ASHESI UNIVERSITY COLLEGE

FLOOD PREDICTION SYSTEM

CAPSTONE PROJECT

B.Sc. Computer Engineering

Nunana Elorm Togo

2020

ASHESI UNIVERSITY COLLEGE

FLOOD PREDICTION SYSTEM

CAPSTONE PROJECT

Capstone Project submitted to the Department of Engineering, Ashesi University College in partial fulfilment of the requirements for the award of Bachelor of Science degree in Computer Engineering.

Nunana Togo

2020

DECLARATION

I hereby declare that this capstone is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:

……………………………………………………………………………………………

Candidate's Name:

……………………………………………………………………………………………

Date: ……………………………………………………………………………………………

I hereby declare that preparation and presentation of this capstone were supervised in accordance with the guidelines on supervision of capstone laid down by Ashesi University College.

Supervisor's Signature:

……………………………………………………………………………………………

Supervisor's Name:

……………………………………………………………………………………………

Date: ……………………………………………………………………………………………

Acknowledgements

To my supervisor, Mr Kofi Adu-Labi whose encouragement and academic advice helped me undertake this project. I thank you very much for your constant encouragement and feedback without which I would not have been able to progress as far as I was able to.

To my family, whose support helped me to persevere in the face of challenges and difficulties during the undertaking of this project. I thank you from the bottom of my heart for without it I would have been able to believe in my ability to produce a work of this nature.

To all my friends, I thank you very much for the alternative ideas and perspective you provided during the progress of the project.

It is because of all these people that this project has come to completion. If any one of these individuals were missing during the execution of this project, it would not have been as successful as it is.

Abstract

Floods have been a problem Ghana has been facing without any kind of solution for a long time. As the years progress the consequence of an occurrence of floods has increased as the amount of infrastructure built in the country increased. In other parts of the world, ways of mitigating the damage floods cause to the citizens of the country has been researched with different methods providing different amounts of success. However most of the methods used in these countries are not applicable to developing countries such as Ghana due to the amount and type of resources they have available to them. In this project, a flood prediction system able to perform the task of predicting floods with the resources readily available is researched and designed. With the ability of knowing the possibility of a flood occurring the necessary disaster management authorities will be able to help reduce the amount of damage caused by the floods.

Table of Content

List of Figures

List of Table

Chapter 1: Introduction

1.1 Background

The country has faced many challenges over the past decade, from fire outbreaks to vehicular accidents and more. However, one of the challenges that Ghana faces consistently without any alleviation is the issue of floods. Over the years, Ghana has experienced the devastation floods wrought on the lives of individuals and the nation. The devastation experienced is indiscriminate of social standing, age, and gender.

The first item explored is the history of floods that has plagued Ghana for so many years. On May 5, 2010, Central Accra, Ofankor and Begoro were deeply submerged in water after two hours of stormy rains.

On June 22, 2010, the nation’s worst flood disaster occurred with a death toll of 35. On June 24, 2010, three bridges connecting the Agona Swedru municipality to neighbouring communities collapsed because of the flooding. The National Disaster Management Organization (NADMO) registered at least 3,000 flood victims in Agona Swedru [1].

On October 14, 2010, the flood displaced 161,000 people across the country during torrential rain and the opening of the Bagre Dam in Burkina Faso. On October 18, 2010, fifty-five (55) communities in the Central Gonja District in the Northern Region, including parts of the district capital, Buipe, were submerged by flood following the overflow of the Volta Lake [1].

On November 2, 2010, two thousand and eight hundred (2,800) people in 120 villages and towns along the Volta Lake in the Kwahu East, Kwahu South, and Kwahu North districts in the Eastern Region were rendered homeless by floods, destroying 850 buildings, farms, markets, and roads [1].

On February 24, 2011, a downpour wreaked considerable havoc on properties in most parts of Accra and some of its surrounding communities [1].

On July 20, 2011, about 10 hours of torrential rain left 105 farmers stranded on farms for three days while it drowned five (5) persons at Akyem Osoroase Krobomu in the Atiwa District in the Eastern Region [1].

