E- TONGUE: A SMART TOOL TO PREDICT SAFE CONSUMPTION OF GROUND WATER

Project ID: CDAP 2020-164

Project Proposal Report

Munaz Mohammed Nibraz

Bachelor of Science (Hons) Degree in Information Technology Specializing in Software Engineering

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

February 2020

E- TONGUE: A SMART TOOL TO PREDICT SAFE CONSUMPTION OF GROUND WATER

Project ID: 2020-164

Project Proposal Report

(Proposal documentation submitted in partial fulfilment of the requirement for the Degree of Bachelor of Science Special (honors) In Information Technology)

Munaz Mohammed Nibraz

IT16165694

Bachelor of Science (Hons) Degree in Information Technology Specializing in Software Engineering

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology Sri Lanka

February 2020

Declaration

I declare that this is my own work and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Registration Number	Signature
Nibraz M M	IT16165694	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Supervisor:

Name	Signature	Date
Ms. Chathurangika Kahandawaarachchi		

Abstract

Water dominates the surface of the earth and it is an indispensable element when making up the human body. A healthy lifestyle explicates the intake of healthy food, water, air and making sure that we are around all these building blocks which are safe and clean is essential. Although different people use different types of water from different sources, it is proven that the groundwater sources satisfy the necessities of the majority of people living in Sri Lanka, and extra care should be grasped when utilizing this direct water from sources to avoid the consumers from acquiring fatal or non-fatal diseases. Taking this problem into consideration, it is comprehended that the importance of water quality must be spoken out to the public, and the process of making sure that the quality of water used in day to day consumption falls under a safe range.

One of the paramount solutions for this issue is to propose a fast and easy method using a technological blend to replace the manual and time-consuming chemical methods which are currently practiced in Sri Lanka. "E-Tongue – a smart device to identify the quality of a ground water sample" is a smart solution which hopes to address the ongoing subject of identifying the water quality with the use of technology. This device uses an Artificial Neural Network (ANN) model by using Water Parameters as the inputs. The output of this model, the Water Quality Index (WQI) will be used to clearly determine the quality of the water sample obtained from a water source. The final outcome of this attempt is to be showcased in the form of a mobile application.

Keywords: Water quality, Artificial Neural Network (ANN), water parameters, Water Quality Index (WQI).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

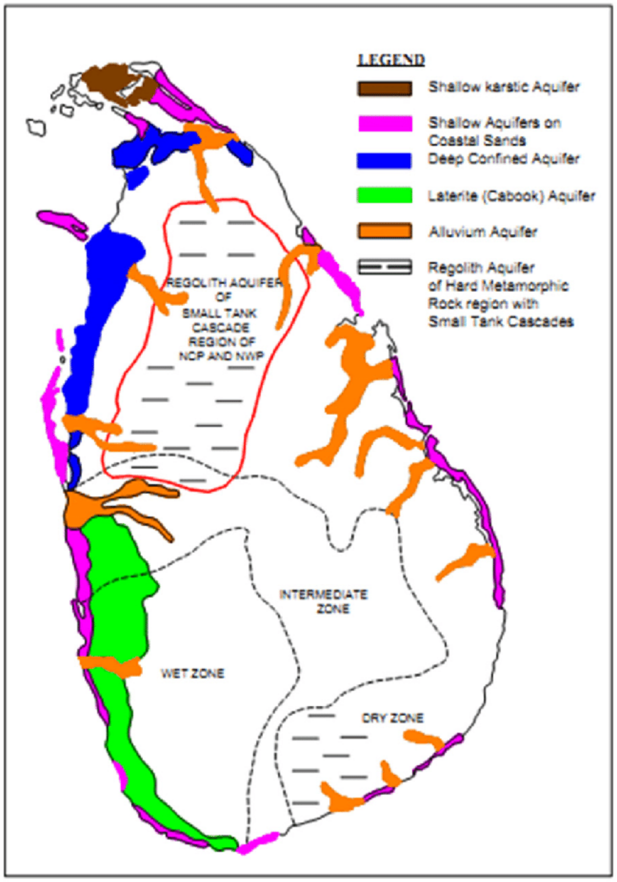
LIST OF ABBREVIATIONS

Abbreviations	Description
WQI	Water Quality Index
WQV	Water Quality Variables
NN	Neural Network
ANN	Artificial Neural Network
CKDu	Chronic kidney disease of unknown etiology
ML	Machine Learning
SVM	Support Vector Machines
Deep NN	Deep Neural Network
WHO	World Health Organization
PFA	Principle Factor Analysis
NWSDB	National Water Supply & Drainage Board
DO	Dissolved Oxygen

