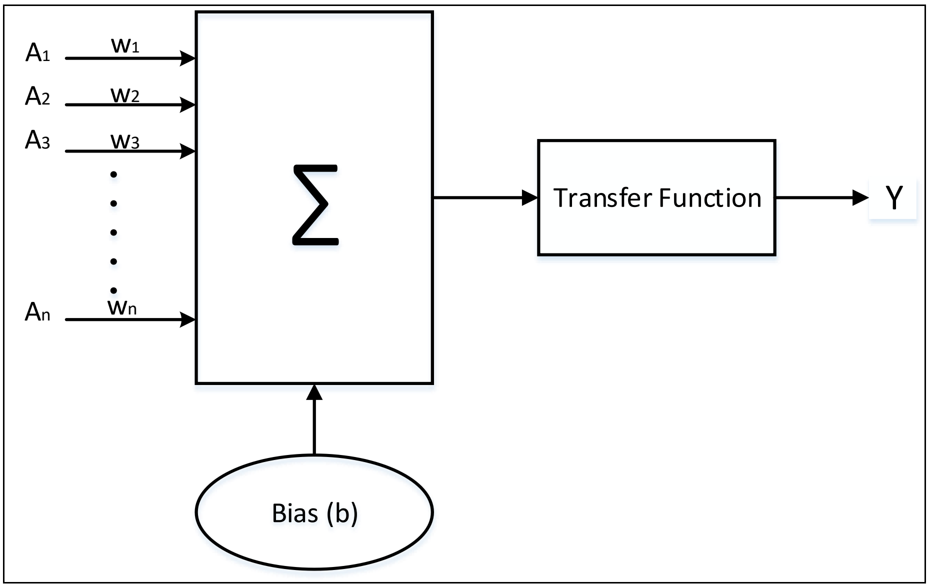


Figure 4.6 Structure of an ANN neuron

A multi-layer perceptron (MLP), the most popular supervised ANN model, is used in this work for water quality rating calculation. It consists of three layers: the input layer, hidden layer and output layer. All three layers have the computational units called nodes which mimics human biological neuron. All the nodes of a layer are connected to the adjoining layer but not connected to each other. There is no feedback connection in the entire architecture, as seen in Figure 4.7. Hence it is known as feed-forward MLP (FF-MLP) architecture.

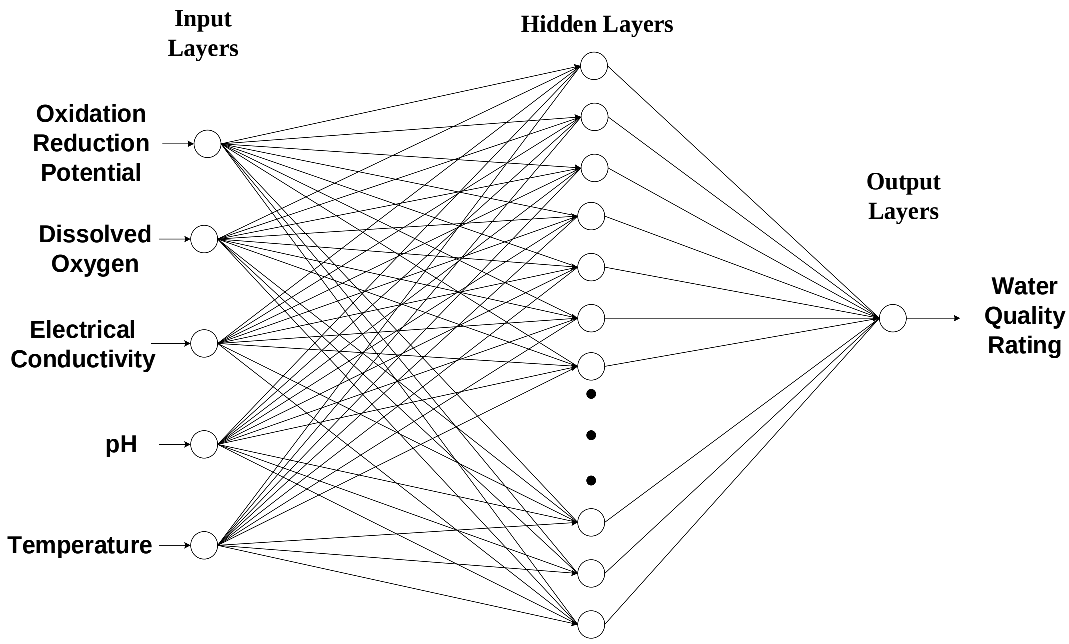


Figure 4.7 3-layered feed-forward MLP architecture

The input layer corresponds to the different water quality parameters (i.e., independent variables), and the output layer corresponds to the water quality rating (WQR) (i.e., dependent variable). Several activation functions have been used in the literature for MLP architecture, such as linear, logistic, tanh, threshold, and Gaussian. The logistic activation function ($f(x) = 1/(1 + e^{-x})$) has been used in this work. In the hidden layer, if the number of nodes is too less, the training of the network will not be proper. And if the number of nodes is too high, the ANN model will be complex. Hence, the mean square error (MSE) or quadratic or L^2 loss function was used for optimization given by Eq. (4.19).

$$MSE = \frac{\sum_{i=0}^{n}(y_i^t - y_i^p)^2}{n} \tag{4.19}$$

Where y_i^t is the target variable and y_i^p is the predicted variable. The number of nodes was selected based on the MSE loss function for training & validation and found to be 30. The same can be observed in Figure 4.8 that with 30 hidden nodes, we are getting minimum MSE-T (Mean Square Error for Training) and MSE-CV (Mean Square Error for Cross Validation). The overall architecture of MLP consists of 5-input layer nodes (water quality parameters, i.e., independent variables), 30-hidden layer nodes (logistic activation function), and 1-output node (water quality rating, i.e., dependent variable).

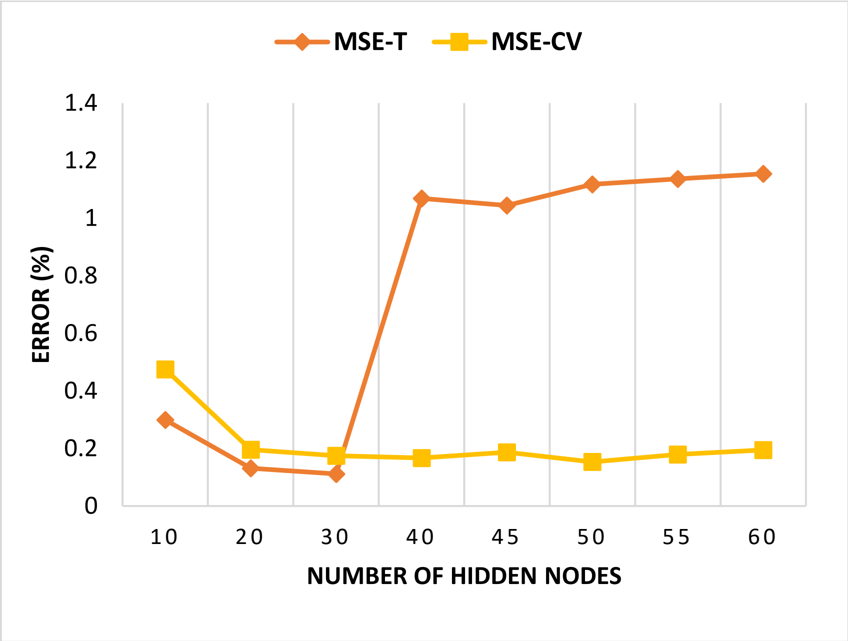


Figure 4.8 Hidden nodes vs. MSE-T and MSE-CV

4.4.2.1. Training and analysis of the Proposed Architecture

The dataset was developed according to the water quality parameter range and their water quality rating, as shown in Table 4.8. The same dataset was used for training in Python using the scikit-learn library. The dataset was divided into a 70:15:15 ratio for training, testing, and validation. The user can choose the range according to the geological conditions and the local water quality. The training and analysis procedure is explained in Figure 4.9. The generated dataset, based on the range of the water quality parameters, was imported and applied as input to the proposed ANN model. After the modeling, 15% of data was used for validation, followed by 15% data for testing of unknown samples. The programming for the presented workflow is done in Python.

Table 4.8 Quality, rating, and range for water quality parameters

Water Quality Parameters						**Water Quality**	**Class**	**Remarks**
E. Coli.	**Temp.**	**DO**	**pH**	**ORP**	**EC**			
Within Desired Limit (4-30)	(20-25)	(10-15)	(6.5-7.5)	(150-400)	(0-250)	Excellent	I	Proceed for Indexing
	(15-20) & (25-30)	(8-10)	(7.5-8.5)	(400-600)	(250-500)	Good	II	
	(10-15) & (30-35)	(4-8)	(6-6.5) & (8.5-9)	(600-800)	(500-1000)	Marginal	III	
	(5-10) & (35-50)	(2-4)	(4-6) & (9-11)	(0-150) & (800-1000)	(1000-2000)	Poor	IV	
	(>50) & (<5)	(0-2)	(0-4) & (11-14)	(>1000)	(>2000)	Unstable	V	
Outside desired limit	-NA-	-NA-	-NA-	-NA-	-NA-	-NA-	-NA-	No need for Indexing

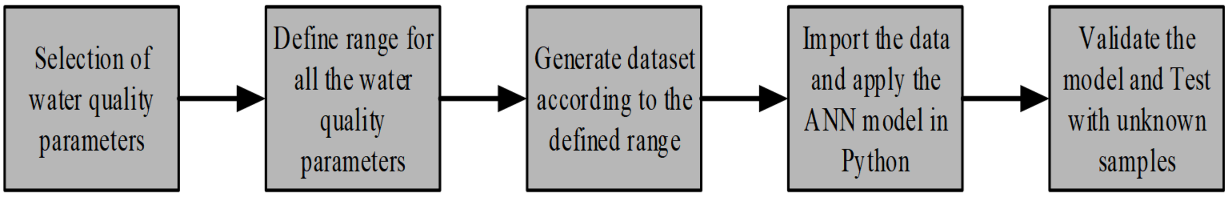


Figure 4.9 Data training and analysis procedure

For the FF-MLP architecture, the dataset was generated for 1800 samples. The ratio of training, validation, and testing was kept at 70:15:15. The proposed ANN model was trained for 1260 samples with the defined water quality rating, and 270 samples were used for each validation and testing. The graph for training data vs. targeted water quality parameters is shown in Figure 4.10, and the plot for testing is shown in Figure 4.11. The R^2 of training and testing is 0.955 and 0.933, respectively, proving ANN to be beneficial for water quality indexing.

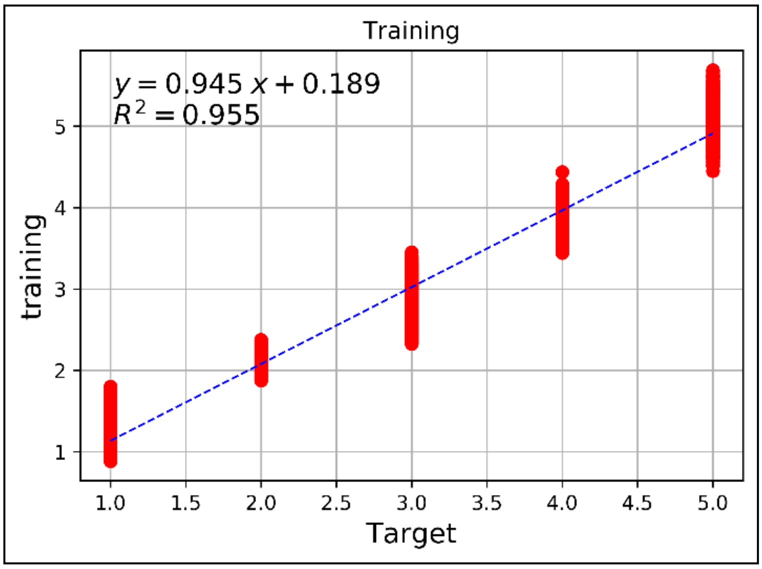


Figure 4.10 Training vs. target regression plot

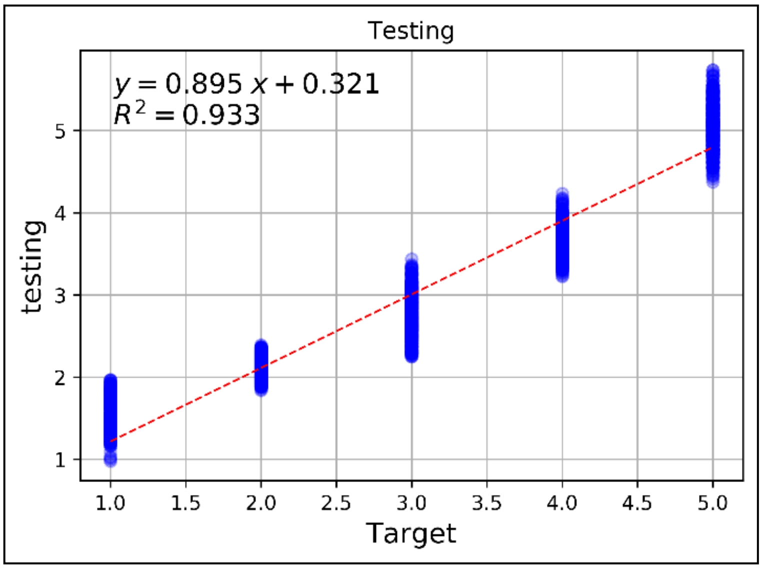


Figure 4.11 Predicted vs. target regression plot

The overall performance of the model is presented by the ROC curve, as shown in Figure 4.12. The ROC (receiver operating characteristics curve) is used to define the model sensitivity, which contains two parameters, namely true positive rate (TPR) and false positive rate (FPR). The TPR and FPR can be defined by Eqs. (4.19) and (4.20).

$$TPR = \frac{TP}{TP+FN} \tag{4.19}$$

$$FPR = \frac{FP}{FP+TN} \tag{4.20}$$

Where TP is the true positive, FN is the false negative, FP is the false positive, and TN is the true negative.

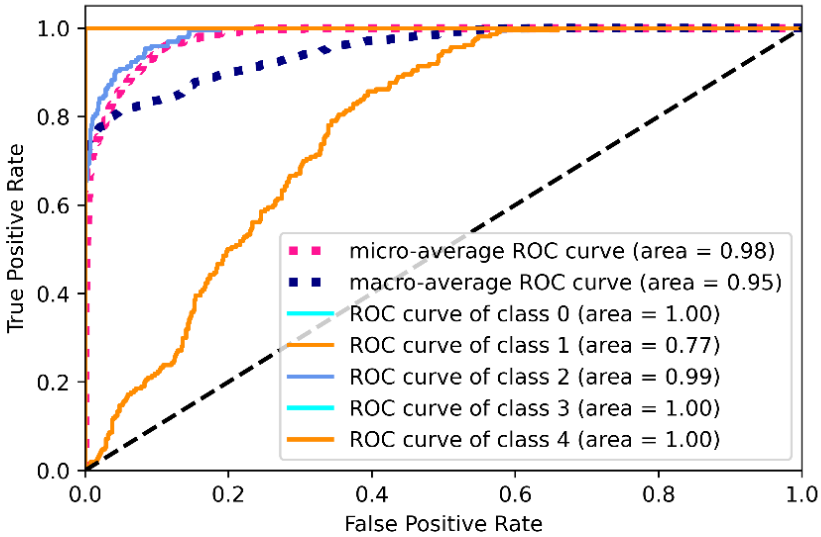


Figure 4.12 ROC for developed ANN model

The suitability of the proposed ANN model for water quality monitoring has been studied in this section. The analysis shows decent results for the training and testing of the proposed ANN model and it can be stated that the proposed ANN model can be quite useful for water quality monitoring in smart water grid development. The alarm or early warning system for any leakage or contamination in the distribution network can be set up based on the water quality rating.

4.5. Summary

The methodologies for data analysis to define overall water quality have been proposed in this chapter. The proposed methodologies were based on indexing tool, statistical modeling, and Artificial Neural Networks (ANN). The water quality was also defined into a single term in the range of 0 to 100, employing the aggregate mean method. Based on the WQI value, it will be easier for the decision-makers to categorize the water quality and take the necessary action. The fuzzy modeling was used to determine the water quality in different categories, such as ‘Excellent’, ‘Good’, ‘Satisfactory’, ‘Poor’, and ‘Bad’ for easily understandable its uses to a layman. The fuzzy modeling employed in this work is an efficient way to predict the water quality in real-time compared to manual approaches. PLSR technique and FF-MLP architecture were also used to define the water quality in different classes. A synthetic dataset was generated to develop the proposed model, and the real-time samples were fed to the developed model, and water quality was defined. The dataset was divided into the ratio of 70:15:15 for training, testing, and validation. Currently, we have not added the E. Coli. in modeling, as mentioned in the category ‘C’ of CPCB water quality standards. No literature has been identified reporting E. Coli. in Rajasthan province, as the study area has dry weather conditions where the chance of growth of E. Coli. is very less. E. Coli. is present only where the storage container is not appropriately cleaned or old distribution pipeline or the pipeline leakage or bad sanitation condition [109], [110].