On November 1, 2011, a downpour occurred in Accra that affected 43,087 people with 14 deaths recorded by the National Disaster Management Organization (NADMO) [1].

On May 31, 2013, heavy rains caused flooding in some parts of Accra such as the Kwame Nkrumah Circle, Darkuman Kokompe, the Obetsebi Lamptey Circle and portions of the Graphic Road, Santa Maria and the Dansoman Roundabout [1]

On June 6, 2014, Accra's poor planning was exposed when a deluge hit the national capital after more than 10 hours of downpour. The heavy rains caused flooding in the city and its environs, including Adabraka, Awoshie, the Kwame Nkrumah Circle, Mallam, North Kaneshie, Abeka, Dansoman and Odorkor [1].

In June 2015, a torrential downpour in Accra claimed over 152 lives as a GOIL Fuel Station exploded at the Kwame Nkrumah Circle [1].

The list above only gives details into a small portion of the different floods experienced by Ghana in the past decade. According to [2], the effects of flooding has been estimated to have been 120,200,000 dollars' worth of economic damage, affected 4,995,149 people and has led to the deaths of about 519 deaths [2].

1.2 Importance of Solving the Problem

Flooding is a problem because it has a national resource cost. These resources could be used in other developmental sectors. The resources in question include both human and material as both property and human lives are destroyed whenever this phenomenon occurs. The information provided above alone makes it a significant problem that needs to be if not fixed immediately have some form of measures to mitigate the effect it has on the economy and the people.

1.3 Project Objectives

The project focuses on the development of an accurate, reliable and consistent way of preventing the occurrence of floods from negatively affecting individuals in the community. A system that will allow for individuals who are in areas prone for this natural disaster to have a way of being forewarned as well as reducing the loss of life experienced any time flooding occurs. The system should be able to increase the efficiency of institutes such as National Disaster Management Organization (NADMO) in responding to such a disaster and putting in place protocols that would increase the probability of being able to help those affected better. In doing so, the system will be able to allow for more efficient systems to be put in place to handle such cases.

1.4 Proposed Solution

The solution described in this paper will be a flood prediction system which would provide early warning to the individuals within a community and inform relevant institutions needed to provide necessary aide to the saving of lives and properties of those affected by the flood. This solution will allow for the collection of important data as well as better map out the risk zones within the country.

Chapter 2: Literature Review

The creation of a possible solution to the flood problem depends on the study of different kinds of solutions put in place to battle similar disasters in different parts of the world. Furthermore, an analysis of these solutions would prove satisfactory in solving the problem plaguing the nation.

The question of how to create a flood prediction system to provide accurate information to the affected community has been debated for several years. Many different ideas have come about, and the progress of technology has allowed for new ideas to surface. This section of the paper seeks to explore the different ideas that have come about.

Before going into how the system would work, many papers have discussed the issue of the cost of the system, especially in developing countries. The paper [3] discussed the different implementation of such a system and noted that a design implemented in developed countries such as the USA is not feasible in developing countries. The absence of resources or institutional structures that handle the different aspects of the system such as getting the information to the public promptly, the monitoring of several areas, and providing reliable maintenance of the system [3]. Kugler and Groeve also made mention of this fact. Developed countries make use of real-time reporting of extreme precipitation and other surface meteorological variables from in situ, radar and in some cases, satellite observations [4]. The stated observations have made it clear. There is no tried and tested prediction system, which is not expensive. Though this is the case, cutting down on cost should not cause a decrease in the accuracy of the system for it is meant to save lives.

In the area of cost-efficient methods of providing a prediction system for floods, Basha and Rus came up with the creation of sensor networks along the Honduras River. The sensor network would take in the measurements of river pressure, rainfall and temperature. The prediction model will use the data to make predictions of when the flood would occur. The output would then be compared with a predefined threshold of what a flood is before notifications would then be sent out [3].