1 INTRODUCTION

The quality of water is an overall description of the biological, chemical and physical characteristics of water in connection with intended uses and a set of standards[14]. Hence, water quality assessment can be defined as the evaluation of the biological, chemical and physical properties of water in reference to natural quality and intended uses. In present days the water quality is solely measured using a global index called Water Quality Index (WQI) which aggregates a set of complex mathematical formulas on certain Water Quality Variables (WQVs). However, in Sri Lanka the use of this WQI is not yet practiced which makes this an attempt to introduce the concept in for more considerations. The main scope of this component is to develop a neural network model which can be used to find the Water Quality Index values that helps in understanding the overall quality of the ground water sample.

1.1 Background Study

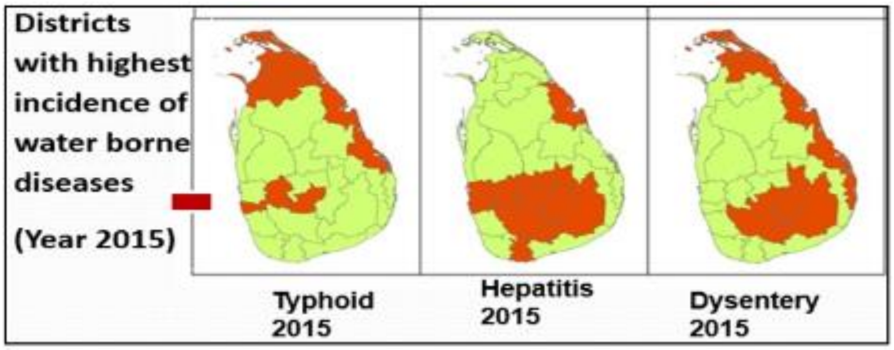


(figure 1.1 - Different Types of Aquifers in Sri Lanka)

Even though Sri Lanka as a country has seen many industrialized development and extravagant systems, the issues that revolve around drinking water has remained virtually untouched. The quality of ground water people consume for drinking as well as the lives of the people who have been consuming this water of questionable quality is very much at stake.

Among the variety of diseases caused by the consumption of polluted and unsafe water, one of the fatal diseases that has been recorded to seize lives of thousands of victims is Chronic Kidney Disease of Unknown Etiology (CKDu). It is mentioned that the incidence of CKDu has been doubling every four to five years, and currently more than 150,000 people are affected by this dangerous disease. An estimated 3% of them lose their lives annually[4].

Apart from CKDu, several articles and documentaries suggest that even some diseases like diarrhea which people consider as a simple and curable ailment can seize lives if left untreated [5]. This extreme situation puts the people of the country to be more cautious when it comes to dealing with water for consumption or any domestic purposes, specially the people whose lives on ground water sources.



(figure 1.2 – Districts with highest water borne diseases)

All these hazardous impact on people and society, questions the overall ideas of water purification, people's knowledge on the precautions, the accuracy of the water quality predictions that are done by the authorities who supplies water and most importantly a ground water consumer's knowledge on the quality of water they are consuming. In 2015, it was mentioned out of 200 participants who contributed to a research which was conducted in Jaffna, 87% of people had been using water from sources which were contaminated with Coliforms and *E.coli*[6].

Real time problem like this should have a real time solution. A smart device that could be used by any person to identify the quality of ground water in a real time period could be an effective solution to help the users to avoid long chemical and mathematical process of identifying the quality of the ground water they are dealing with. The above-mentioned background studies suggest a smart tool and the smart tool needs a proper Machine learning model to output the water quality in real time. A suitable Machine learning model with a less error and high accuracy is what makes the smart device more efficient and reliable.

Even though this smart solution won't be a complete elucidation for the ongoing water quality issue and waterborne diseases, if it can make a change in one individual's life to help identify the safeness of the ground water for consumption and domestic usage, this could be considered as a significant success.