Although sensor technology has achieved the manufacturing of low-cost and portable water quality sensors, the sensors face drift sooner or later after installation. The drift may occur due to sensor aging, temperature & humidity variation, poisoning among the sensor array, or due to a combination of all. This sensor drift will demolish the calibration model of any instrument. This issue can be solved by calibrating the sensors, which is also a challenge for field-deployable instruments. An alternate solution is provided for the drift compensation based on ANN modeling in chapter 5.

Chapter 5

Drift Analysis and Compensation of Commercial Water Quality Sensors

Preamble

Although sensor technology has achieved the manufacturing of low-cost and portable sensors, there are specific issues in the multi-sensor systems used for water quality monitoring, which prevents these systems from the routine measurement of water samples. An important issue is drift; related to sensor readings, which may refute the calibration of sensors leading to the necessity of frequent recalibration of the sensors that required effort as well as shut down the system. An alternative approach for drift correction is based on the mathematical correction method. In this chapter, a regression calibration method employing an Artificial Neural Network (ANN) is devised. A feed-forward ANN-based regression model has been used to extend the calibration lifetime of sensors. The evaluation of the model was performed based on the Root Mean Square Error (RMSE) and the RMSE for cross-validation (RMSE-CV). The proposed model is also compared with the traditional statistical method and proved superior. The experimental results demonstrate the best performance with a negligible error rate. Based on the results of the current study, ANN appears to be more adaptive for data analysis in environmental monitoring applications. The introduction to drift, its causes, and compensation methods have already been discussed in chapter 2. In this chapter, only the adopted methodology and results will be discussed in detail.

5.1. Reference Solutions

The standard reference solutions used for the measurement were pH solution (4 pH, 7 pH, and 10 pH), Electrical Conductivity (EC) solution (1000 µS/cm), Dissolved Oxygen (DO) solution (0 mg/l), and Oxidation Reduction Potential (ORP) solution (225 mV). These solutions were purchased along with the water quality sensors from ATLAS Scientific, USA [111]. All the reference solutions used in the proposed work were of analytical grade and non-toxic. The reference solutions used in this work are presented in Appendix A.

5.2. Sensors and Their Measurement Procedure

A total of four sensors were used for the record of measurement in this work: pH sensor, EC sensor, DO sensor, and the ORP sensor. The pH and ORP sensors are the electrochemical sensors; the DO sensor is membrane-based, and the EC sensor is electrode-based. All these sensors were calibrated with reference solutions (mentioned earlier) in the laboratory environment to avoid uncertainty in measurement. A 2-point and 3-point calibration were performed for the conductivity and pH sensor, respectively. The ORP and DO sensors were calibrated using 1-point calibration. The calibration procedure is explained in [148].

A summary of water quality sensors, their sensor technology, recalibration time, and reference solution has been presented in Table 5.1. The recalibration time for pH, DO and ORP sensor depends on the uses (there is no fix time for recalibration), whereas the conductivity sensor does not need the recalibration once it is properly calibrated with reference solution as claimed by the manufacturer.

Prior to measurement, the sample temperature was required to maintain at 25°C to prevent the deviation in readings due to temperature variation. During the experimental period, a total of 20 sessions for measurement were performed within 120 measurement days in the laboratory conditions. The experiment was uniformly distributed for a period of four months.

Table 5.1 Water quality sensors, their sensor technology, recalibration time and reference solutions

Sensor	Sensor Technology	Recalibration Time	Reference
pH Sensor	Reference Electrode (Ag/AgCl)	There is no set schedule for recalibration. Depends on different uses.	7 moles/L
ORP Sensor	Glass Electrode (Reference Solution not Specified)	There is no set schedule for recalibration. Depends on different uses.	225 mV
Conductivity Sensor	Two Probe Measurement Technique (Graphite Plates)	Conductivity probes work by measuring the electrical current of the water between two graphite plates. The plates do not go bad or change, so recalibration is not necessary. After the first calibration, conductivity probe is good to go.	1000 µS/cm
DO sensor	Membrane-PTFE (Polytetrafluoroethylene)	There is no set schedule for recalibration. Depends on different uses.	0 mg/L

5.3. Spyder Platform-The Python Development Environment

Spyder is a scientific development environment written in Python and designed for data analysts, engineers, and scientists. It is an open-source platform with inbuilt libraries that supports object-oriented dynamic programming and graphical programming as well. It has many in-built features, such as editing, analysis, debugging, data exploration, deep inspection, and data visualization as well. The environment is integrated with many scientific packages, including NumPy, SciPy, Matplotlib, SymPy and many more [149] so, these libraries need not be installed separately. This python environment is used for the ANN implementation in this work. The Spyder platform is shown in Figure 5.1. It has an editor window with syntax highlighting, introspection, and code completion. In the help window, the user can get detailed information about any command and function at the time of programming.

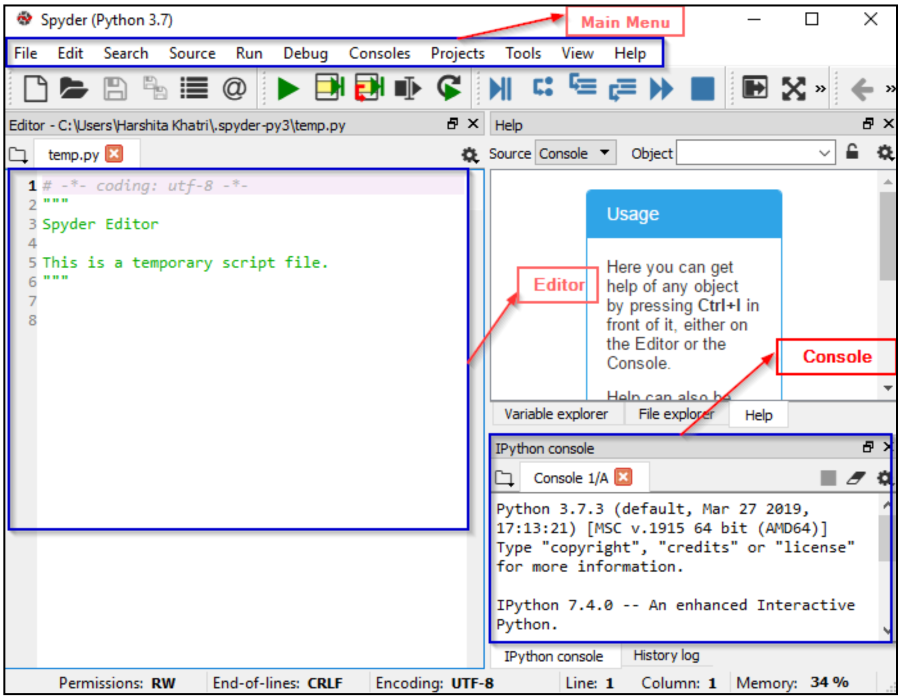


Figure 5.1 Spyder platform

5.4. Proposed Artificial Neural Network (ANN) Architecture

A fully feed-forward artificial neural network (FF-ANN) architecture with 3 layers (input, output, and hidden) has been used in this study, as shown in Figure 5.2. It consists of three consecutive layers: the input layer, output layer, and hidden layer. All three layers consist of computational nodes, which are represented by the circle. The input layer corresponds to the independent variables, which in this case, are drifted and targeted water quality parameters. The output layer corresponds to the corrected water quality parameters. Several activation functions can be used in the ANN architecture, such as linear, Gaussian, logistic, tanh, and threshold. In this work, three different activation functions were used in the ANN architecture given below.

ReLu function: $$f(x) = \begin{cases} 0 & for \quad x < 0 \\ x & for \quad x \geq 0 \end{cases} \tag{5.1}$$

Logistic function: $$f(x) = \frac{1}{1+e^{-x}} \tag{5.2}$$

tanH function: $$f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (5.3)$$

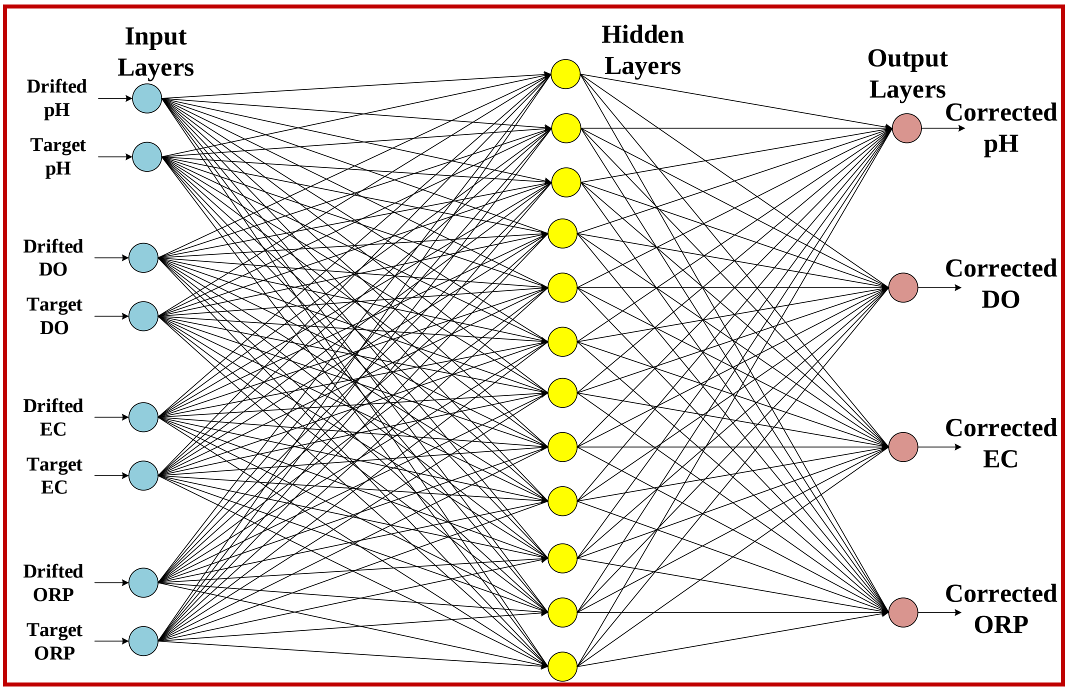


Figure 5.2 An 8-input (drifted pH, DO, EC, ORP, and targeted pH, DO, EC, ORP) and 4-output (corrected pH, DO, EC, ORP) 3-layer feed-forward ANN architecture

An optimization strategy was used to fix the number of hidden layer nodes, and it was found that for ReLu and tanH neurons, 45 nodes were giving the best results. Whereas, for logistic neurons, the number of hidden layer nodes could not be decided. The reason behind this will be discussed in the next section. The overall architecture has 8-input layer nodes (drifted and targeted water quality parameters, i.e., independent variables), 45-hidden layer nodes, and 4-output layer nodes (corrected water quality parameters, i.e., dependent variables).

5.5. Data Pre-processing

Data Preprocessing is a step in any Machine Learning process in which the data is transformed, or encoded, to get it to such a state that the algorithm can easily interpret the features of the data. In the sensor fusion method, the measured values from the sensors are on different scales. So, these measured parameters should be brought on the same scale for further analysis of data. It is crucial to identify the suitable preprocessing technique so that the desired results are met. There are different preprocessing techniques available as discussed below [150].

5.5.1. Standardization

Data standardization is the process of converting any distributed data into a normal distribution. In data standardization, the mean is subtracted from the data, and then the result is divided by standard deviation. The normal distribution will always have '0' as mean and '1' as standard deviation. Assume a data set $X = \{x_1, x_2, x_3, \cdots, x_n\}$ with standard normal variate (z) defined by

$$z = \frac{(x-\bar{x})}{\sigma} \tag{5.4}$$

Where z represents the area covered by the distribution curve, positive values of z denote that the value is the right side of the mean and vice versa.

5.5.2. Normalization

In normalization, all the variables are converted to the same scale without changing the data pattern or losing any information. Min-max scaling is the basic method, in which all the values are converted between -1 and 1 {-1, 1} or 0 and 1 {0, 1}. If $X = \{x_1, x_2, x_3, \cdots, x_n\}$ is the data set, then the normalized value X_n can be represented by

$$X_n = \frac{X-X_{min}}{X_{max}-X_{min}} \tag{5.5}$$

Where X_n is the normalized value, X_{min} is the minimum value from the column of dataset and X_{max} is the maximum value.

5.5.3. Multiplicative Scattering Correction (MSC)

MSC is used in the spectroscopic data analysis for both offset variation correction (baseline correction) and scaling. The variation is common and is caused by the light scattering in the measurement sample. An additive and a multiplicative term are used in the equation defining the scattering contributions in addition to the spectral signal for a single spectrum.

$$a = x + y\hat{a} + e \tag{5.6}$$

Here, $\hat{a}$ is the reference spectrum, and x & y are the regression coefficients obtained from fitting the measurement sample spectrum (a) with the reference ($\hat{a}$). Now the corrected spectrum is defined by the following equation.

$$a_{msc} = \frac{(a-x)}{y} \tag{5.7}$$

5.5.4. Data Smoothing

The data smoothing technique is used for revealing patterns from data, removing outliers, and reducing noise. It also improves the Signal-to-Noise Ratio (SNR). The most popular technique for smoothing is moving average. This can be explained by a simple example of five points moving average filter. It takes the current value and four previous values and takes the average, then replaces the current value with the average value. The final smoothed data point is given by

$$Y_m = \frac{\sum_{i=-n}^{n} X_i Y_m}{\sum_{i=-n}^{n} X_i} \tag{5.8}$$

In this work, Normalization is done to change the values of all columns in the dataset to a common scale without changing the differences in the range [151]. In our case, since all the water quality parameters are on a different scale, normalization is used to bring all the water quality parameters on the same scale. As all of the data is positive, we can normalize it between {0, 1}. The min-max scaling is used to normalize the data in the range of 0 to 1. The normalization in Python has been done using the sklearn MinMaxScaler library.

5.6. Training and Analysis Procedure

The simulation was done in the Spyder platform [152] using a scikit-learn (sklearn) library [143]. The sklearn library is built on NumPy, SciPy and Matplotlib. It is used in classification, regression, clustering, dimensionality reduction, model selection and preprocessing. The data set was generated using the recorded readings for every sensor during measurement days. The complete data was divided into training, testing, and validation in the ratio of 70:15:15. The training was performed up to a maximum of

2000 epochs. The learning rate was fixed at 0.0001. During the testing of 15% data, the trained model was used for prediction. The reason for analyzing the test set was to ensure whether the training model could be used for later analysis of the samples. For the remaining 15% data, 10-fold cross-validation was performed in the validation procedure in order to see that the model was working fine.