While Basha and Rus provided a solution implemented with a sensor network on a local scale, Kugler and Groeve's method of predicting flood was on the global scale. In the methodology, they described their method of flood prediction as water surface change detection. The method was explained to use a modified version of technique first developed at Dartmouth Flood Observatory where a microwave radiation difference of the land and water was used to detect flood [4]. Once this kind of detection system is up, it would be able to serve many different locations. However, it does not have the element of prediction only detection. The second flaw of this system is the duration between scans in different places; hence, it would not be ideal for any rapid response action.

Another method was the rain-induced disaster alarm system. The method qualifies to be made mention of because the majority of floods are caused by extreme rainfall in any given area for a prolonged period. Zamora and Ching et al. proposed the use of existing infrastructure to measure rain-induced data as opposed to the traditional methods of using rain gauges and other sensors. The method brought forward was the use of a network of wireless such as microwave links. For the past three years, the group had gathered several rain-induced data from 26 GHz microwave link and acoustic recorders to better understand the characteristics of tropical rain in the sub-kilometre scale [5]. The data collected would then be transmitted to a central database for analysis. The measurement of rain data is made possible due to the attenuation, which occurs when a signal passes through rainy conditions. Better measurement is obtained by getting the acoustic power of the rain by acoustic sensors in the form of low-cost MP3 recorders [5]. The method described is based on the idea that there exists an already existing wireless infrastructure that covers all the necessary areas required, which most at times is the case.

Hung Ngoc and Minh-Thanh et al designed a flood prediction system using an ARM microcontroller specifically an ATmega chip [6]. The microcontroller would obtain data from water level and rain gauge sensors which are stored in a memory card and later transmitted via GPRS and GSM to a central centre which would be in charge of monitoring and analysis. The individuals obtain information via SMS [6].

All possible solutions provided have one core issue; the accuracy of the prediction will always be dependent upon the prediction model used in the analysis of any form of data obtained. Some use statistical analysis. Others use real-time data prediction as in the case of Hung Ngoc and Minh-Thanh system, which is solely dependent on the data that is collected by the system in real-time. Then there is a fusion of the two where historical data is used in conjunction with real-time data collection to provide better accuracy when it comes to the topic of flood prediction. The case of Basha and Rus, where historical data was needed before any prediction could form so the model could adjust is such an instance. Hence, a discussion on what kind of model used in the flood prediction model is needed.

Flood prediction models are categorized into two different categories, long term flood prediction ranging from days to weeks to months and short term flood prediction, which deals with the range of hours. According to Mosavi, Ozturk, and Chau physically-based models often require various types of hydro-geomorphological monitoring datasets, requiring intensive computation which prohibits short-term prediction [7]. For a reliable long term prediction, at least, a decade of data from measurement gauges must be analysed [7]. The idea of having at least a decade of data to provide an accurate long term prediction is reasonable because such a large dataset would allow for the minimization of outliers having a tremendous impact in the creation of the model. To reduce the amount of complex computation that is required to develop a prediction model, the number of inputs into the model is required [7]. The reduction in the number of inputs would reduce the amount of computational power that would be needed as well as increase the response time of the model. In this regard, data-driven prediction models using machine learning are practical tools since it is quicker to develop with minimal inputs [7]. "ML is a field of artificial intelligence used to induce regularities and patterns, providing easier implementation with low computation cost, as well as fast training, validation, testing, and evaluation, with high performance, compared to physical models, and relatively less complexity"[7].

Chapter 3: Design

3.1 User Requirements

For the system to be able to perform the function that it will be designed for, certain requirements must be met. After conducting research the following requirements were crafted. These requirements will allow for individuals within the society as well as the authorities involved to carry out necessary actions to save lives and properties.

- Should be able to deliver warning of flood with enough time for preventive measure to be taken. (2 hours minimum)
- Should be able to notify both community and authority of impending flood.
- The error of false prediction should be less than three percent
- Should be easy to maintain
- Should be able to obtain relevant data (water level, pressure, moisture levels etc.) for future solution refinement.

3.2 Design Requirements

The system itself must be able to perform under different conditions hence certain requirements must also be met before the system will be able to work efficiently in the environment that it will find itself in. An analysis of the environment that the system will be deployed in created the requirements listed below.