1.2 Literature Review

According to the Linkoping University, department of applied physics, Prof. Fredrik Winquist was the inventor of the voltammetry based electronic tongue concept [7] which indeed was an inspiration for the development of an electronic tongue based smart device to address an ongoing environmental issue. Even though the concept was used in many industrial purposes like pharmaceutical drug testing and whisky taste testing, a chemical model to analyze the performance of the electrodes of a custom made voltammetric e-tongue for the evaluation of infused tea gives a sole chemical approach to the idea behind the voltammetry based electronic tongue concept[8][9] which then provided the way for more advanced mathematical model implementations for identification of tea samples[10].

According to Shafi U in[11], classical Machine learning algorithms like Su*pport Vector Machines (SVM)*, *Deep Neural Networks (Deep NN)* and *Neural Networks (NN)* were used to measure the water quality with a highest accuracy of 93%. This level of accuracy distinctly indicates the importance of training the selected mathematical model under both controlled and open field conditions. It can also be stated that NN and Deep NN models are highly fitting in training a model which includes complex functions. Furthermore, it is important to note here that out of 30 water quality variables which were defined by World Health Organization (WHO), 25 variables are used in order to achieve this highest accuracy. However, using 25 different types of sensors makes this system economically infeasible due to budgetary restraints. Sakizadeh, M[12] used 16 water quality parameters along with an *Artificial Neural Network (ANN)* with Bayesian regularizations. This study capitulated correlation coefficients between the observed and the predicted values of 0.94 and 0.77. Even though the reduction of number of used variables didn't impact on a vast difference in the accuracy and mean error, using 16 sensors on respective variables puts the progress of the study into a tight spot.

In a comprehensive overview [13], the study suggests a different and more efficient and scalable approach when it comes to selecting the variables for the WQI calculation. Once a data set has been collected, it is initially passed through a *Principal Factor Analysis (PFA)* where all the variables present in the dataset will be preprocessed to select the best suiting variables while preserving the overall variance as much as possible. This step addresses a principal issue when it comes to designing the device with using suitable sensors that could be within the planned budget. This paper also states the importance of parallelly training multiple algorithms, selecting the best algorithm with the highest accuracy and a least mean error to proceed with an effective mathematical model for a reliable WQI output[13][14].

1.3 Research Gap

After conducting a thorough background study and literature review, it was identified that similar research has been carried out for several instances which mainly focuses on moving bodies such as rivers[2][3]. However, the approaches that had been taken were very specific to the moving body of water, and the real issue related to the water quality of groundwater sources has not been resolved with a proper solution. The approaches that the water source supply firms taking are manual ones. The lack of a Machine learning approach to compute the water quality faster and more efficiently has created a research gap, to which a solution is being proposed through this component.

Therefore, the main idea behind this particular component is to carry out an extensive study on the Water Quality Variables and approach with a suitable algorithm to find the Water Quality of groundwater sources of a specific location through the use of a best possible mathematical model.

Name	Physical and Chemical	Bacteriology	Metals and Heavy Metals	Algae	TOC	THM	Pesticides	Oil and Grease	Waste Water	Filter Media
National Water Supply & Drainage Board, Thelawala Road, Ratmalana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
National Water Supply & Drainage Board, Wathhimi Road, Kurunegala	✓	✓								✓
National Water Supply & Drainage Board, New Vishaka Road, Bandarawela.	✓	✓	✓	✓				✓	✓	
National Water Supply & Drainage Board, Manager Office, Nupe, Matara.	✓	✓						✓	✓	✓
Regional Laboratory, National Water supply &Drainage Board, Vavunia.	✓	✓								

National Water Supply & Drainage Board, Sivan Pannai Road, Jaffna.	✓	✓	✓		✓			✓	✓	

(Table 1.1 – few manual water quality testing labs present in Sri Lanka [15])

1.4 Research Problem

After realizing the importance of a healthy lifestyle and the importance of the quality of water we use for drinking and domestic purposes, new techniques and methodologies have been discovered to ensure the safety of consumable water. With the advancement of technological facets, institutes and researchers have put thought in combining technology with mathematical processes to replace the manual chemical methods that have been used for a significant time period to increase the quality of water.

The techniques that are currently in use to identify water quality in Sri Lanka vary from the methodologies followed by other developed countries and are also less accurate as it still remains a developing country. The concept of WQI is not utilized here as it involves a major issue in using a machine learning based approach for the identification of water quality. Therefore, the introduction of WQI into water quality identification will be an essential to help mitigate issues related to the accuracy of the identified grade of water. A suitable mathematical model must be identified and trained in order to perform an accurate task. Furthermore, the prediction of what will be the quality of that water source in future will help the user to draw up necessary precautions to avoid any unnecessary consumption.