5.7. Traditional Statistical Method

The drift correction has also been attempted using a statistical method for testing and comparison. The principal component analysis based drift correction (PCA-DC) has been used in this study to compare with the proposed ANN model. If the sensors show significant drift in reference solutions, the first principal component of PCA of reference describes the drift direction [153]. The loading vector P_r of reference PCA model is used to determine the projection T_r of the new measuement samples X_{new}. Now, subtracting the drift component from new measurement samples will give the new measurement matrix as given in Eq. (5.9). The scaling and transformation should be the same for both the reference and the measurement samples.

$$X_{new}^{corrected} = X_{new} - T_r P_r' \tag{5.9}$$

5.8. Performance Evaluation of the Proposed ANN Model

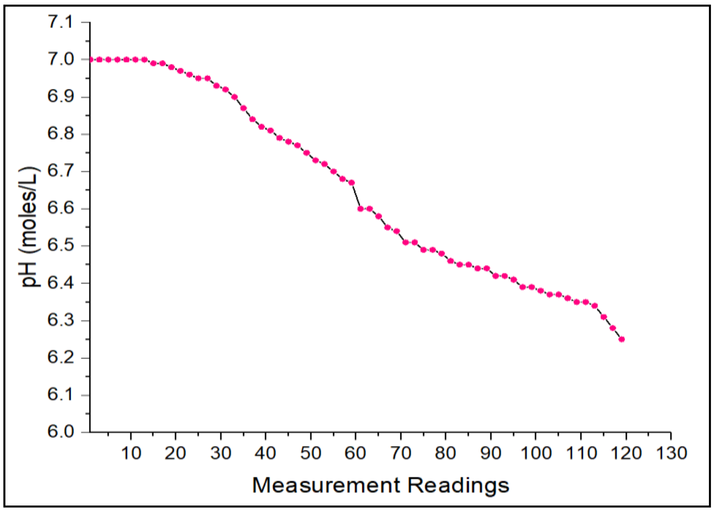
To evaluate the performance of proposed ANN model, root mean square error (RMSE) was calculated using Eq. (5.10) [154]. RMSE is the standard deviation that is used to measure the difference between the predicted and the observed values.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_p - y_o)^2}{n}} \tag{5.10}$$

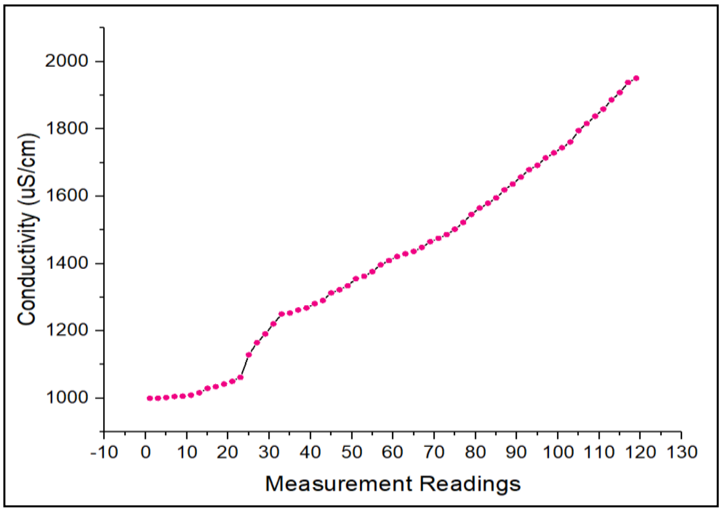
Where y_p is the predicted value and y_o is the observed one. The Root Mean Square Error for Cross-Validation (RMSECV) was also calculated for every activation function. The RMSECV was used to construct the proposed ANN model, and RMSE was used to test the model against new data that the model has not seen.

5.9. Results and Discussion

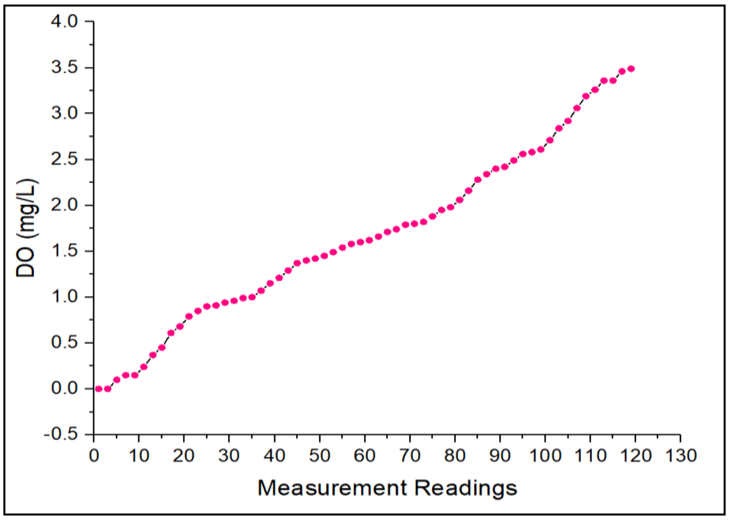
The development of an experimental test-bed setup has already been discussed in detail in chapter 3. The sensors used were pH, conductivity, dissolved oxygen, and oxidation-reduction potential sensors. The responses of these sensors for their standard solutions were recorded over 120 days. The deviation of the sensors in standard solution between the experimental study was observed that can lead to the biased measurement for target samples, thus resulting in unreliable readings. The measurement records for each sensor are shown in Figure 5.3. It can be observed from the measurement record that for pH, DO, and EC sensors, there is more deviation in sensor readings, whereas in the ORP sensor, the deviation is less as compared to others.



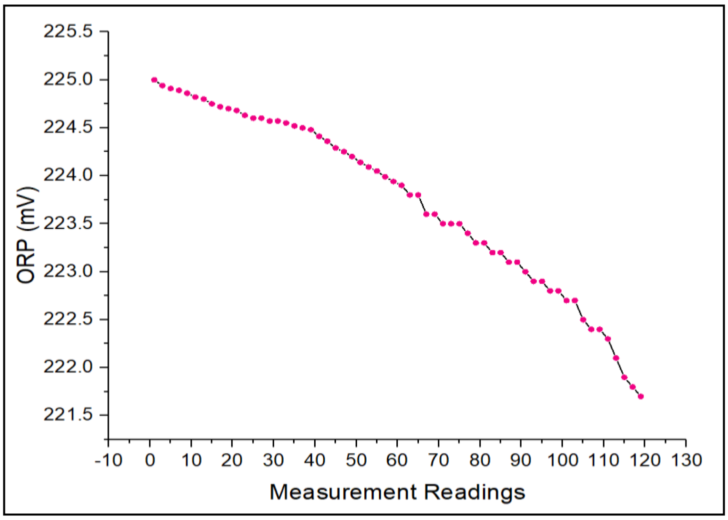
(a) pH Readings



(b) EC Reading



(c) DO Readings



(d) ORP Readings

Figure 5.3 Drift readings for different water quality sensors

5.9.1. Effect of drift variation on water quality

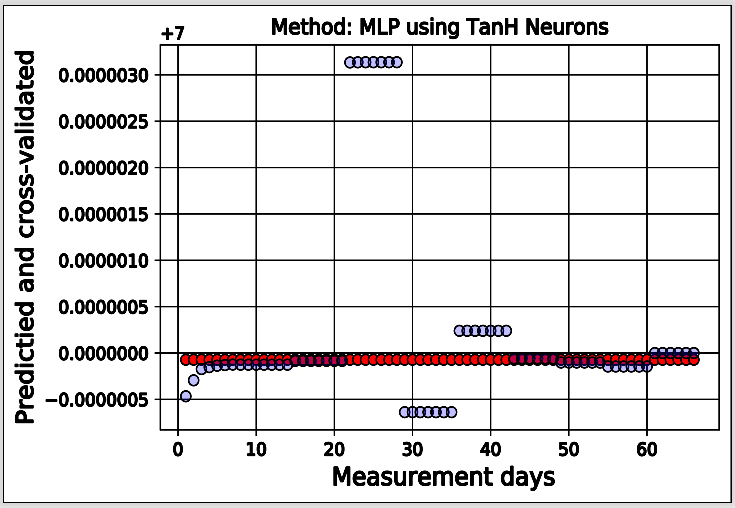
As discussed in the chapter 4, the water quality has been defined by statistical and soft computing methods. We will take an example of the fuzzy water quality analysis and see how the drift in the sensors may affects the overall water quality. The water quality parameters have been acquired from the sensors without any drift correction or calibration as well as from the reference instrument, which was regularly calibrated. A confusion matrix is presented in Table 5.2 showing the overall water quality of the samples with timely calibration (reference instrument) and without calibration (the sensors used in this work). It can be observed from the table that the some of the samples have been misclassified as compared to the true water quality class, which may refute any developed model for precisely predicting the water quality. Thus, it is essential to compensate the sensor deviation over time and the same has been proposed in this chapter.

Table 5.2 Confusion matrix for overall water quality of different samples without drift compensation

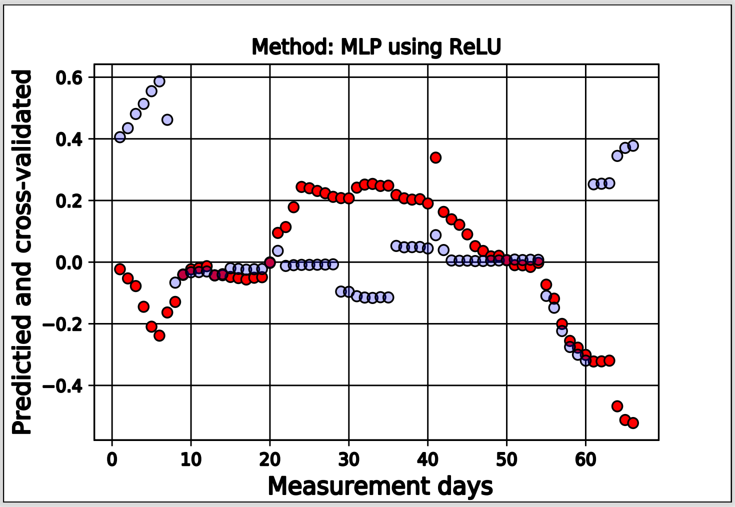
True Water Quality \ Predicted Water Quality	Excellent	Good	Satisfactory	Poor
Excellent				
Good	2	13		
Satisfactory		2		
Poor			2	1

5.9.2. Drift compensation of water quality sensors

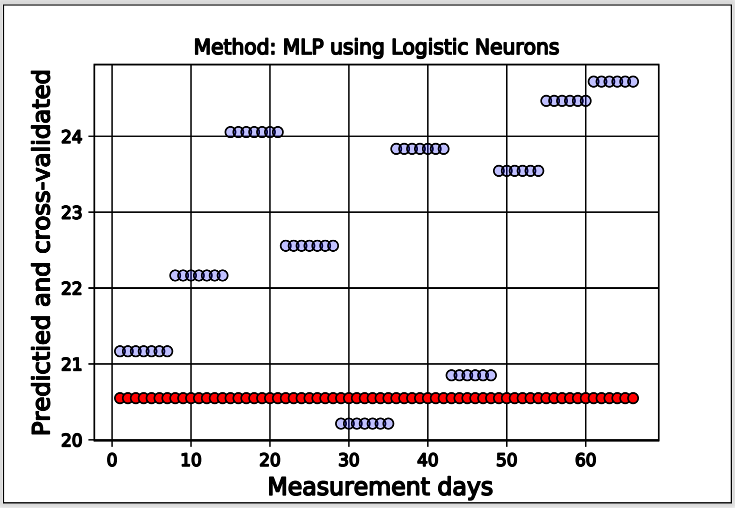
In order to correct the deviation, we have applied the ANN method in Spyder, as described in section 5.3. We simulated the Python program for different activation functions, as discussed earlier, and predicted and cross-validated results were plotted with respect to measurement days. Some of the plots of predicted and cross-validated values for different water quality sensors are shown in Figure 5.4. The red and blue points correspond to the predicted and cross-validated values, respectively. The X-axis represents the measurement days, and the Y-axis corresponds to the values for different water quality sensors.



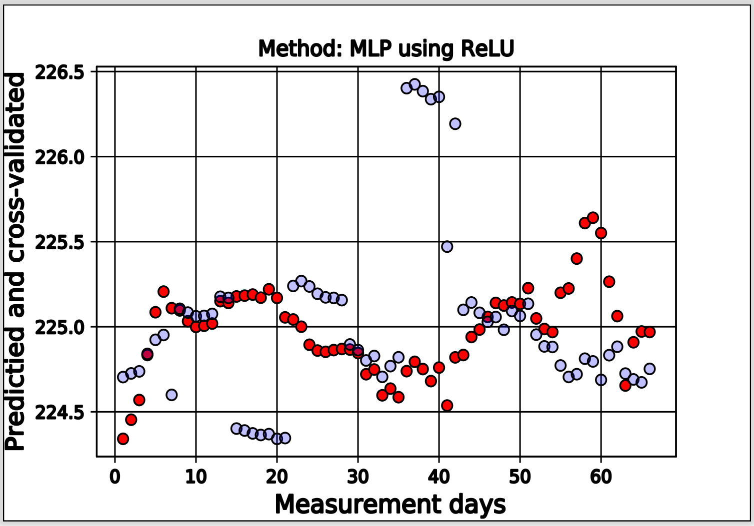
(a) pH values for tanh activation function



(b) DO values for ReLu activation function



(c) EC values for logistic activation function



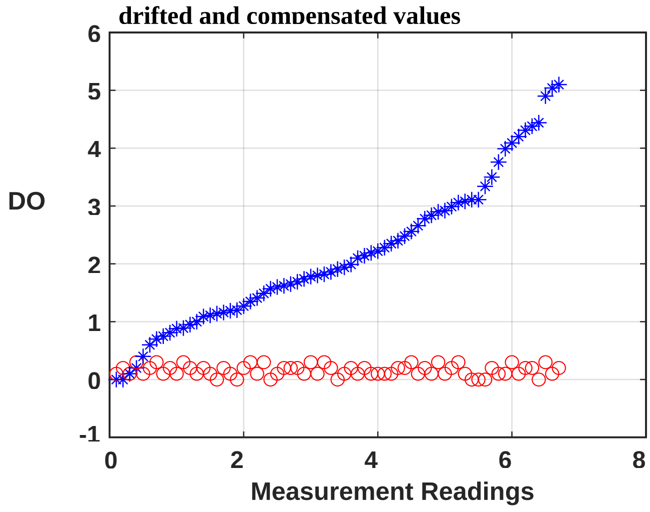
(d) ORP values for ReLu activation function

Figure 5.4 Predicted and cross-validated values of water quality sensors using different activation functions

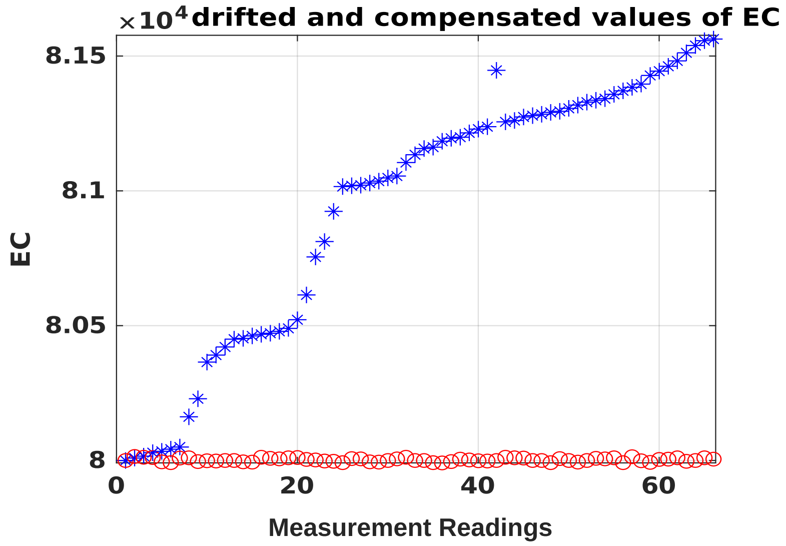
For ReLu and tanH functions, we are getting the best results with 45 hidden nodes, whereas for the logistic function, the RMSE is too high. The minimum RMSE and RMSECV are 0.00 and 0.00, and the maximum RMSE and RMSECV are 416.87 and 820.05, respectively. It is observed from the plot that the drift compensation has been achieved in a quite efficient manner using different activation functions except for the logistic function. For the logistic function, the simulation was performed using every combination of the number of hidden layer nodes and the number of iterations. But the optimization does not converge in any case.

5.9.2. Comparison of the proposed model with a statistical method

The PCA-based drift correction (PCA-DC) has been implemented in this study, as discussed in section 3.4. Some of the plots of drifted and compensated values for water quality sensors are shown in Figure 5.5. The blue and red points correspond to the drifted and compensated values for water quality sensors, respectively. Table 5.1 shows the RMSE and RMSECV calculated for both the proposed ANN model and the PCA-based drift correction. The PCA-DC method has the maximum and minimum RMSE of 2.3683 and 0.4750 for DO and pH sensors. It can be observed from the Table 5.3 that the proposed ANN model is superior to the PCA-DC method.