- Should be deployed quickly
- Should be energy efficient

- Should report changes to a remote monitor
- Should be able to operate in water and wet conditions
- Should be cost efficient as it would be deployed to many areas
- Should be able to last without replacement for about two years
- Should be able to handle different weather conditions

3.3 High Level Design

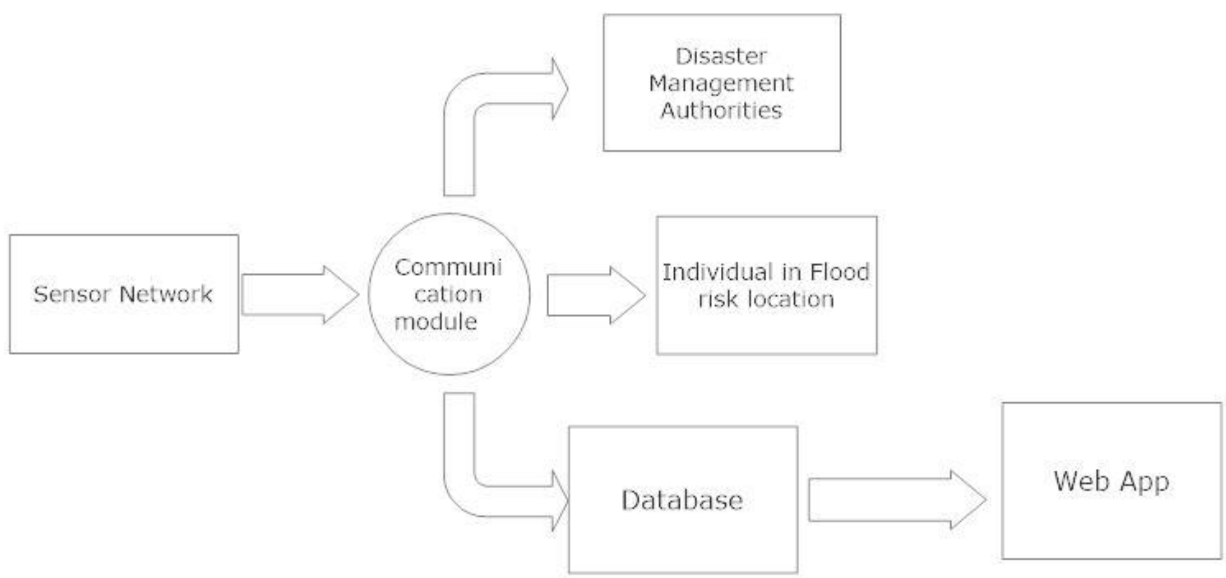

Figure 3.1 System Functional Diagram

Figure 3.1 shows a high level view of how the entire system is going to operate. The sensor network which is spread across the country is responsible for gathering environmental data which is then passed onto the communication module. The communication module is responsible for sending the data that is gathered from the sensor network to three different recipients; the data base which keeps storage of all the data that is collected by the sensor network, the individuals who are at risk of an approaching flood, and the authorities that are in charge of disaster management with locations that are at risk of a flood in the near future. The Database then feeds the collected data to a web app that is responsible for analysing the collected data from the different locations the sensor network is deployed at.