2 OBJECTIVES

2.1 Main Objective

The main objective of this function is to identify the Water Quality Index (WQI) of a sample of water which belongs to a specific location that the user is intending to examine. The identified WQI value could then be used to identify the quality of the water sample which was obtained from the water source.

In order to identify the numeric value of WQI, a set of water quality parameters should be processed and the readings of those parameters should be obtained from the sensors. The ultimate goal would be to make sure that the people are using safe water for consumption and domestic purposes.

2.2 Specific Objectives

- To identify the intended requirement to perform a thorough study on the past work and to understand the feasibility of the requirement.
- Identify a data set which could be used to progress with the ML model.
- To select the set of water parameters that has the highest impact on identifying the Water Quality Index while preserving overall variance as much as possible.
- Selecting appropriate algorithms to train the mathematical model.
- Analyze the output from all the used algorithms and identify the prime algorithm with high accuracy and low error in order to select the best model.
- Testing the model with a known set of data to make sure that the WQI output is accurate as intended.

3 RESEARCH METHODOLOGY

The basic functionality of this component is to analyze the water sample and predict the quality of the sampled groundwater. Once a sample is collected from the water source, a further study is performed to get the specific readings of the water quality values through the use of appropriate sensors. The water quality values determine the actual quality of the sample water determining its feasibility to drink or not. Once the sensor readings are obtained, the values will be used by the neural Network model to predict the Water Quality Index (WQI).

Parameter	Units	Standards[a]	Mean±SD[b]	Range	Suitability[c]
pH	pH units	6.5 – 8.5	8.07±0.08	7.91 – 8.20	S
DO	mg L^{-1}	4 – 6	4.05±0.72	3.35 – 4.73	S
Turbidity	NTU	5	10.7±0.21	8.63 –14.2	NS
TDS	mg L^{-1}	500	117±15.9	105 – 133	S
Total Coliforms (TC)	MPN·100 ml^{-1}	50	$1.7\times10^5\pm7.9\times10^4$	$7.2\times10^4 - 3.1\times10^5$	NS

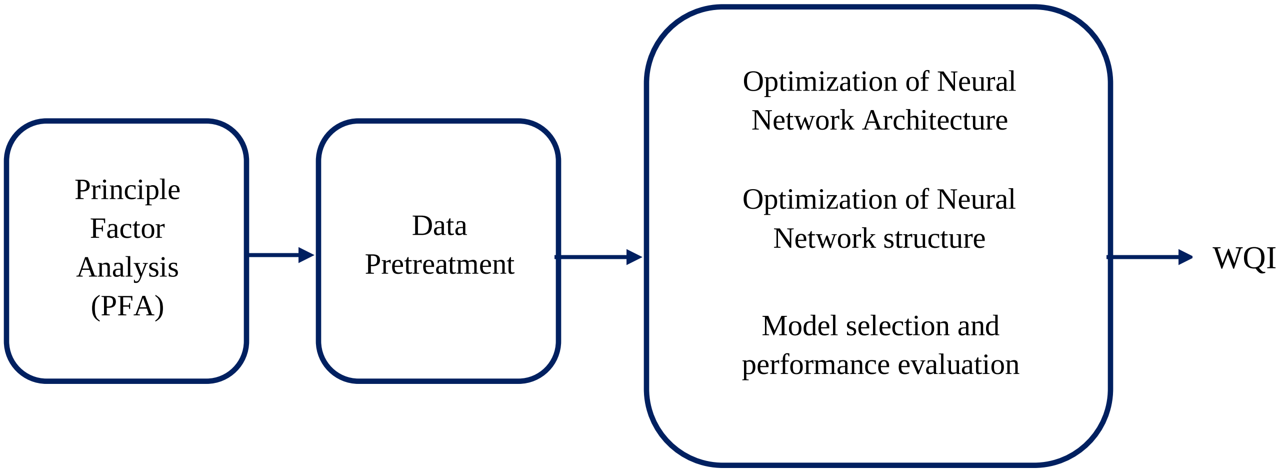
[a] WHO suggested water quality standards (Gray 2008, WHO 2004)
[b] Values are averaged from at least three consecutive measurements. SD: standard deviation
[c] Suitability for drinking as compared with WHO suggested water quality standards. 'S', suitable; 'NS', not-suitable.