(a) DO Sensor



(b) EC Sensor

Figure 5.5 PCA based drift correction of water quality sensors

Table 5.3 RMSE and RMSECV for different activation functions

Water Quality Sensors	Proposed Method (FF-ANN)						PCA-DC Method	
	ReLu		*Logistic*		*tanH*			
	RMSE	***RMSECV***	***RMSE***	***RMSECV***	***RMSE***	***RMSECV***	***RMSE***	***RMSECV***
pH	1.3101	0.2386	411.18	0.0014	0.00	0.0003	0.4750	1.9087
DO	0.5781	0.3315	416.87	0.0001	0.00	0.00	2.3683	2.4289
EC	0.6924	2.5688	409.41	820.05	0.00	0.0242	1.3345	1.2058
ORP	0.2923	0.8333	410.47	77.66	0.00	0.00	2.0039	2.6014

5.10. Summary

The application of a multi-sensor system for water quality monitoring is problematic as the uncertainty increases due to sensor drift over time. This drift may refute established models and reduce the lifetime of model results in imprecise prediction in measurement. In this work, we have used an FF-ANN model for drift compensation of water quality sensors. The RMSE is achieved zero for activation function tanH for the proposed ANN model, which is far better than the PCA-DC method (0.4750%). It was evident that the ANN model was superior to the statistical method. As discussed earlier, the proposed ANN model can be implemented in our developed hardware setup at no additional cost. Our results showed that machine learning could be an alternative approach to traditional statistical methods for environmental monitoring applications. Since the MSS is multivariate, ANN proved to be quite an efficient methodology for drift correction. This proposed work can prevent frequent calibration of the sensors and increase the calibration lifetime.

Chapter 6

Water Quality Monitoring in Distribution Networks

Preamble

There are many challenges while developing a smart or sustainable city, such as air/water quality monitoring, water resource management, power grid implementation, and transport management. Water quality monitoring is one of them in which many researchers and scientists showed interest. The current distribution systems always face leakage, failure, illegal connections, delay in maintenance. The solution to this problem is the implementation of a smart water grid. A smart water grid can manage the water supply in the distribution systems by real-time monitoring of water quality, flow, pressure, and distribution network status. In this chapter, a real-time assessment of water quality is proposed in the smart water grid employing and machine learning algorithms. The proposed model can analyze various water quality parameters such as temperature, pH, dissolved solids, electrical conductivity (EC), salinity, turbidity, dissolved oxygen (DO) and oxidation reduction potential (ORP). The proposed architecture can log, analyze data and remotely monitor the data. The data obtained from various sensing nodes were uploaded to the cloud, a service provided by ThingSpeak®. Experimental results show that the proposed low-cost sensing network can be an ideal early warning system in smart cities.

6.1.Introduction

60% of the world's population will be living in cities by the end of 2050. With the growth in urbanization, many problems arise, jeopardizing the environmental sustainability of the cities. The rapid growth of urbanization also raises numerous challenges, such as water and air pollution, waste disposal, saturated transport, and more energy consumption, resulting in poor public health. These problems can be solved by implementing Information and Communication Technologies (ICT) [155], [156]. Ensuring water quality monitoring in the distribution network is a challenge due to frequent failures and pipeline leakages. The current distribution system consists of different components, such as a pump, pipeline network, and valves. The performance and reliability of these components decrease over time, and the distribution systems have a higher risk of pipeline leaks, failures and wastage of water. It is difficult to access the leakage and consumes time, which results in high wastage of water. To overcome these problems, scientists and researchers have introduced a smart water grid for the management of distribution systems. A smart water grid is capable of real-time as well as online water quality, flow, and pressure monitoring, failure detection, leakage detection in distribution systems. The smart water grid implementation can also manage water consumption and supply. A smart grid includes a wireless network to cover the entire distribution network, sensors for monitoring connected to the wireless network, smart meters to monitor, control, and automate the water supply. This enables the real-time status monitoring of distribution networks to locate leakages and timely maintenance. The management can be made easy with a smart grid platform to supply water 24/7 to the consumers. A smart water grid is an integration of various sensing and communication technologies (SCT), which are driven by the Internet of Things (IoT), cloud computing, and big data analysis [157].

In this work, an attempt has been made to form a smart water grid employing a sensing platform, a Wireless Sensor Network (WSN), and an Internet of Things (IoT) to monitor the water quality in the distribution network. Cyber-Physical Systems are a network of interconnecting individual elements that fulfill sensing, computing, monitoring, multiple communication among modules, sensory output and data analytics. The integration of the above-mentioned individual modules of the smart grid forms a CPS structure. The water quality parameters to be monitored were chosen based on the criteria defined by the Central Pollution and Control Board (CPCB), India [14]. The parameters targeted in this work are temperature, pH, Electrical Conductivity (EC), Dissolved Oxygen (DO), Oxidation Reduction Potential (ORP) and E. Coli.

Two sensing nodes and one server node were developed for demonstration in the star network in this work. A server based on Raspberry Pi has been used as an IoT platform for updating the acquired data to the cloud. Water quality sensors were interfaced with the NodeMCU through the signal conditioning board at each sensing node. The water quality parameters were obtained from the developed setup. Based on the acquired parameter, the water quality is defined by employing an artificial neural network (ANN) into five categories based on their water quality rating.

In the experimental procedure, a tri-stage attempt was made. Different water quality parameters were acquired in the first stage, based on sensor readings interfaced with the NodeMCU Panel. In the second stage, the water quality rating was determined using the ANN model based on the results collected, and in the last stage, the data was sent to the ThingSpeak cloud platform and published.

The contributions of the presented work can be stated as follows:

- A network is formed to demonstrate the functioning of a smart water grid employing two sensing nodes and a server node.
- Different water quality parameters were acquired using the developed setup
- ANN was used to define the water quality rating based on acquired data from sensors
- Online monitoring is enabled using the ThingSpeak cloud platform

Different hardware modules, including water quality sensors, Zigbee, Arduino, NodeMCU, have already been discussed in chapter 3. Here, we will discuss the proposed distribution network architecture for wireless sensing and the ANN architecture for water quality monitoring, and the online monitoring of water quality parameters.

6.2.Proposed Hardware Architecture

A block diagram of the proposed distribution network is shown in Figure 6.1. The water from the water tank is supplied to the household through the distribution network's pumping station. In the network, nodes 2, 3, 4, and 5 are the sensing node for water quality parameter acquisition, and node 1 is the server node, which is located at the water supply station.

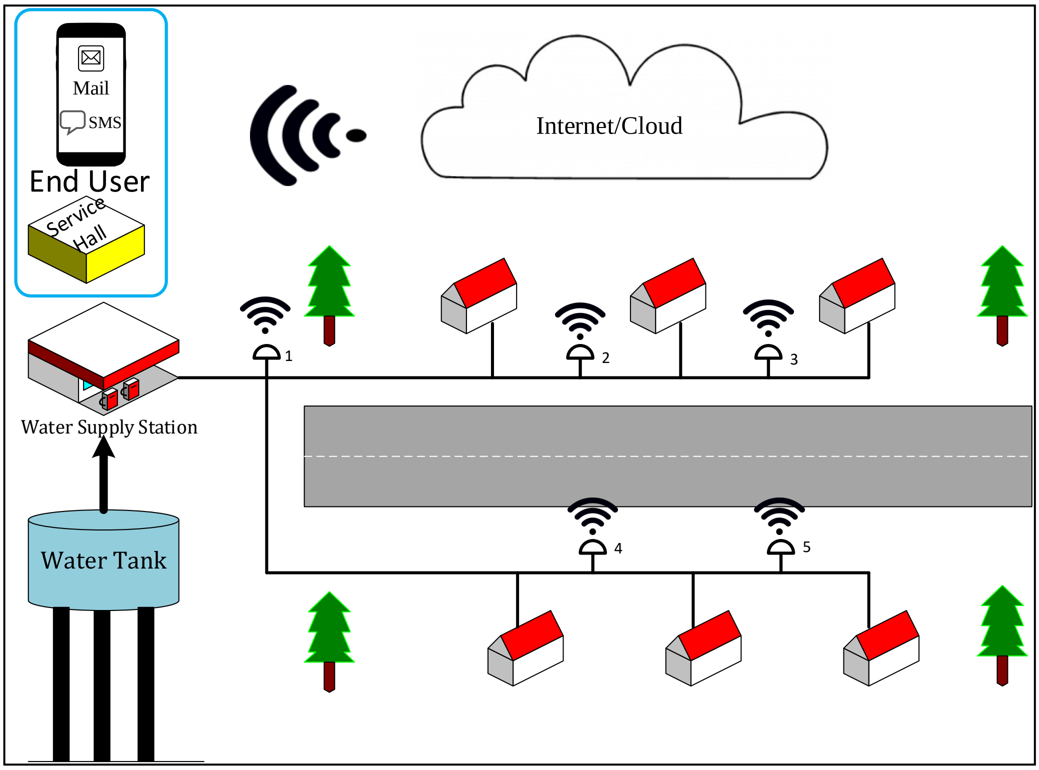


Figure 6.1 Proposed distribution network architecture

As discussed in the introduction, the water quality monitoring in distribution networks is attempted in 3-layered architecture as shown in Figure 6.2. The first layer, which is the physical layer (sensing nodes), comprises the NodeMCU, sensors and dedicated signal conditioning circuits. The NodeMCU development board is used as a core controller for the sensing node. The physical layer performs the water quality parameter acquisition. The communication between sensors and NodeMCU is done through the I2C protocol. The sensor array generates a data matrix that contains water quality parameters. The second layer is the edge computing layer or the server node, where Raspberry Pi setup is used for data acquisition and ANN model implementation to calculate the water quality rating. A separate database is also generated on the server node, where the authority can access the data of different nodes. In the third layer (Application layer), the water quality parameters are uploaded on the ThingSpeak cloud service. In the application layer, real-time node data can be accessed through the App, API request or terminal.

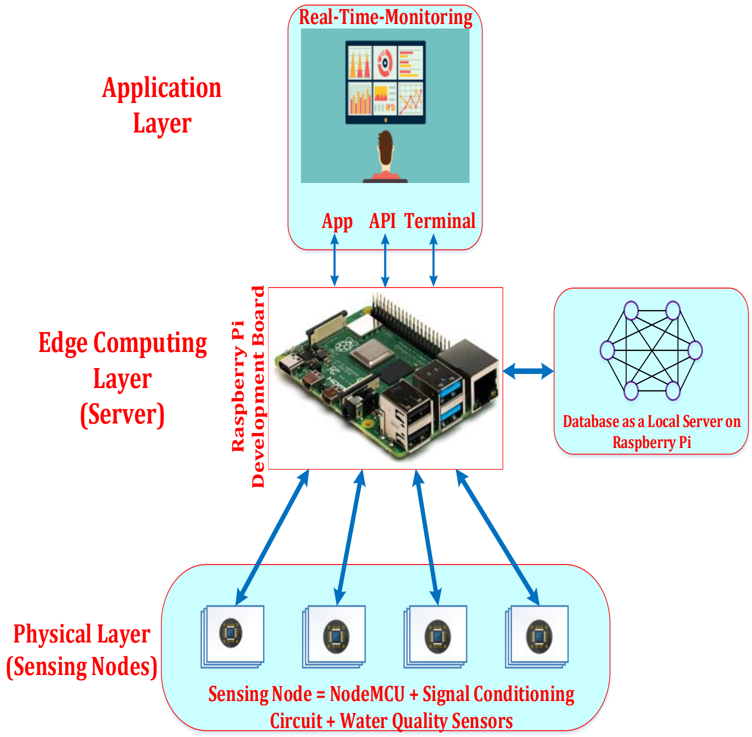


Figure 6.2 A 3-layered functional architecture of the proposed work

The sensing node block diagram and the experimental setup are shown in Figures 6.3 and 6.4. Five water quality sensors (temperature, EC, pH, ORP, and DO) were interfaced with the NodeMCU development board through a signal conditioning circuit (SCC). The signal conditioning circuit works as a mediator between the sensors and NodeMCU. The sensors and signal conditioning circuit require a 3.3 V, which is available on NodeMCU. Only the NodeMCU needs to power up with a 5 V DC adaptor.

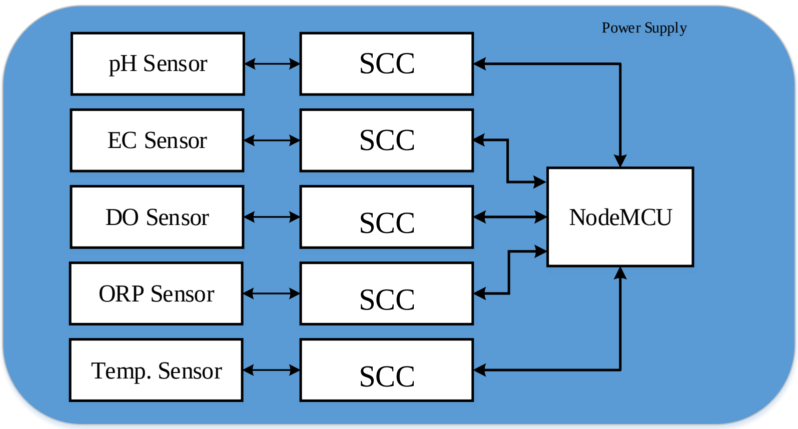


Figure 6.3 Sensing node block diagram

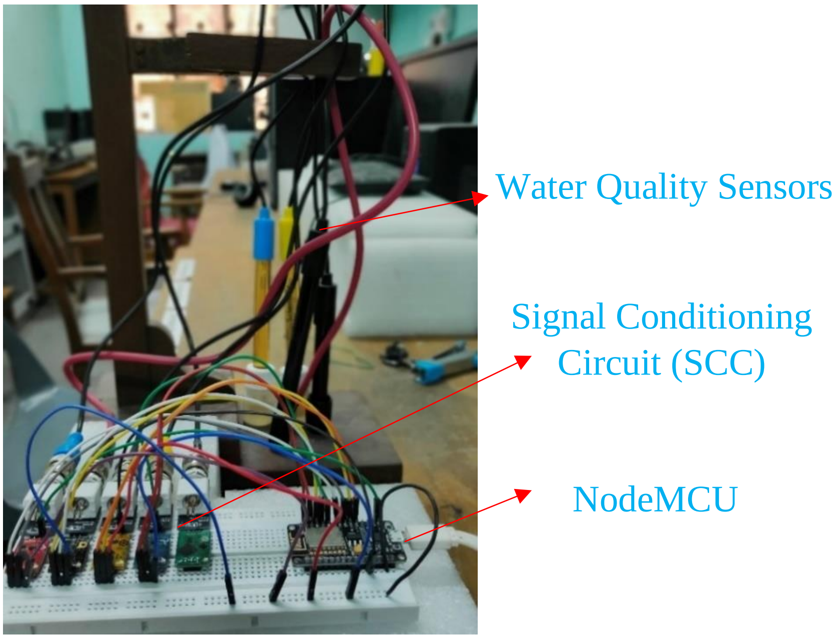


Figure 6.4 Sensing node experimental setup

The server node architecture and experimental setup are shown in Figures 6.5 and 6.6, respectively. A NodeMCU, keyboard, mouse, and 7 inch LCD touch screen from the Waveshare, were interfaced with the Raspberry Pi 3 development board at the server node. The server receives the water quality parameters as a data matrix, which is then used as input to the ANN model for water quality rating calculation. Although the developed sensing and server node setup was demonstrated in the laboratory environment using a 5V adaptor power supply, they are also tested using a power bank for field deployment and functioning correctly.

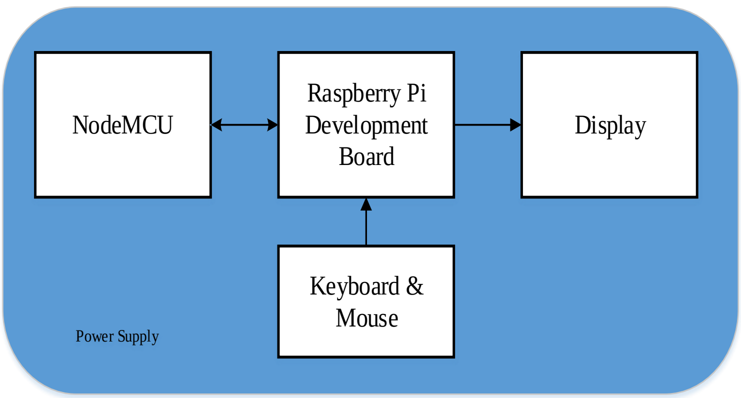


Figure 6.5 Server node block diagram

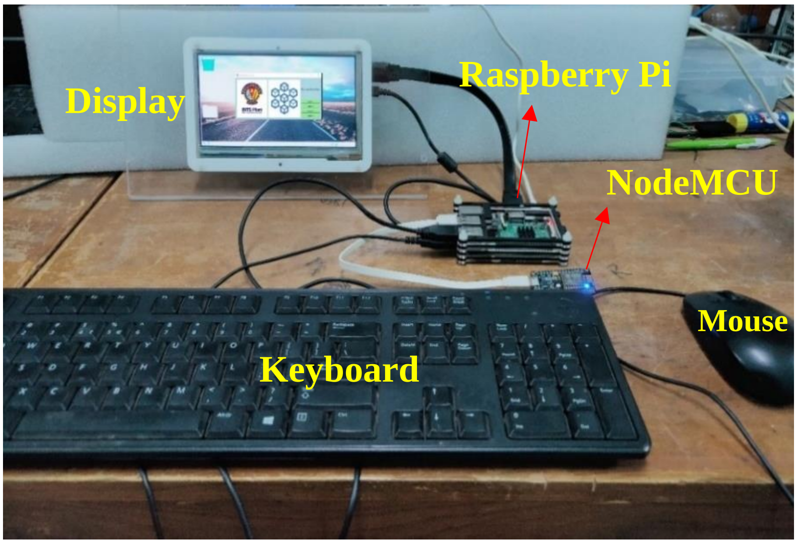


Figure 6.6 Server node experimental setup

6.3. Software framework

In addition to hardware development, a software framework for water quality parameter acquisition and further analysis has been designed. The programming for the sensing node is written in Arduino programming software (IDE) (which supports the NodeMCU programming) for data acquisition from the sensor array and to send the data over the wireless network. Python was used to program the Raspberry Pi for sensing nodes data acquisition and the implementation of the proposed neural network model with the help of the scikit-learn library [143].