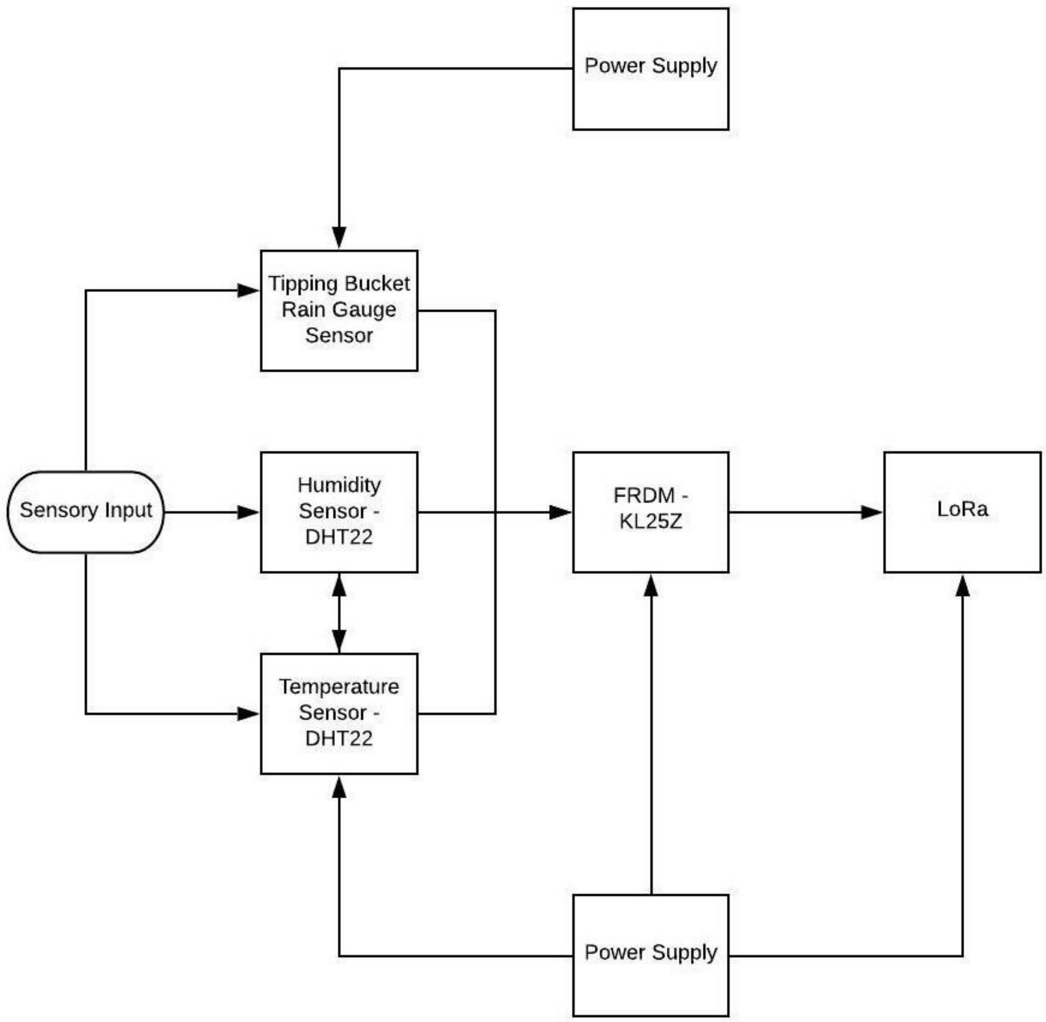

Figure 3.2 Functional Diagram of Sensor Object

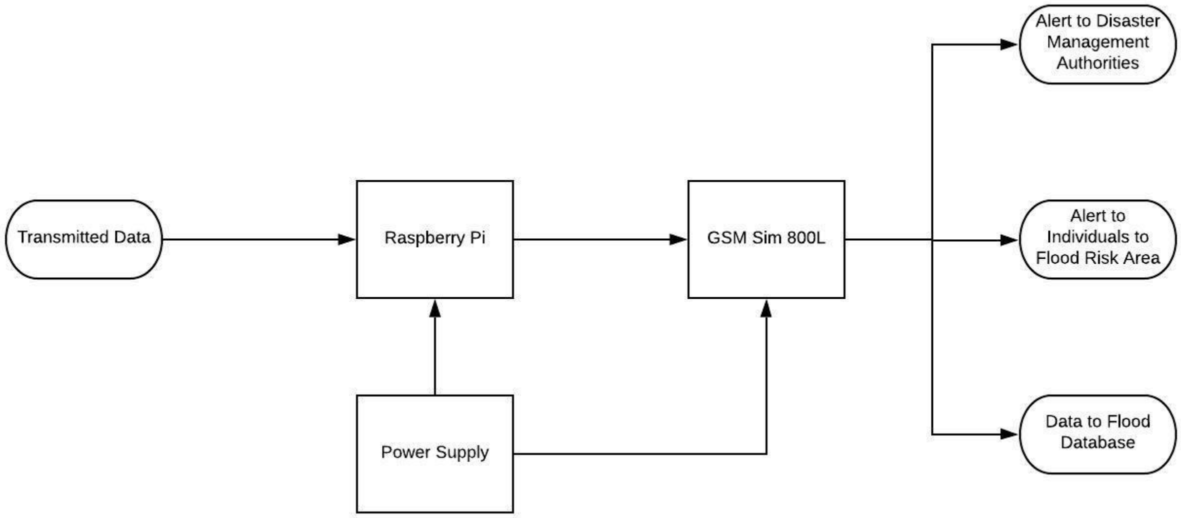

Figure 3.3 Functional Diagram of Sensor Network Hub

Zooming into the device that makes up the sensor network, figure 3.2 is a proposed block diagram of the components that makes up the environmental data collection system. It is made up of four sections, the sensor section which is responsible for capturing environmental data, the microcontroller section which is responsible for processing the data that is collected from the sensors into meaningful information, the communication module which is responsible for the delivery of the data from the sensor object to the sensor network hub and the power module responsible for providing power to the entire system.

All sensor object transmit their data to a Raspberry Pi which contains the predictive algorithm responsible for detecting the possibility of a flood in the area. The microcontroller is also responsible for further processing of data so as to prepare it for transmission to a database for further analysis. The alert of a possible flood when detected is sent to the authorities in charge of disaster management and the individuals in the flood risk zone. The last section is the power supply which is responsible for providing power to all the sections of the sensor object and network hub.

3.4 Components and Technologies

In deciding on the types of components to use for the different sections, a Pugh chart was utilized. The Pugh chart is one of the most common ways of selecting one out of several choices given some specific criteria. Each of the components will be compared based on the criteria refined from the requirements stated in 3.1 and 3.2. When the criteria being compared is the same the value given will be 0, when it is worse then it will be the negative of the weight of that criteria and when better, it will be the positive of the weight of that criteria.