(Table 3.1 – Water Quality Value standards according to WHO)

3.1 Procedure:

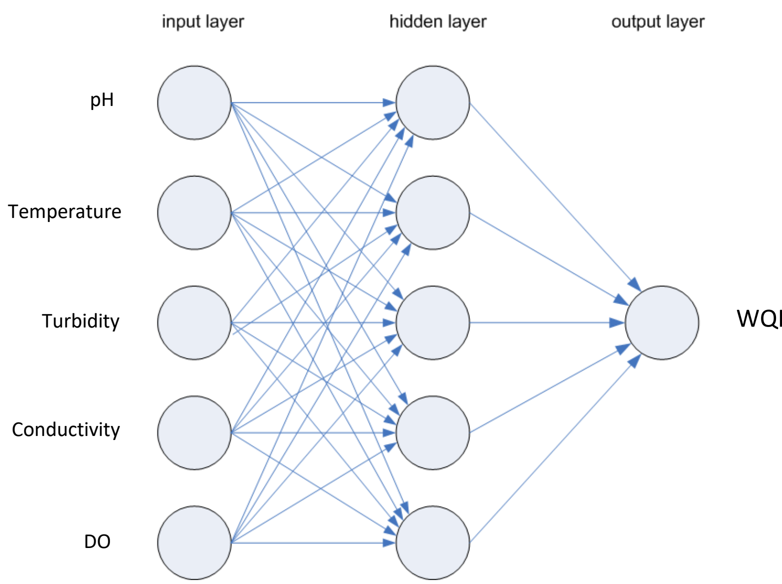
In general, the water quality index is determined using a set of complex mathematical formulas by processing the values of the water quality elements. This component is implemented by using an advanced ***Artificial Neural Network (ANN)*** *technique* where a set of neural network models are trained and a best model is selected on a basis where high accuracy of the output and a low percentage error occurs. The data set which will be used to train the ANN model will be collected from the Department of Irrigation of Sri Lanka and Water Resources Board. It shall include the past data of the water quality and the water quality variable values of a specific location which are used to determine the quality of the water sources.

Once the data has been collected, a *Principal Factor Analysis (PFA)* will be immediately performed on the data to simply reduce a somewhat large set of WQVs to a smaller and more manageable number while preserving as much of the overall variance as possible[13].



(figure 3.1 – methodology flow diagram)

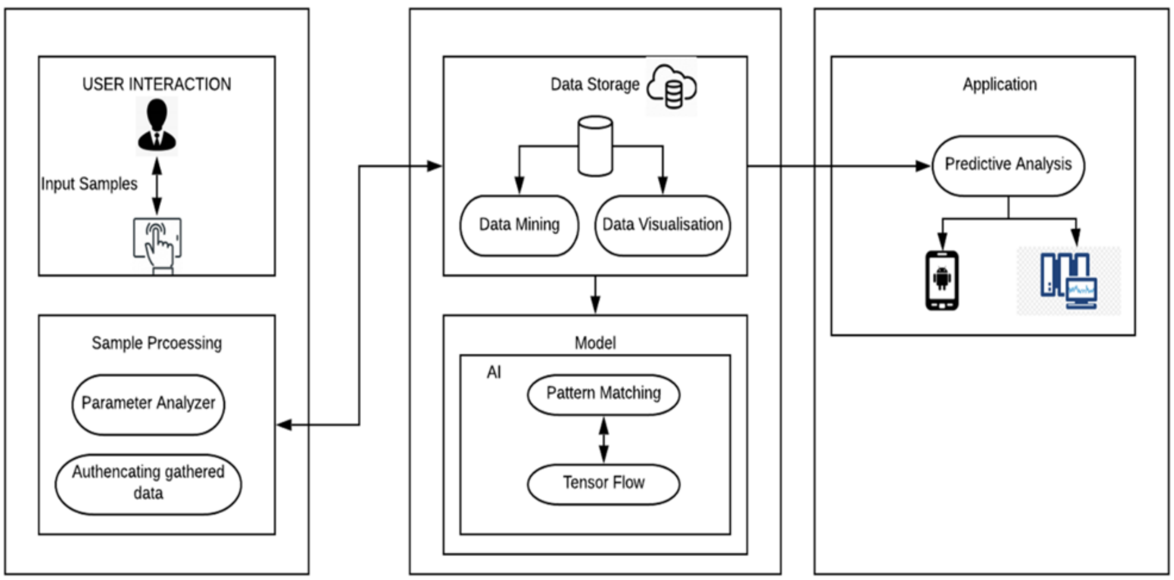
The selected water quality variables will then be standardized in the data pretreatment stage and used to develop the ANN model. When focusing on the development of the mathematical model, a series of steps are considered to make the process more understandable. ***Optimization of network architecture*** solely defines the architectural idea of the ANN model specifying the characteristics of the nodes present in each layer.