6.4. Experimental procedure

The layered architecture shown in Figure 6.2 has been followed for the experimental procedure. Initially, the sensors were calibrated with the reference before measurement to avoid any uncertainty. The first step involves the water quality parameter measurement in the physical layer. The measurement was carried out after the sensor stabilized with minimum variation in readings. Later the acquired water quality parameters were sent to the server node and the ANN model proposed in chapter 4 was applied to define overall water quality based on the acquired data. The ANN model has already been discussed, thus not described here. The necessary steps for the experimental procedure are as follows:

- Wireless sensor network setup.
- Sensing node data acquisition and transfer it to the server node.
- Apply the ANN model on received data at the server node.
- Upload the data to cloud ‘ThingSpeak’ for real-time monitoring.

The data updating and real-time monitoring on the cloud server is shown in Figure 6.7. The acquired data was regularly updated and stored on the “Thingspeak” cloud by the server node. Thingspeak helps consumers archive the data, interpret the data, and analyze the website’s data.

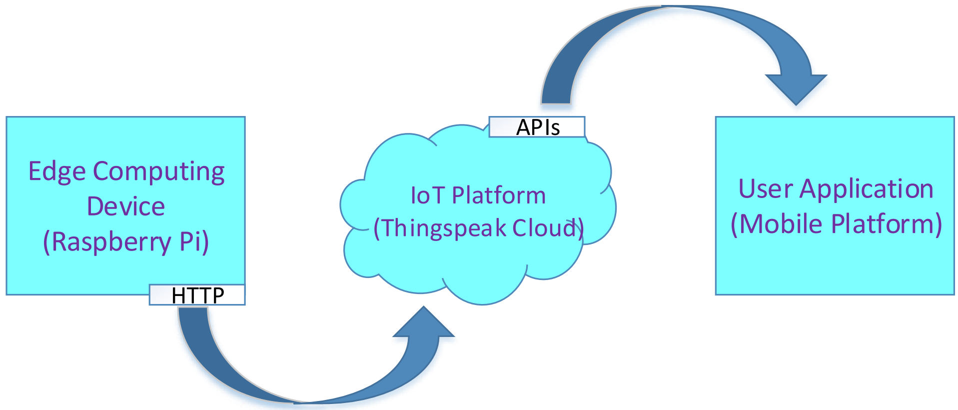
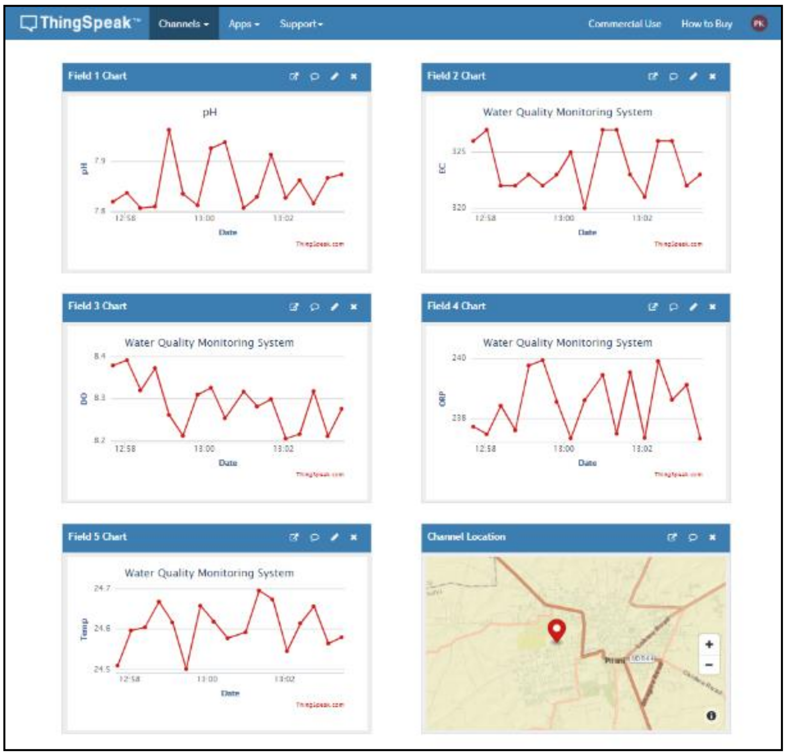


Figure 6.7 Data uploading and real time monitoring

A channel is created with the name ‘‘Water Quality Monitoring System’’ at the ThingSpeak server. Different data fields of the channel are named as different water quality parameters. The water quality parameters are updated on the ThingSpeak server by channel write API key and the ‘urllib2’ on Raspberry Pi. The parameters can be monitored from any digital device (either PC or Tablet, or Mobile). For Mobile, an apkfile (ThingView) needs to be installed, which is verified by MATLAB. After installing the ThingView, the water quality parameters can be monitored by accessing the channel by entering the channel ID and reading the API key of the channel. A screenshot of the IoT platform on PC, as well as mobile, is shown in Figure 6.8.



(a) Online parameters monitoring in PC

(b) Online parameters monitoring in Mobile

Figure 6.8 Online monitoring of water quality parameters

In addition to the data updating and monitoring on the cloud, a local server (without internet) has been set up on the edge computing layer where data from all the nodes have been saved on the Raspberry Pi as shown in Figure 6.2. The data from any node can be requested from the server at any time. If the node number and password match the saved password, the data for the requested node will be displayed; otherwise, an error message will be displayed. The procedure for server setup and requesting data is shown in Figure 6.9.

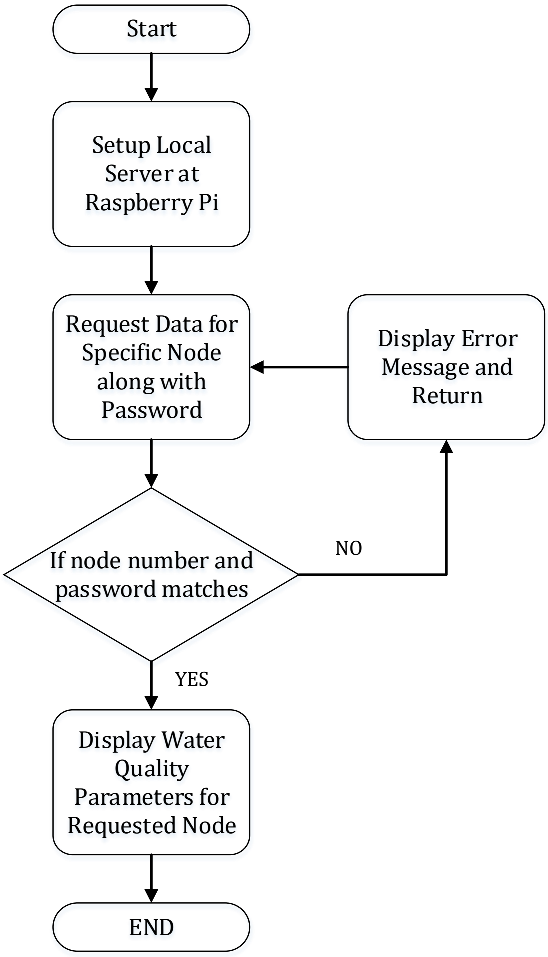
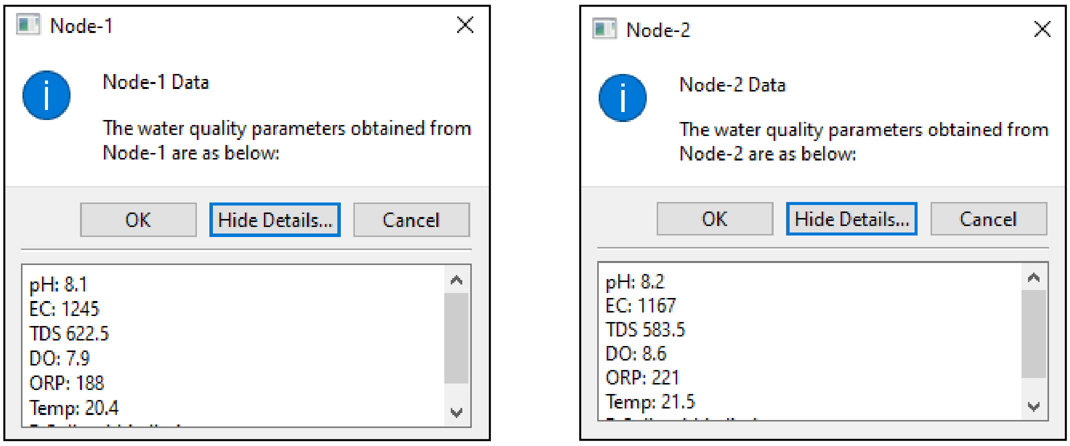


Figure 6.9 Local server set up and data request procedure

The user interface is developed on the application layer, as shown in Figure 6.10. The user can select the individual node and monitor the water quality parameter in a real-time environment. When the user clicks on a particular node, a pop-up window appears and displays the water quality parameters from the selected node. An example of a pop-up window is shown in Figure 6.11.

Figure 6.10 Server user interface



(a) (b)

Figure 6.11 Pop-up window GUI and water quality parameters for selected node

6.5.Discussion

The implementation and practicability of different modules of a smart water grid are investigated in this work. An example of a distribution network was presented for a successful demonstration. As discussed earlier, two sensing nodes and one server node were developed for the same. The water quality parameters used for ANN modeling are pH, DO, EC, ORP, and Temperature. The suitability of the proposed ANN model for water quality monitoring is also studied. The analysis shows decent results for the training and testing of the proposed ANN model. According to the experiment results, it can be stated that the proposed modules and ANN model can be quite useful for smart water grid development. The alarm or early warning system for any leakage or contamination in the distribution network can be set up based on the water quality rating. The early warning system requires a different decision-making algorithm, which can be implemented using fuzzy logic. The proposed work’s advantages are low-cost hardware compared to high-end conventional instruments and open-source software for programming, data analysis, and ANN implementation. To avoid any uncertainty in sensor readings, all the water quality sensors were calibrated with standard solutions before the measurement.

The maintenance can be easy-going if any sensing node malfunctioning, which can easily be spotted due to real-time monitoring of the pipeline network. The development board (Raspberry Pi) is advanced in the proposed work as any machine learning algorithm can be implemented in Python with open source libraries, and there is no memory limit in the Raspberry Pi. More parameters can be monitored in the proposed work. There are many remote locations in rural areas where there is no internet connection for real-time monitoring. In that case, the GSM module can be interfaced to send the data directly to the cloud for real-time monitoring.

6.5.1. Critical Challenges for a Smart Water Grid

Despite having multiple experimental trials, a smart water grid will face various challenges, as mentioned below. First, the cost of updating the current distribution system architecture is too high, which is not possible without government funding. Second, the integration and communication among different sensing nodes in the WSN will be a challenge in a smart water grid. The data generated by different sensing nodes will have a massive amount of data, which requires high storage and big data analytics. The job redesign of existing staff will also be a challenge as the old rules need to be redundant, and a new one will be imposed on them. Society must accept the technology and the fact that it is going to benefit them in the long term.

6.6. Summary

A smart water grid is part of sustainable and smart city implementation in which information and communication technologies (ICT) play a key role. A smart water grid integrates different individual modules, such as sensor interfacing, data acquisition, cloud updation, real-time, and online monitoring. The proposed ANN model can be a practical tool for water quality assessment. The ANN can also be helpful for big data analytics as the smart water grid will generate a massive amount of data. The presented work can be the foundation of smart cities where water quality monitoring, distribution systems status, water pressure, and flow can easily be monitored in real-time. The developed system may also be helpful in monitoring the water quality in remote areas where there is no internet connection. GSM module can be interfaced to send the data directly to the cloud for real-time monitoring. The data acquired are translated into a single term to define the water quality rating. The developed system is also competent in monitoring, processing the data, decision-making based on the results obtained from the data analysis, and displaying the water quality parameters and water quality rating. The smart city program has already been implemented in some cities worldwide, such as the Australian SEQ water grid and the United States National smart water grid project. Many of the countries are also investing in smart city implementation by replacing the current distribution systems. The smart water grid’s remaining features are a smart water meter, end-user intimation in terms of either SMS or email, which we plan to implement in the future.

Chapter 7

Conclusions and Future Recommendations

Preamble

The conclusions drawn from the work are presented in this chapter covering the design and development of a water quality monitoring system, data analysis, drift compensation and water quality monitoring in distribution networks. Despite the work presented in the thesis, there is a scope for improvement and further extension of the work done. This chapter also discusses the future recommendation that can enhance the functionality of the developed prototype.

7.1.Conclusions

Water quality monitoring before consumption is essential nowadays as the available water is severely polluted due to domestic and industrial waste discharge into the natural resources (either surface or ground), runoff from agricultural land. Traditional water quality monitoring methods are quite tedious as these require sample collection on-site and subsequent lab-based chemical analysis, which is laborious and cost-intensive. These methods are offline, so the user has to wait for the report outcome of the chemical analysis. The following conclusions can be made based on the work presented in the thesis.

- There are various methods for water quality parameter measurement, including traditional laboratory-based and In-situ techniques. The conventional techniques involve on-site sample collection followed by laboratory testing, which is time and cost-consuming. In comparison, in-situ measurements are fast & accurate and save both time as well as cost. The overall water quality can be defined by a single term or a numeric value, obtained by different techniques employing statistical and soft computing methods.

- The overall water quality depends on available water resources in the specific geological region; hence, it is required to identify the region-specific quality parameters before developing a hardware framework. In addition to that, selecting a core controller and related modules and peripherals is essential for hardware development. The methodology for adapting the specific water quality parameters, selection of COTS modules, software, and experimental methodologies has been presented in this thesis.

- The water quality can be defined by either statistical method or soft computing method. The data analysis employing different methods for water quality monitoring on the developed setup in a real-time environment right after the data acquisition has been presented in this thesis work. The classification of water quality employing fuzzy modeling has also been presented in this thesis.

- Although sensor technology has achieved the manufacturing of low-cost and portable water quality sensors, the sensors face drift sooner or later after installation. The drift may occur due to sensor aging, temperature & humidity variation, poisoning among the sensor array, or a combination of all. The soft computing techniques for drift compensation of commercial water quality sensors have been presented in this thesis work.

- The current distribution systems always face leakage, failure, illegal connections, delay in maintenance. The solution to this problem is the implementation of a smart water grid. A smart water grid can manage the water supply in the distribution systems by real-time monitoring of water quality, flow, pressure, and distribution network status. A real-time assessment of water quality in the smart water grid employing and machine learning algorithms has been presented in this thesis work.

7.2.Contribution of Thesis

The specific contributions of this thesis work are as follows:

- The existing methods for the assessment of water quality parameters (conventional methods and real-time measurement) have been investigated. Also, various significant water quality monitoring techniques have been investigated in this work. This work has also investigated different methods for drift analysis and compensation.

- In the thesis work, a prototype is developed based on different commercial off-the-shelf (COTS) modules for real-time drinking water quality monitoring. The developed prototype is capable of real-time water quality display, logging the results, and uploading the results on the cloud. An interactive human-machine interface is given for ease of operation to the end-user where the user can measure individual water quality parameters or overall water quality.

- In the developed setup, we have observed the drift in water quality sensors after a certain period, which has been rectified in this work. Initially, the sensors were calibrated with the available reference solution, and after that, their measurements were recorded for a period of 120 days. The drift was observed and compensated employing the Feed Forward-Artificial Neural Network (FF-ANN) model. The proposed work can also extend the calibration time of the commercial water quality sensors.