3.4.1 Sensor Selection

The physical quantities being measured are temperature, humidity and rainfall. For temperature the sensors being compared are DHT22, and LM35

Table 3-1 Temperature Sensor Pugh Chart

		Weight	DHT22	LM35
Criteria				
Cost		2	-2	+2
Energy Efficiency		3	+3	-3
Precision		4	+4	-4
Sum			5	-5

The sensors being compared for humidity are DHT22, and HR202

Table 3-2 Humidity Sensor Pugh Chart

		Weight	HR202	DHT22
Criteria				
Cost		2	+2	-2
Energy Efficiency		3	-3	+3
Precision		4	-4	+4
Sum			-5	+5

Taking the two Pugh charts together it will be better to use the DHT22 as the temperature and humidity sensor due to the fact that it measures both quantities hence no extra cost is incurred from adding another sensor. Also DHT22 has a temperature range between -40 and 125 degree Celsius.

The tipping bucket rain gauge was chosen to perform the act of measuring amount of rainfall because the device is going to be deployed at an area where an individual will not need to constantly empty the bucket. Hence the tipping bucket mechanism will allow for the rain gauge to measure the rainfall amount continuously without the need of human intervention to empty the bucket when it is full.

3.4.2 Communication Module

The communication module is one of the most important aspect of the system. Without the correct communication module the entire system will not be able to perform the task it was meant to. Therefore choosing the correct communication module is very important. What must be taken into consideration is where the sensor object will be deployed at as well as the distance the data has to cover before it arrives at its destination. This informed the criteria under which the selection of the communication module was made.