(figure 3.2 – structure of an ANN model)

Figure 3.2 explains the structure of the ANN model, giving a rough idea of the number of input, hidden and output nodes that could be used in the model design. The next step, ***Optimization of the network structure*** aims mainly at identifying the optimal training algorithm, learning rate, number of iterations and training stopping conditions. Once a suitable algorithm is identified, then the training of the network model will commence in the ***Model selection and performance evaluation*** phase. Selecting and training an optimal ANN model then could be used to obtain WQI as the output.

The results are sent to the cloud from where the mobile application could access and display the WQI value and the reference range to exhibit to the users to which category the water sample belongs to.



(figure 3.3 – System Overview Diagram)

3.2 Technology Selection

Software component:-

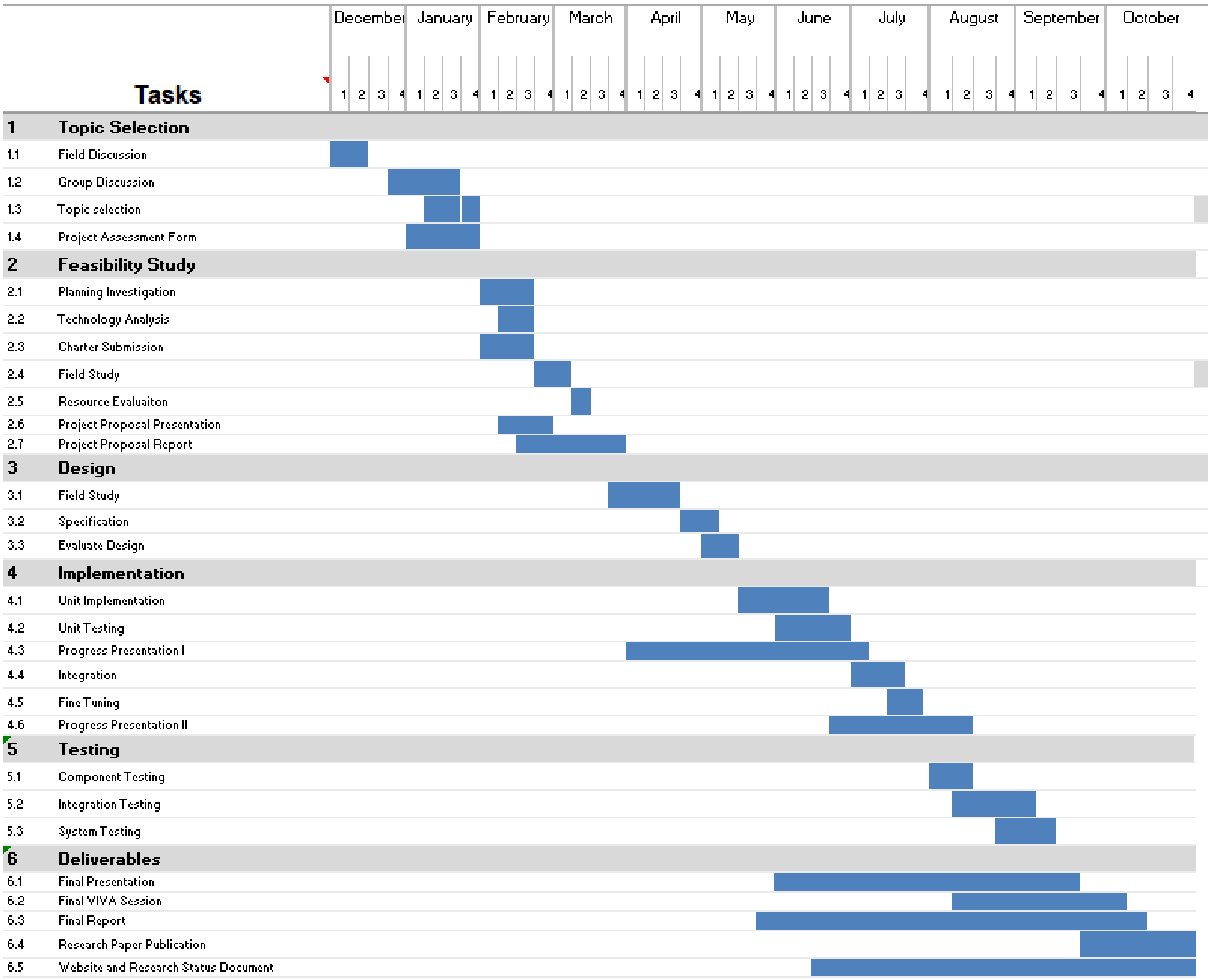
Mobile Development	:Android
Machine learning and Artificial Neural Network	:Tensorflow on Python
Continuous Integration and Continuous Delivery	:Jenkins
Version controlling	:Git