- In this work, a distributed network architecture has been proposed for real-time water quality monitoring, preventing delays in maintenance and high wastage of water. The preliminary results show that the proposed architecture can be helpful in smart water grid implementation.

7.3. Future Recommendations

The thesis work paved the way for new research directions. The following are some of the captivating aspects that can be addressed as an extension of the current work.

- The developed hardware prototype has been tested for the samples collected from the campus. This can be further tested for the water samples outside the campus (locality).

- The proposed algorithm for drift compensation of the water quality sensors can be implemented on the Raspberry Pi. Further investigation & validation of the algorithm on hardware can be performed.

- While developing any system, self-diagnosis and self-calibration are always important, which can be carried out as an extension of the work. Predictive maintenance is also an important aspect, which can be addressed in future work.

- E. Coli. has not been included in the current ANN modeling for water quality analysis as the study area has dry weather conditions and has very little chance of E. Coli. Growth. There are shreds of evidence that E. Coli. is found only in old distribution pipelines, pipeline leakage, bad sanitation conditions, or where the storage container is not properly cleaned [109], [110]. There are methods available for the prediction of E.Coli based on the ANN model and statistical analysis [158]–[161], which will be added in the upgraded version of our system. Such a system can be used in other areas of Rajasthan or India.

Bibliography

[1] D. W. Clark and D. W. Briar, “What is ground water?,” 1993.

[2] P. G.-P. I. for S. in Dev., & E., and undefined 1993, “Water in crisis,” *ibiologia.unam.mx*.

[3] F. and A. O. of the UN, “AQUASTAT,” 2016. [Online]. Available: http://www.fao.org. [Accessed: 25-Aug-2017].

[4] K. Anan, “‘Water-Related Diseases Responsible For 80 Per Cent of All Illnesses, Deaths In Developing World’, Says Secretary-General In Environment Day Message,” *UN*, 2003. [Online]. Available: http://www.un.org/press/en/2003/sgsm8707.doc.htm. [Accessed: 05-Dec-2017].

[5] P. Khatri, K. Kumar Gupta, and R. Kumar Gupta, “Raspberry Pi-based smart sensing platform for drinking-water quality monitoring system: A Python framework approach,” *Drink. Water Eng. Sci.*, vol. 12, no. 1, pp. 31–37, Jun. 2019.

[6] “Union Minister of Consumer Affairs, Food & Public Distribution Shri Ram Vilas Paswan releases Water Quality Report for State Capitals & Delhi as analysed by BIS.” [Online]. Available: https://pib.gov.in/newsite/PrintRelease.aspx?relid=194516. [Accessed: 05-Apr-2020].

[7] World Resources Institute, “India Water Tool Version 2.1,” 2016. [Online]. Available: http://maps.indiawatertool.in/iwt/av/IWTUserGuide.pdf. [Accessed: 05-Dec-2017].

[8] “Progress on drinking water, sanitation and hygiene 2000–2017 | UNICEF.” [Online]. Available: https://www.unicef.org/reports/progress-on-drinking-water-sanitation-and-hygiene-2019. [Accessed: 18-Nov-2019].

[9] J. Bartram and R. Ballance, *Water quality monitoring : a practical guide to the design and implementation of freshwater quality studies and monitoring programmes*. E & FN Spon, 1996.

[10] “Water Quality — The Collaborative on Health and the Environment.” [Online]. Available: https://www.healthandenvironment.org/environmental-health/environmental-risks/global-environment/water-quality. [Accessed: 19-Sep-2021].

[11] UNEP, *A Snapshot of the World ’ s Water Quality : Towards a global assessment*. 2016.

[12] "United States Environmental Protection Agency | US EPA," 2010. [Online]. Available: https://www.epa.gov/. [Accessed: 15-Nov-2019].

[13] "National Water Quality Standard for Malaysia." [Online]. Available: http://wepa-db.net/en/topic/waterstandard/Malaysia_1_surface.pdf. [Accessed: 15-Nov-2019].

[14] Central Polution Control Board, "Central Polution Control Board; Environmental Standards; Water Quality Criteria," 2007. [Online]. Available: http://cpcb.nic.in/Water_Quality_Criteria.php. [Accessed: 25-Nov-2017].

[15] "Water Testing | CARO.ca." [Online]. Available: https://www.caro.ca/water-testings/. [Accessed: 31-Jan-2018].

[16] "What is TDS and its relation to water? | APEC Water." [Online]. Available: http://www.freedrinkingwater.com/water-education2/72-tds-water-page2.htm. [Accessed: 31-Dec-2017].

[17] Fundamentals of Envirnmental Measurements, "Dissolved Oxygen - Environmental Measurement Systems," *Fundam. Envirnmental Meas.*, pp. 1–13, 2015.

[18] C. Std, "Turbidity Meter." [Online]. Available: http://www.who.int/water_sanitation_health/hygiene/emergencies/fs2_33.pdf. [Accessed: 31-Dec-2017].

[19] T. H. Hansen *et al.*, "The effect of drinking water pH on the human gut microbiota and glucose regulation: results of a randomized controlled cross-over intervention," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Nov. 2018.

[20] "Total Dissolved Solids Measurements." [Online]. Available: https://iccontrols.com/wp-content/uploads/art-v1400001_total_dissolved_solids_measurement.pdf.

[21] J. Bhardwaj, K. K. Gupta, and R. Gupta, "A review of emerging trends on water quality measurement sensors," in *Proceedings - International Conference on Technologies for Sustainable Development, ICTSD 2015*, 2015, pp. 1–6.

[22] P. Khatri, K. K. Gupta, R. K. Gupta, and P. C. Panchariya, "Towards the Green Analytics: Design and Development of Sustainable Drinking Water Quality Monitoring System for Shekhawati Region in Rajasthan," *Mapan - J. Metrol. Soc. India*, 2021.

[23] C. M. Hussain and R. Keçili, "Electrochemical techniques for environmental analysis," in *Modern Environmental Analysis Techniques for Pollutants*, Elsevier, 2020, pp. 199–222.

[24] R. Bhardwaj *et al.*, "Temperature compensation of ISFET based pH sensor using

artificial neural networks," in *Proceedings of the 2017 IEEE Regional Symposium on Micro and Nanoelectronics, RSM 2017*, 2017.

[25] P. Khatri, K. K. Gupta, and R. K. Gupta, "Drift compensation of commercial water quality sensors using machine learning to extend the calibration lifetime," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 2, pp. 3091–3099, Aug. 2021.

[26] A. Rasmussen and M. E. Zaghloul, "In the flow with MEMS," *IEEE Circuits Devices Mag.*, vol. 14, no. 4, pp. 12–25, Jul. 1998.

[27] A. Boisen, J. Thaysen, H. Jensenius, and O. Hansen, "Environmental sensors based on micromachined cantilevers with integrated read-out," *Ultramicroscopy*, vol. 82, no. 1–4, pp. 11–16, Feb. 2000.

[28] Z. Wu, J. Wang, C. Bian, J. Tong, and S. Xia, "A MEMS-based multi-parameter integrated chip and its portable system forwater quality detection," *Micromachines*, vol. 11, no. 1, 2020.

[29] X. D. Wang and O. S. Wolfbeis, "Optical methods for sensing and imaging oxygen: Materials, spectroscopies and applications," *Chemical Society Reviews*, vol. 43, no. 10. Royal Society of Chemistry, pp. 3666–3761, 21-May-2014.

[30] P. Leonard *et al.*, "Advances in biosensors for detection of pathogens in food and water," *Enzyme Microb. Technol.*, vol. 32, no. 1, pp. 3–13, Jan. 2003.

[31] S. Rodriguez-Mozaz, M. J. Lopez de Alda, and D. Barceló, "Biosensors as useful tools for environmental analysis and monitoring," *Anal. Bioanal. Chem. 2006 3864*, vol. 386, no. 4, pp. 1025–1041, Jun. 2006.

[32] J. Y. Lai, Y. T. Chen, T. H. Wang, H. S. Chang, and J. L. Lai, "Biosensor integrated with transducer to detect the glucose," *Proc. - IEEE Int. Symp. Circuits Syst.*, pp. 2015–2018, 2007.

[33] S. M. Z. Hossain and N. Mansour, "Biosensors for on-line water quality monitoring – a review," *https://doi.org/10.1080/25765299.2019.1691434*, vol. 26, no. 1, pp. 502–518, Jan. 2019.

[34] F. Ustaoğlu, Y. Tepe, and B. Taş, "Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index," *Ecol. Indic.*, p. 105815, Nov. 2019.

[35] R. K. Horton, "An Index Number System for Rating Water Quality," *J. Water Pollut. Control Fed.*, vol. 37, no. 3, pp. 300–306, 1965.

[36] R. Brown, N. McClelland, R. Deininger, and R. Tozer, "A WATER QUALITY INDEX- DO WE DARE," 1970.

[37] J. M. Landwehr, R. A. Deininger, and R. D. Harkins, "An objective water quality

index," *J. Water Pollut. Control Fed.*, vol. 46, no. 7, pp. 1804–1807, 1974.

[38] N. I. McClelland, "Water Quality Index Application in the Kansas River Basin. EPA-907/9-74-001. Contract No. 68-01-0761.," 1974.

[39] T. M. Walski and F. L. Parker, "CONSUMERS WATER QUALITY INDEX.," *ASCE J Env. Eng Div*, vol. 100, no. EE3, pp. 593–611, Jun. 1974.

[40] S. H. Dinius, "DESIGN OF AN INDEX OF WATER QUALITY," *JAWRA J. Am. Water Resour. Assoc.*, vol. 23, no. 5, pp. 833–843, Oct. 1987.

[41] D. Swaroop Bhargava, "Use of water quality index for river classification and zoning of Ganga river," *Environ. Pollution. Ser. B, Chem. Phys.*, vol. 6, no. 1, pp. 51–67, Jan. 1983.

[42] R. A. Smith, R. B. Alexander, and M. G. Wolman, "Analysis and interpretation of water quality trends in major US rivers, 1974-81.," *US Geol. Surv. Water-Supply Pap.*, vol. 2307, 1987.

[43] R. A. Smith, R. B. Alexander, and M. G. Wolman, "Water-quality trends in the nation's rivers," *Science (80-.).*, vol. 235, no. 4796, pp. 1607–1615, 1987.

[44] N. Dalkey, "An experimental study of group opinion: The Delphi method," *Futures*, vol. 1, no. 5. pp. 408–426, 1969.

[45] J. Dojlido, J. Raniszewski, and J. Woyciechowska, "Water quality index applied to rivers in the Vistula river basin in Poland," *Environ. Monit. Assess.*, vol. 33, no. 1, pp. 33–42, Oct. 1994.

[46] S. Dwivedi, I. C. Tiwari, and D. S. Bhargava, "Classification of Water Quality of River Ganga at Varanasi," *Indian J. Environ. Prot.*, vol. 20, no. 9, pp. 688–697, 2000.

[47] S. B. Jonnalagadda and G. Mhere, "Water quality of the odzi river in the Eastern Highlands of Zimbabwe," *Water Res.*, vol. 35, no. 10, pp. 2371–2376, Jul. 2001.

[48] A. A. Bordalo, R. Teixeira, and W. J. Wiebe, "A water quality index applied to an international shared river basin: The case of the Douro River," *Environ. Manage.*, vol. 38, no. 6, pp. 910–920, Dec. 2006.

[49] S. M. Avvannavar and S. Shrihari, "Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India," *Environ. Monit. Assess.*, vol. 143, no. 1–3, pp. 279–290, Aug. 2008.

[50] M. K. Chaturvedi and J. K. Bassin, "Assessing the water quality index of water treatment plant and bore wells, in Delhi, India," *Environ. Monit. Assess.*, vol. 163, no. 1–4, pp. 449–453, Apr. 2010.

[51] G. Singh and R. Kamal, "Application of Water Quality Index for Assessment of Surface Water Quality Status in Goa," *Curr. World Environ.*, vol. 9, no. 3, pp. 994–1000, Dec. 2014.

[52] R. Bhutiani, D. R. Khanna, D. B. Kulkarni, and M. Ruhela, "Assessment of Ganga river ecosystem at Haridwar, Uttarakhand, India with reference to water quality indices," *Appl. Water Sci.*, vol. 6, no. 2, pp. 107–113, Jun. 2016.

[53] K. A. Shah and G. S. Joshi, "Evaluation of water quality index for River Sabarmati, Gujarat, India," *Appl. Water Sci.*, vol. 7, no. 3, pp. 1349–1358, Jun. 2017.

[54] A. K. Chaurasia, H. K. Pandey, S. K. Tiwari, R. Prakash, P. Pandey, and A. Ram, "Groundwater Quality assessment using Water Quality Index (WQI) in parts of Varanasi District, Uttar Pradesh, India," *J. Geol. Soc. India*, vol. 92, no. 1, pp. 76–82, Jul. 2018.

[55] H. A. Ameen, "Spring water quality assessment using water quality index in villages of Barwari Bala, Duhok, Kurdistan Region, Iraq," *Appl. Water Sci.*, vol. 9, no. 8, pp. 1–12, Nov. 2019.

[56] S. Tyagi, B. Sharma, P. Singh, and R. Dobhal, "Water Quality Assessment in Terms of Water Quality Index," *Am. J. Water Resour.*, vol. 1, no. 3, pp. 34–38, Oct. 2020.

[57] A. M. Legendre, *Nouvelles Méthodes Pour La Détermination Des Orbites Des Comètes*. 1805.

[58] J. D. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics | Princeton University Press*. 2009.

[59] V. G. Christensen, P. P. Rasmussen, and A. C. Ziegler, "Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams," in *Water Science and Technology*, 2002, vol. 45, no. 9, pp. 205–211.

[60] K. P. Singh, A. Malik, N. Basant, and P. Saxena, "Multi-way partial least squares modeling of water quality data," *Anal. Chim. Acta*, vol. 584, no. 2, pp. 385–396, Feb. 2007.

[61] I. Chenini and S. Khemiri, "Evaluation of ground water quality using multiple linear regression and structural equation modeling," *Int. J. Environ. Sci. Technol.*, vol. 6, no. 3, pp. 509–519, 2009.

[62] R. Koklu, B. Sengorur, and B. Topal, "Water quality assessment using multivariate statistical methods-a case study: Melen river system (Turkey)," *Water Resour. Manag.*, vol. 24, no. 5, pp. 959–978, Jul. 2010.

[63] M. A. Shareef, A. Toumi, and A. Khenchaf, “Estimation of Water Quality Parameters Using the Regression Model with Fuzzy K-Means Clustering,” 2014.

[64] A. K. Shrestha and N. Basnet, “The Correlation and Regression Analysis of Physicochemical Parameters of River Water for the Evaluation of Percentage Contribution to Electrical Conductivity,” *J. Chem.*, vol. 2018, 2018.

[65] B. Dutta and B. Sarma, “Correlation Study and Regression Analysis of Ground Water Quality Assessment of Nagaon Town of Assam , India,” *Int. J. Eng. Res. Technol.*, vol. 7, no. 06, pp. 320–331, 2018.

[66] K. U. Ahamad, P. Raj, N. H. Barbhuiya, and A. Deep, “Surface Water Quality Modeling by Regression Analysis and Artificial Neural Network,” in *Advances in Waste Management*, Springer Singapore, 2019, pp. 215–230.

[67] D. Ha *et al.*, “Recent achievements in electronic tongue and bioelectronic tongue as taste sensors,” *Sensors and Actuators, B: Chemical*, vol. 207, no. PB. Elsevier, pp. 1136–1146, 01-Feb-2015.

[68] M. Śliwińska, P. Wiśniewska, T. Dymerski, J. Namieśnik, and W. Wardencki, “Food analysis using artificial senses,” *Journal of Agricultural and Food Chemistry*, vol. 62, no. 7. American Chemical Society, pp. 1423–1448, 19-Feb-2014.

[69] T. P. Lambrou, C. C. Anastasiou, C. G. Panayiotou, and M. M. Polycarpou, “A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems,” *IEEE Sens. J.*, vol. 14, no. 8, pp. 2765–2772, 2014.

[70] A. Lee *et al.*, “Online monitoring of drinking water quality in a distribution network: a selection procedure for suitable water quality parameters and sensor devices,” *Int. J. Syst. Assur. Eng. Manag.*, vol. 3, no. 4, pp. 323–337, Dec. 2012.