Table 3-3 Communication Module Pugh Chart

		Weight	GSM sim800L	LoRa SX1278
Criteria				
Cost		2	+2	-2
Energy Efficiency		3	-3	+3
Distance		4	+4	-4
Sum			3	-3

From the comparison between the two communication modules the GSM sim800L module is better in this case due to distance being a key criteria for communicating data to where it needs to be. With the GSM module as long as a cellular network is available in the area, transmission of data is not an issue. However for communication between sensor object and the sensor network hub within the same zone LoRa is used to reduce cost of transmitting data from the sensor object to the sensor network hub. Due to the LoRa being able to transmit data in the range of 15 km the sensor object can be spread to cover an area of about 500 kilometre squared circumference if the range is taken as the radius with the central hub as the centre of the circle.

3.4.3 Microcontroller Unit

The brain of the entire operation is the microcontroller unit. Without it, processing the information obtained by the sensors is impossible. The criteria used in judging which MCU to use are the cost, processing power and energy efficiency.

Table 3-4 Microcontroller Pugh Chart

		Weight	ARDUINO UNO	FRDM KL25Z
Criteria				
Cost		2	-2	+2
Energy Efficiency		3	-3	+3
Processing Power		4	-4	+4
Sum			-9	+9

From the Pugh chart FRDM KL25Z is the best to work with. Due to lower cost, higher processing power and energy consumption the MCU will be able to handle the computation that will be carried out. Also the more precise clock count will allow for better tracking of the values captured by the sensors hence allow for the database to have more details in regards to the physical quantities measured. In carrying out the predictive analysis of all the data acquired at the edge of the network, a microcontroller capable of performing machine learning is needed. Since the most tried and tested machine learning platforms are based on Python a Raspberry Pi is the most suitable single-board computer to act as the hub of the sensor network. Also the ability to remotely connect to the Raspberry Pi will also allow for necessary upgrades to be performed whenever a better algorithm for predicting flood occurrence is created.

Chapter 4: Design Implementation

4.1 Hardware Implementation

The hardware section of the system as mentioned in chapter three is made up of three main sections excluding the power module. The sensor module, processor module and the communication module.