Project Management:-

Trello

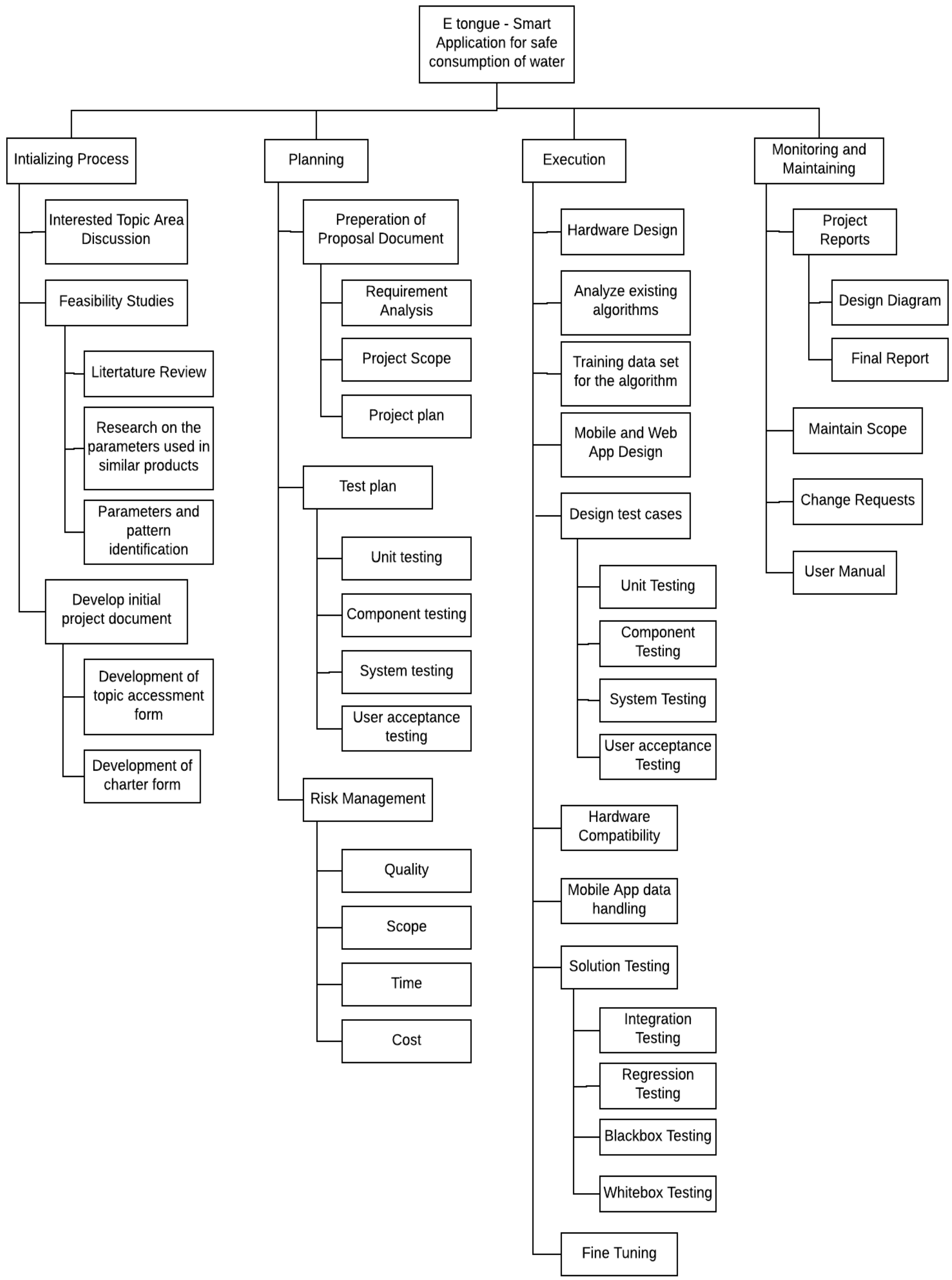
Slack

3.3 Gantt Chart



(figure 3.4 – Gantt Chart)