[71] Y. A. Papadimitrakis, “Automated Water Quality Monitoring In Water Distribution Networks,” in *Integrated Technologies for Environmental Monitoring and Information Production*, Springer Netherlands, 2003, pp. 103–118.

[72] F. Williamson *et al.*, “Online water quality monitoring in the distribution network,” *Water Pract. Technol.*, vol. 9, no. 4, pp. 575–585, Dec. 2014.

[73] D. Zeng, L. Gu, L. Lian, S. Guo, H. Yao, and J. Hu, “On cost-efficient sensor placement for contaminant detection in water distribution systems,” *IEEE Trans. Ind. Informatics*, vol. 12, no. 6, pp. 2177–2185, Dec. 2016.

[74] Y. Wang, S. M. S. M. Rajib, C. Collins, and B. Grieve, “Low-Cost Turbidity Sensor for Low-Power Wireless Monitoring of Fresh-Water Courses,” *IEEE Sens. J.*, vol. 18, no. 11, pp. 4689–4696, Jun. 2018.

[75] R. Yue and T. Ying, "A water quality monitoring system based on wireless sensor network & solar power supply," in *2011 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, CYBER 2011*, 2011, pp. 126–129.

[76] A. Al-Dahoud, M. Fezari, and H. Mehamdia, "Water Quality Monitoring System Using WSN in Tanga Lake," in *Advances in Intelligent Systems and Computing*, 2020, vol. 987, pp. 1–9.

[77] A. T. Demetillo, M. V. Japitana, and E. B. Taboada, "A system for monitoring water quality in a large aquatic area using wireless sensor network technology," *Sustain. Environ. Res.*, vol. 1, no. 1, p. 12, 2019.

[78] P. de Bièvre, "The 2012 International Vocabulary of Metrology: 'VIM,'" *Accreditation and Quality Assurance*, vol. 17, no. 2. Springer, pp. 231–232, 09-Apr-2012.

[79] I. Helm, L. Jalukse, and I. Leito, "Measurement Uncertainty Estimation in Amperometric Sensors: A Tutorial Review," *Sensors*, vol. 10, no. 5, pp. 4430–4455, Apr. 2010.

[80] M. L. Hitchman, *Measurement of Dissolved Oxygen (Chemical Analysis)*. John Wiley & Sons Inc.

[81] D. P. Lucero, "Design of Membrane-Covered Polarographic Gas Detectors," *Anal. Chem.*, vol. 41, no. 4, pp. 613–622, Apr. 1969.

[82] P. R. Warburton, R. S. Sawtelle, A. Watson, and A. Q. Wang, "Failure prediction for a galvanic oxygen sensor," *Sensors Actuators, B Chem.*, vol. 72, no. 3, pp. 197–203, Feb. 2001.

[83] J. M. Barcelo-Ordinas, M. Doudou, J. Garcia-Vidal, and N. Badache, "Self-calibration methods for uncontrolled environments in sensor networks: A reference survey," *Ad Hoc Networks*, vol. 88, pp. 142–159, May 2019.

[84] L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," Aug. 2015.

[85] S. Wold, H. Antti, F. Lindgren, and J. Öhman, "Orthogonal signal correction of near-infrared spectra," in *Chemometrics and Intelligent Laboratory Systems*, 1998, vol. 44, no. 1–2, pp. 175–185.

[86] X. Li, W. Cai, and X. Shao, "Correcting multivariate calibration model for near infrared spectral analysis without using standard samples," *J. Near Infrared Spectrosc.*, vol. 23, no. 5, pp. 285–291, Oct. 2015.

[87] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, and A. Perera, "Drift compensation of gas sensor array data by common principal component

analysis," *Sensors Actuators, B Chem.*, vol. 146, no. 2, pp. 460–465, 2010.

[88] R. Laref, D. Ahmadou, E. Losson, and M. Siadat, "Orthogonal Signal Correction to Improve Stability Regression Model in Gas Sensor Systems," *J. Sensors*, vol. 2017, 2017.

[89] Z. Yi, "Discriminative dimensionality reduction for sensor drift compensation in electronic nose: A robust, low-rank, and sparse representation method," *Expert Syst. Appl.*, vol. 148, p. 113238, Jun. 2020.

[90] M. Holmberg, F. Winquist, I. Lundström, F. Davide, C. Dinatale, and A. D'Amico, "Drift counteraction for an electronic nose," *Sensors Actuators, B Chem.*, vol. 36, no. 1–3, pp. 528–535, 1996.

[91] S. Holmin, C. Krantz-Rülcker, I. Lundström, and F. Winquist, "Drift correction of electronic tongue responses," *Meas. Sci. Technol.*, vol. 12, no. 8, p. 1348, 2001.

[92] Y. Luo, S. Wei, Y. Chai, and X. Sun, "Electronic nose sensor drift compensation based on deep belief network," in *Chinese Control Conference, CCC*, 2016, vol. 2016-Augus, pp. 3951–3955.

[93] L. Zhang, Y. Liu, Z. He, J. Liu, P. Deng, and X. Zhou, "Anti-drift in E-nose: A subspace projection approach with drift reduction," *Sensors Actuators, B Chem.*, vol. 253, pp. 407–417, 2017.

[94] G. Xu, W. Tian, Z. Jin, and L. Qian, "Temperature drift modelling and compensation for a dynamically tuned gyroscope by combining WT and SVM method," *Meas. Sci. Technol.*, vol. 18, no. 5, pp. 1425–1432, 2007.

[95] M. J. Wenzel, A. Mensah-Brown, F. Josse, and E. E. Yaz, "Online drift compensation for chemical sensors using estimation theory," *IEEE Sens. J.*, vol. 11, no. 1, pp. 225–232, 2011.

[96] K. J. Brunner *et al.*, "Compensation of phase drifts caused by ambient humidity, temperature and pressure changes for continuously operating interferometers," *J. Instrum.*, vol. 14, no. 11, pp. P11016–P11016, 2019.

[97] J. Andreu and P. Angelov, "An evolving machine learning method for human activity recognition systems," *J. Ambient Intell. Humaniz. Comput.*, 2013.

[98] F. Piccialli, S. Cuomo, V. S. di Cola, and G. Casolla, "A machine learning approach for IoT cultural data," *J. Ambient Intell. Humaniz. Comput.*, vol. 1, p. 3, Sep. 2019.

[99] J. Seo, T. H. Laine, and K. A. Sohn, "Machine learning approaches for boredom classification using EEG," *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 10, pp. 3831–3846, Oct. 2019.

[100] T. J. B. Durga Devi, A. Subramani, and P. Anitha, “Modified adaptive neuro fuzzy inference system based load balancing for virtual machine with security in cloud computing environment,” *J. Ambient Intell. Humaniz. Comput.*, pp. 1–8, Feb. 2020.

[101] T. Bikmukhametov and J. Jäschke, “Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models,” *Comput. Chem. Eng.*, vol. 138, p. 106834, Jul. 2020.

[102] S. K. Behera, A. K. Rath, and P. K. Sethy, “Maturity status classification of papaya fruits based on machine learning and transfer learning approach,” *Inf. Process. Agric.*, May 2020.

[103] J. M. D. Pereira, O. Postolache, P. M. B. S. Girão, and M. Cretu, “Minimizing temperature drift errors of conditioning circuits using artificial neural networks,” *IEEE Trans. Instrum. Meas.*, vol. 49, no. 5, pp. 1122–1127, 2000.

[104] K. R. Kashwan and M. Bhuyan, “Robust electronic-nose system with temperature and humidity drift compensation for tea and spice flavour discrimination,” in *2005 Asian Conference on Sensors and the International Conference on New Techniques in Pharmaceutical and Biomedical Research - Proceedings*, 2005, vol. 2005, pp. 154–158.

[105] R. Uthra, N. Kalaiarasi, and A. Rathinam, “Neural network based sensor drift compensation of induction motor,” in *International Conference and Workshop on Emerging Trends in Technology 2011, ICWET 2011 - Conference Proceedings*, 2011, pp. 571–575.

[106] S. Adhikari and S. Saha, “Multiple classifier combination technique for sensor drift compensation using ANN & KNN,” in *Souvenir of the 2014 IEEE International Advance Computing Conference, IACC 2014*, 2014, pp. 1184–1189.

[107] S. Sinha, R. Bhardwaj, N. Sahu, H. Ahuja, R. Sharma, and R. Mukhiya, “Temperature and temporal drift compensation for Al2O3-gate ISFET-based pH sensor using machine learning techniques,” *Microelectronics J.*, vol. 97, p. 104710, Mar. 2020.

[108] R. M. Coyte, A. Singh, K. E. Furst, W. A. Mitch, and A. Vengosh, “Co-occurrence of geogenic and anthropogenic contaminants in groundwater from Rajasthan, India,” *Sci. Total Environ.*, vol. 688, pp. 1216–1227, Oct. 2019.

[109] N. Soni and A. Bhatia, “Analysis of Quality of Drinking Water of Private Bore-well and Piped water Supply in Jaipur city, Rajasthan, India,” *Res. J. Recent Sci. Res. J. Recent. Sci*, vol. 4, pp. 313–316, 2014.

[110] S. Chandra, T. Saxena, • Sampat Nehra, and • M Krishna Mohan, “Quality assessment of supplied drinking water in Jaipur city, India, using PCR-based

approach."

[111] Atlas Scientific LLC, "Atlas Scientific Environmental Robotics," 2019. .

[112] "Lab Grade pH Probe | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/ph-probe/. [Accessed: 15-Jul-2021].

[113] "Conductivity Probe K 0.1 | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/conductivity-probe-k-0-1/. [Accessed: 15-Jul-2021].

[114] "Lab Grade Dissolved Oxygen Probe | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/dissolved-oxygen-probe/. [Accessed: 15-Jul-2021].

[115] "Lab Grade ORP Probe | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/orp-probe/. [Accessed: 15-Jul-2021].

[116] "PT-1000 Temperature Kit | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/pt-1000-temperature-kit/. [Accessed: 15-Jul-2021].

[117] "EZO™ pH Circuit | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/ezo-ph-circuit/. [Accessed: 23-Jul-2021].

[118] "EZO™ Dissolved Oxygen Circuit | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/ezo-dissolved-oxygen-circuit/. [Accessed: 23-Jul-2021].

[119] "EZO™ ORP Circuit | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/ezo-orp-circuit/. [Accessed: 23-Jul-2021].

[120] "EZO™ Conductivity Circuit | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/ezo-conductivity-circuit/. [Accessed: 23-Jul-2021].

[121] "EZO™ RTD Temperature Circuit | Atlas Scientific." [Online]. Available: https://atlas-scientific.com/ezo-rtd-temperature-circuit/. [Accessed: 23-Jul-2021].

[122] Raspberry Pi Foundation, "Raspberry Pi Documentation," 2014. [Online]. Available: https://www.raspberrypi.org/documentation/. [Accessed: 15-Mar-2018].

[123] S. Diaz, T. Krohmer, A. Moreira, S. E. Godoy, and M. Figueroa, "An Instrument for Accurate and Non-Invasive Screening of Skin Cancer Based on Multimodal Imaging," *IEEE Access*, vol. 7, pp. 176646–176657, 2019.

[124] A. Monteiro, M. De Oliveira, R. De Oliveira, and T. Da Silva, "Embedded application of convolutional neural networks on Raspberry Pi for SHM," *Electron. Lett.*, vol. 54, no. 11, pp. 680–682, May 2018.

[125] R. A. Rodriguez, P. Cammarano, D. A. Giulianelli, P. M. Vera, A. Trigueros, and L. J. Albornoz, "Using Raspberry Pi to Create a Solution for Accessing Educative Questionnaires from Mobile Devices," *Rev. Iberoam. Tecnol. del Aprendiz.*, vol. 13, no. 4, pp. 144–151, Nov. 2018.

[126] A. Pal, A. Mehta, H. Goonesinghe, D. Mirshekar-Syahkal, and H. Nakano, "Conformal Beam-Steering Antenna Controlled by a Raspberry Pi for Sustained High-Throughput Applications," *IEEE Trans. Antennas Propag.*, vol. 66, no. 2, pp. 918–926, Feb. 2018.

[127] R. E. Sweeney, V. Nguyen, B. Alouidor, E. Budiman, R. K. Wong, and J. Y. Yoon, "Flow Rate and Raspberry Pi-Based Paper Microfluidic Blood Coagulation Assay Device," *IEEE Sens. J.*, vol. 19, no. 13, pp. 4743–4751, Jul. 2019.

[128] "7inch HDMI LCD (B) (Firmware Rev 2.1) User Manual - Waveshare Wiki." [Online]. Available: https://www.waveshare.com/wiki/7inch_HDMI_LCD_(B)_(Firmware_Rev_2.1)_User_Manual. [Accessed: 04-Mar-2018].

[129] "L298N Datasheet(PDF) - STMicroelectronics." [Online]. Available: https://www.alldatasheet.com/datasheet-pdf/pdf/22440/STMICROELECTRONICS/L298N.html. [Accessed: 19-Jul-2021].

[130] Digi International, "Product Manual Xbee / Xbee-PRO RF Modules," 2009.

[131] "Arduino Uno Rev3 | Arduino Official Store." [Online]. Available: https://store.arduino.cc/usa/arduino-uno-rev3. [Accessed: 22-Jul-2021].

[132] NodeMcu, "NodeMcu -- An open-source firmware based on ESP8266 wifi-soc.," *NodeMcu*, 2018. [Online]. Available: https://www.nodemcu.com/index_en.html. [Accessed: 24-May-2021].

[133] "YSI EXO1 Water Quality Sonde | YSI | ysi.com." [Online]. Available: https://www.ysi.com/exo1. [Accessed: 30-May-2020].

[134] Python.org, "Python 2.7.14 documentation," 2018. [Online]. Available: https://docs.python.org/2.7/. [Accessed: 10-Mar-2018].

[135] "Conductivity K 10 Kit | Atlas Scientific." [Online]. Available: https://www.atlas-scientific.com/kits/conductivity-k-10-kit/. [Accessed: 05-Jun-2020].

[136] US Environmental Protection Agency, "Environmental Impact and Benefits Assessment for Final Effluent Guidelines and Standards for the Construction and Development Category," 2009.

[137] T. D. Banda and M. Kumarasamy, "Aggregation Techniques Applied in Water Quality Indices (WQIs)," *Pollut. Res.*, vol. 39, no. 2, pp. 400–41, 2020.

[138] Q. Wen, Z. Ge, Z. Song, and P. Wang, "Dynamic process calibration based on sparse partial least squares," in *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, 2015, vol. 2015-March, no. March, pp. 1366–1371.

[139] P. Khatri, K. K. Gupta, and R. K. Gupta, "A review of partial least squares modeling (PLSM) for water quality analysis," *Modeling Earth Systems and Environment*, vol. 7, no. 2. Springer, pp. 703–714, 15-Oct-2021.

[140] A. Sargaonkar and V. Deshpande, "Development of an overall index of pollution for surface water based on a general classification scheme in Indian context," *Environ. Monit. Assess.*, vol. 89, no. 1, pp. 43–67, Nov. 2003.

[141] J. H. Holland, "Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence," *Ann Arbor Univ. Michigan Press 1975*, pp. viii, 183 p., 1975.

[142] MathWorks, "Fuzzy Inference System Modeling - MATLAB & Simulink - MathWorks India." .

[143] Scikit-learn Developers, "Machine Learning in Python — scikit-learn 0.22.2," *http://scikit-learn.org*, 2020. .

[144] B. Kosko, *Fuzzy thinking : the new science of fuzzy logic*. Hyperion, 1993.

[145] J. Zhao and B. K. Bose, "Evaluation of Membership Functions for Fuzzy Logic Controlled Induction Motor Drive," *IEEE 2002 28th Annual Conference of the Industrial Electronics Society*. pp. 229–234, 2002.

[146] L. A. Zadeh, "Fuzzy Logic," *Computer (Long. Beach. Calif).*, vol. 21, no. 4, pp. 83–93, Apr. 1988.

[147] S. Lee, "PERCENTAGE ERROR," *MathWorld*, 2016. [Online]. Available: http://mathworld.wolfram.com/PercentageError.html. [Accessed: 25-Mar-2018].

[148] Yellow Springs Inc, "EXO User Manual : Advanced Water Quality Monitoring Platform." pp. 1–154, 2016.

[149] "Spyder Website." [Online]. Available: https://www.spyder-ide.org/. [Accessed: 09-Mar-2020].

[150] B. Tudu, B. Kow, N. Bhattacharyya, and R. Bandyopadhyay, "Normalization techniques for gas sensor array as applied to classification for black tea," *Int. J. Smart Sens. Intell. Syst.*, vol. 2, no. 1, pp. 176–189, Nov. 2017.