3.4 Work Breakdown Structure



(figure 3.5 – Work Breakdown Structure)

4 DESCRIPTION OF PERSONAL AND FACILITIES

Member	Component	Tasks
Nibraz M M	Identification of the Water Quality Index by selecting the best algorithm and the best Artificial Neural Network model.	• Carrying out an extensive study on water quality and the water quality variables. • Selecting the best set of water quality parameters to be used in training the model. • Finding a suitable set of algorithms that could be used in this specific scenario. • Setting up Python 3.5 and Tensorflow Lite through the Anaconda distribution. • Training the mathematical model to obtain an optimum result. • Pulling the sensor readings which are stored in cloud. • Collecting water samples from specific location to test the accuracy of the real time performance.

(Table 4.1 –Task Allocation)

REFERENCES

[2] Amir Hamzeh Hagiabi, Ali Heider Nasrolahi and Abbas Parsaie, Water quality prediction using machine learning model, Water quality research journal 2018.

[3] H. B. Chu1, W. X. Lu1∗, and L. Zhang2, Application of Artificial Neural Network in Environmental Water Quality Assessment, J. Agr. Sci. Tech. (2013) Vol. 15: 343-356

[4] Kashi Kafle, Soumya Balasubramanya, Ted Horbulyk, Prevalence of chronic kidney disease in Sri Lanka: A profile of affected districts reliant on ground water.

[5] Waterborne Diseases by Dr Manjula Rathnabharathie cover story 02, Volume 29, Number 03 Vidurava available at: http://dl.nsf.ac.lk/bitstream/handle/1/9962/VIDU_29%283%29_19.pdf?sequence=2&isAllowed=y

[6] Sri Lankan Journal of Infectious Diseases - Contaminated Public well water as a source of sporadic outbreak of enteric infection in Northern Sri Lanka.

[7] Linkoping University, Department of Physics, Chemistry and Biology (IFM), available at: http://www.ifm.liu.se/applphys/biosensors-and-bioelectro/group-members/fredrik-winquist/

[8] Abhishruti Bhuyan, Bipan Tudu, Rajib Bandyopadhyay, Arunangshu Ghosh, and Sanjeev Kumar, ARMAX Modeling and Impedance Analysis of Voltammetric E-Tongue for Evaluation of Infused Tea, IEEE SENSORS JOURNAL, VOL. 19, NO. 11, JUNE 1, 2019.

[9] Palash K. Kundu, Madhusree Kundu, A comparative study on tea samples classification by E-Tongue using Cross correlation based PCA , PCA and Shamon's Nonlinear mapping method, 2012 7th International Conference on Electrical and Computer Engineering.

[10] Palash K. Kundu, Madhusree Kundu, Classification of Tea Samples Using SVM as Machine Learning Component of E-Tongue, 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI).

[11] Shafi, U.; Mumtaz, R.; Anwar, H.; Qamar, A.M.; Khurshid, H. Surface Water Pollution Detection using Internet of Things. In Proceedings of the 2018 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT), Islamabad, Pakistan, 8–10 October 2018; pp. 92–96.

[12] Sakizadeh, M. Artificial intelligence for the prediction of water quality index in groundwater systems. Model. Earth Syst. Environ. 2016, 2, 8.

[13] Nabeel M. Gazzaz, Mohd Kamil Yosoff, Ahmad Zaharin Aris, Hafizan Juhair, Mohammad Firzul Ramil - Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors, Marine Pollution Bulletin 64 (2012) 2409–2420.

[14] Umair Ahmed, Rafia Mumtaz , Hirra Anwar , Asad A. Shah, Rabia Irfan and José García-Nieto, Efficient Water Quality Prediction Using Supervised Machine Learning, Water 2019, 11, 2210; doi:10.3390/w11112210.

[15] Tests with laboratory details, National Water Supply & Drainage Board, available at: https://drive.google.com/file/d/1D7IGpSsKe1Hs8OTwuFWr0n5PgGB7yOwC/view