[151] S. Lakshmanan, “How, When and Why Should You Normalize / Standardize / Rescale Your Data?,” *Https://Medium.Com/@Swethalakshmanan14/How-When-and-Why-Should-You-Normalize-Standardize-Rescale-Your-Data-3F083Def38Ff*. p. https://medium.com/@swethalakshmanan14/how-when-an, 2017.

[152] The Spyder Website Contributors, “Spyder Website,” 2018. [Online]. Available: https://www.spyder-ide.org/.

[153] T. Artursson, T. Eklöv, I. Lundström, P. Mårtensson, M. Sjöström, and M. Holmberg, “Drift correction for gas sensors using multivariate methods,” *J. Chemom.*, vol. 14, no. 5-6, pp. 711–723, Sep. 2000.

[154] Wikipedia, “Root-mean-square deviation - Wikipedia,” 2020. [Online]. Available: https://en.wikipedia.org/wiki/Root-mean-square_deviation.

[155] S. E. Bibri and J. Krogstie, “Smart sustainable cities of the future: An extensive interdisciplinary literature review,” *Sustainable Cities and Society*, vol. 31. Elsevier Ltd, pp. 183–212, 01-May-2017.

[156] B. N. Silva, M. Khan, and K. Han, “Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities,” *Sustainable Cities and Society*, vol. 38. Elsevier Ltd, pp. 697–713, 01-Apr-2018.

[157] “Managing the water distribution network with a Smart Water Grid,” *Smart Water*, vol. 1, no. 1, p. 4, Dec. 2016.

[158] M. I. Yesilnacar, E. Sahinkaya, M. Naz, and B. Ozkaya, “Neural network prediction of nitrate in groundwater of Harran Plain, Turkey,” *Environ. Geol.*, vol. 56, no. 1, pp. 19–25, 2008.

[159] H. A. Zaqoot, M. Hamada, and S. Miqdad, “A Comparative Study Of Ann For Predicting Nitrate Concentration In Groundwater Wells In The Southern Area Of Gaza Strip,” *Appl. Artif. Intell.*, vol. 32, no. 7–8, pp. 727–744, 2018.

[160] D. Dwivedi, B. P. Mohanty, and B. J. Lesikar, “Estimating Escherichia coli loads in streams based on various physical, chemical, and biological factors,” *Water Resour. Res.*, vol. 49, no. 5, pp. 2896–2906, May 2013.

[161] S. N. Shamsudin, M. H. F. Rahman, M. N. Taib, W. R. W. A. Razak, A. H. Ahmad, and M. M. Zain, “Analysis between Escherichia coli growth and physical parameters in water using Pearson correlation,” in *2016 7th IEEE Control and System Graduate Research Colloquium, ICSGRC 2016 - Proceeding*, 2017, pp. 131–136.

Appendix-A

A.1. Sensors and Signal Conditioning Circuit

Conductivity Probe (ENV-40-EC-K1.0)

DO Probe (ENV-40-DO)

ORP probe (ENV-40-ORP)

pH Probe (ENV-40-pH)

Temperature Probe (#PT-1000)

A.2. Reference Solutions

ORP reference (225 mV)

DO reference (0 mg/l)

EC reference (84 µS/cm and 1413 µS/cm)

pH reference (4, 7 and 10)

Appendix-B

Specification of the Developed Prototype

- **Software:** Python (open source)
- **Logging capabilities:** Yes
- **Memory:** depends on Raspberry Pi memory (more than 100000 logged readings are possible)
- **Power:** 230 V Adapter
- **Smart sensor/ports:** No
- **Measurement unit:** Parameter dependent
- **User calibratable:** Yes
- **Waterproof:** No
- **Measured parameters:** pH, EC, ORP, DO, and temperature
- **Derived parameters:** TDS, salinity
- **Sampling rate:** up to 2 Hz
- **Online Monitoring:** Available on ThingSpeak Cloud Platform

Appendix-C

Fuzzy Sets Theory

Assume that X is a set of objects and x is an element of X. A classical set (A) can be defined such that each element belongs to set A. So, the membership function of a classical set can be defined as 1 if it belongs to A and 0 if it does not belong to A. The membership function for a classical set can be represented as follows.

$$\mu_A(x) = \begin{cases} 1, & if \quad (x \in A) \\ 0, & if \quad (x \notin A) \end{cases} \tag{C.1}$$

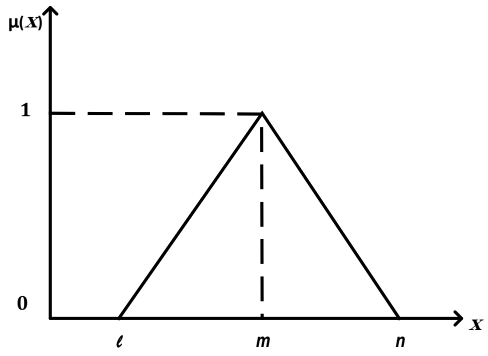
A fuzzy set, unlike the classical set described above, expresses the degree to which an element belongs to a set. As a result, the membership function of a fuzzy set can have values between 0 and 1, denoting an element's degree of membership in the set.

A.1. Fuzzy Sets and Membership Functions

Let X be a collection of x objects; then a fuzzy set can be defined as $A = \{(x, \mu_A(x)) | x \epsilon X\}$. In the given fuzzy set, $\mu_A(x)$ is the membership function (MF). Each element of X is assigned a membership value between 0 and 1, which can be called as an extension of a crisp set. The property of membership functions is subjective, meaning it can vary from person to person for the same idea. X is partitioned into different fuzzy sets, each with MFs that cover X more or less uniformly. Linguistic values or linguistic labels are fuzzy sets that usually have names that conform to descriptors that arise in our everyday verbal usage, such as “small,” “medium,” or “large.”

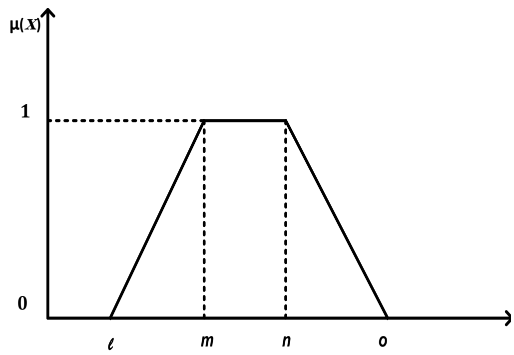
A.1.1. Membership functions

MFs can be specified in several ways, and they can be based on different functions. Following are some functions that are commonly used. The standard triangular and trapezoidal membership functions are shown in Figures C.1 and C.2. The triangular membership function can be given as Eq. (C.2).



$$f(x; l, m, n) = \begin{cases} 0 & \text{for } x < l \\ \dfrac{x-l}{m-l} & \text{for } l \le x \le m \\ \dfrac{n-x}{n-m} & \text{for } m \le x \le n \\ 0 & \text{for } x > n \end{cases}$$

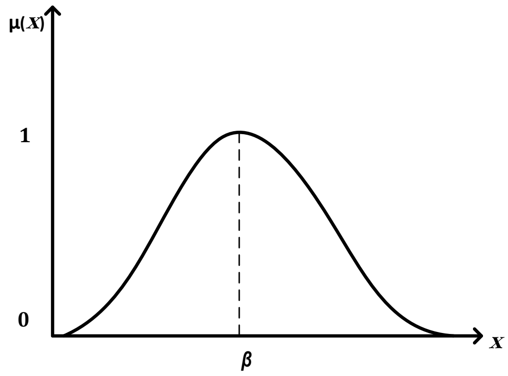
Figure C.1 Triangular Membership Function **Equation C.2**



$$f(x; l, m, n) = \begin{cases} 0 & \text{for } x < l \\ \dfrac{x-l}{m-l} & \text{for } l \le x \le m \\ 1 & \text{for } m < x < n \\ \dfrac{o-x}{o-n} & \text{for } n \le x \le o \\ 0 & \text{for } x > n \end{cases}$$

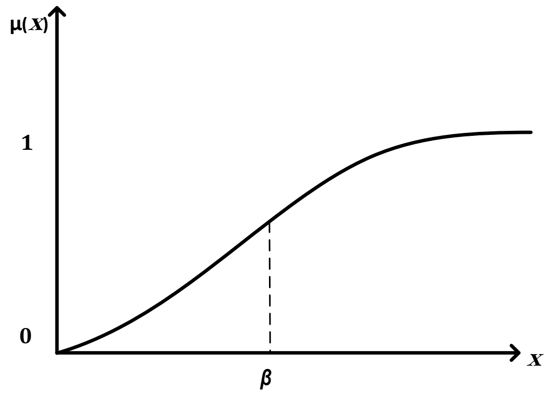
Figure C.2 Trapezoidal Membership Function **Equation C.3**

The membership function for the given trapezoidal membership function can be defined as Eq. (C.3). The gaussian and sigmoid membership functions are shown in Figures C.3 and C.4. The gaussian MF ($G: X \rightarrow \{0; 1\}$) is given in Eq. (C.4). Where, γ is the slope and β is the midpoint. The slope must be positive and should never reach '0'. The sigmoid MF (S-MF) ($S: X \rightarrow \{0; 1\}$) is given by Eq. (C.5). The value of S-MF neither reaches '1' nor '0'. In the S-MF, γ is the slope value at inflexion point and β is the midpoint.



$$G(x;\beta,\gamma) = \exp\left(-\gamma(x-\beta)^2\right)$$

Figure C.3 Gaussian Membership Function **Equation C.4**



$$S(x;\beta,\gamma) = 1/(1+\exp(-\gamma(x-\beta)))$$

Figure C.4 Sigmoid Membership Function **Equation C.5**

Appendix-D

Partial Least Squares Regression (PLSR)

The PLS Regression algorithm estimates the matrices W, T, O, and P through the following steps.

Algorithm D.1 Partial Least Squares Regression pseudo algorithm

1. Initialize the residuals matrices $R_0 = X_{nxp}$ and $S_0 = Y_{nxk}$;

   for $i = 1$ to p **do**

2. Calculate PLS weights vector

   $\mathbf{W_i = R_0{}^T S_0;}$

3. Calculate and normalize scores vector

   $\mathbf{T_i = R_0 W_i (W_i{}^T R_0{}^T R_0 W_i)^{\frac{-1}{2}}}$;

4. Calculate X loading vector

   $\mathbf{O_i = R_0{}^T T_i}$;

5. Calculate Y loading vector

   $\mathbf{P_i = S_0{}^T T_i}$;

6. Update the **X** residuals vector

   $\mathbf{R_0 = R_0 - T_i O_i{}^T}$;

7. Update the Y residuals vector

   $\mathbf{S_0 = S_0 - T_i P_i{}^T}$;

   end for

8. Obtain output matrices $\mathbf{W}, \mathbf{T}, \mathbf{O}, \mathbf{P}$.

List of Publications

- ***Work Specific***

Journal Paper:

- **Khatri P**, Gupta K.K. and Gupta R.K. Raspberry Pi-based smart sensing platform for drinking-water quality monitoring system: a Python framework approach. *Drink Water Eng Sci*. 2019;12(1):31-37. doi:10.5194/dwes-12-31-2019

- **Khatri P**, Gupta K.K. and Gupta R.K. A Comprehensive Study on the Effects of Water Quality Parameter Variation on Water Quality and Water Quality Index. *Res. J. Chem. Environ.*, vol. 23, no. 12, pp. 114–123, 2019.

- **Khatri P**, Gupta K.K. and Gupta R.K. Drift compensation of commercial water quality sensors using machine learning to extend the calibration lifetime. *J Ambient Intell Humaniz Comput*. August 2020. doi:10.1007/s12652-020-02469-y

- **Khatri P**, Gupta K.K. and Gupta R.K. Assessment of Water Quality Parameters in Real-Time Environment. *SN Comput Sci*. 2020;1(6):340. doi:10.1007/s42979-020-00368-9

- **Khatri P**, Gupta K.K. and Gupta R.K. (20xx) 'Real-time Water Quality Monitoring for Distribution Networks in IoT Environment', '*Int. J. Environment and Sustainable Development*', In Press.

- **Khatri P**, Gupta K.K., Gupta R.K. and Panchariya P.C., Towards the Green Analytics: Design and Development of Sustainable Drinking Water Quality Monitoring System for Shekhawati Region in Rajasthan, MAPAN. (2021). doi:10.1007/s12647-021-00465-x.

Conference and Proceedings:

- **Khatri P,** Gupta K.K. and Gupta R.K. Development of Cyber-Physical Systems for Water Quality Monitoring in Smart Water Grid. AICTE Sponsored Online 3rd International Conference on Recent Trends in Communication & Intelligent Systems (ICRTCIS), Oct 2021. *In press*.

- **Khatri P**, Gupta K.K. and Gupta R.K. "Water Quality Index Calculation: Switching from MATLAB Fuzzy Toolbox to Python for Real-Time Implementation," 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI), 2020, pp. 1-5, doi: 10.1109/ICATMRI51801.2020.9398318.

- **Khatri P**, Gupta K.K. and Gupta R.K. Smart Water Quality Monitoring System for Distribution Networks. SSRN Electron J. 2019. doi:10.2139/ssrn.3352296

- Bhardwaj J, Gupta K.K., and **Khatri P**. Real Time Assessment of Potable Water Quality in Distribution Network based on Low Cost Multi-Sensor Array. In: IOP Conference Series: Materials Science and Engineering. Vol 331. Institute of Physics Publishing; 2018. doi:10.1088/1757-899X/331/1/012027J.

Book Chapter:

- **Khatri P**, Gupta K.K. and Gupta R.K. (2021) Drift Compensation of a Low-Cost pH Sensor by Artificial Neural Network. In: Bansal J.C., Paprzycki M., Bianchini M., Das S. (eds) Computationally Intelligent Systems and their Applications. Studies in Computational Intelligence, vol 950. Springer, Singapore. https://doi.org/10.1007/978-981-16-0407-2_8

➢ ***Others***

- **Khatri P**, Gupta K.K. and Gupta R.K. A review of partial least squares modeling (PLSM) for water quality analysis. *Model Earth Syst Environ*. 2020;1:3. doi:10.1007/s40808-020-00995-4

- **Khatri P**, Gupta K.K. and Gupta R.K. Multi-way Partial Least Squares (MPLS) Calibration of Low-Cost Water Quality Instrument. *J Ambient Intell Humaniz Comput*. Revision Under Review

Mr. Punit Khatri joined BITS-Pilani in 2017 as a Junior Research Fellow in the Department of Electrical and Electronics Engineering. In 2019, he was awarded CSIR SRF-Direct fellowship. He has completed his Diploma (Electronics), B. Tech. (E & C), M. Tech. (Instrumentation) in 2007, 2010, and 2015 respectively. He has R & D work experience of almost 8 years. He has published many papers in conferences and International Journals of repute. His area of interest is intelligent system design, multivariate data analysis, instrumentation system, and prototype development.

Prof. Karunesh Kumar Gupta completed his Ph.D. in digital image processing. He has completed his M.E. in control and instrumentation from NIT Allahabad, UP, India, and B.Tech. in electronics from Lucknow University in 1994 and 1991, respectively. His major field of study is acoustic and vibration-based machine health monitoring.

Currently, he is working as an **Associate professor** in the Department of Electrical and Electronics Engineering, Birla Institute of Technology and science, Pilani, India. He has published many research articles in conferences and peer-reviewed journals. He has also served as a Reviewer in many reputed International Journals. His current research interests are high-speed vision application in robotics, acoustic and vibration-based machine health monitoring, drinking water quality measurement, biometrics, and compressed domain image analysis.

Prof. Raj Kumar Gupta completed his Ph.D. in Physical Sciences from Raman Research Institute affiliated to Jawaharlal Nehru University, New Delhi, India in 2006. He has completed his M.Sc. and B.Sc. in Physics from Calcutta University in 1999 and 1997, respectively. His major field of study is understanding the change in properties of materials in the ultrathin film regime and its application for device fabrication.

Currently, he is working as a **Professor** in the Department of Physics, Birla Institute of Technology and Science, Pilani, India. He has filed an Indian patent on the topic entitled "A novel optomechanical system for measuring surface plasmon resonance". He has published many research articles in peer-reviewed journals. He has also served as a Reviewer in many reputed International Journals. His current research interests are surface plasmon resonance, instrumentation, and related software development, scanning probe microscopy of thin film of nanoparticles, controlling parameters for defect formation in thin films, scanning tunneling microscopy/spectroscopy, and application of ultrathin films for device development etc.