4.1.1 Schematic

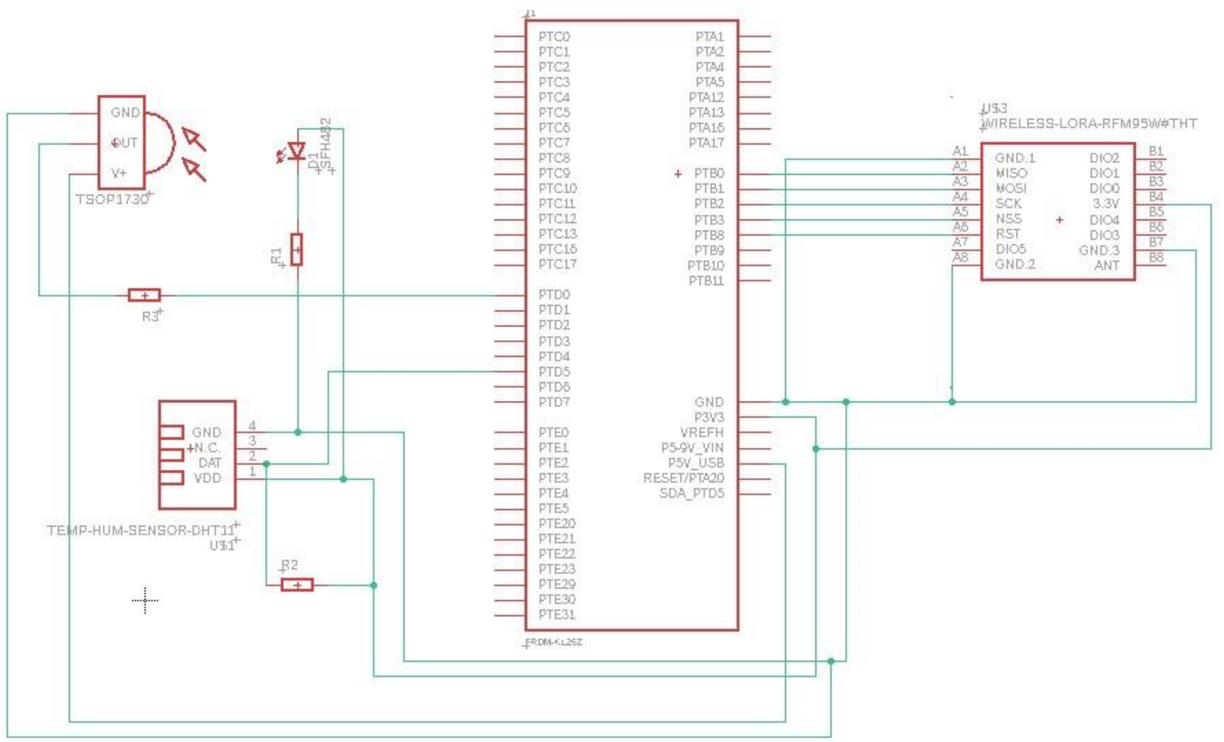

Figure 4.1 Circuit Diagram of Sensor Object

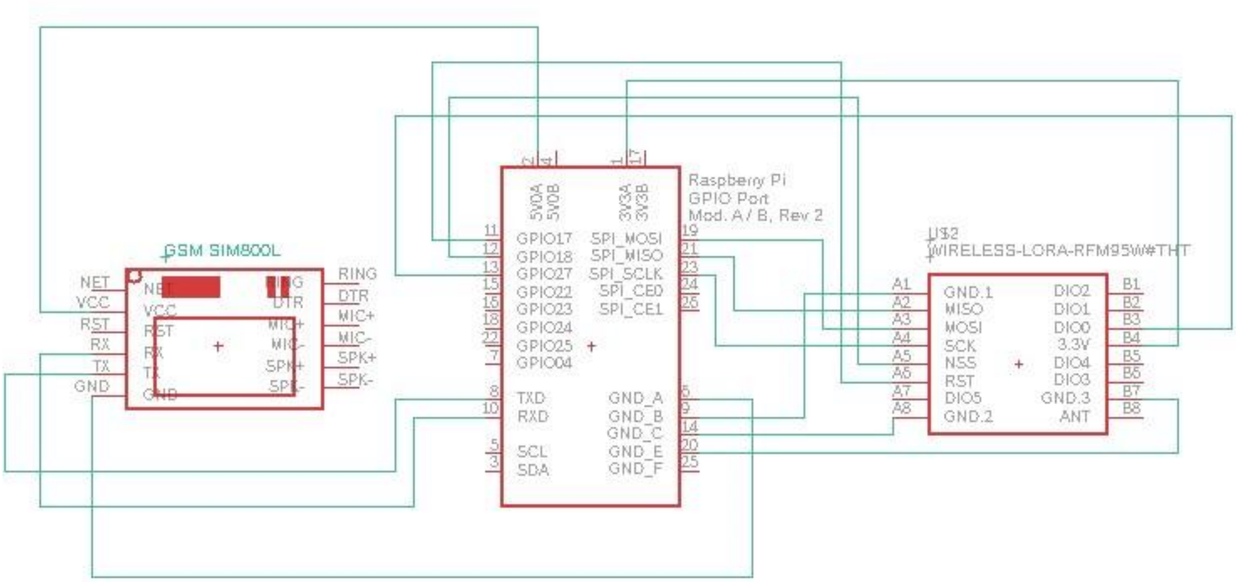

Figure 4.2 Circuit Diagram of Sensor Network Hub

4.1.2 Sensor Module Setup

The sensor module is made up of two sensors, DHT22 which records the temperature and humidity in the area every two seconds and the tipping bucket rain gauge. For the DHT22 temperature and humidity sensor an hourly average will be calculated for transmission. The tipping bucket rain gauge on the other hand is made up of two parts working together. The first part is the bucket connected to a stepper motor which turns one hundred and thirty five degrees any time it has filled up to empty its content. In order to detect if the container is full or not an IR sensor will be place at the edges of the container and when it has filled up it will break the connection hence serving as a trigger for the motor. Any time it triggers a count is kept and when an hour has passed the number of counts is multiplied by the volume of the container and that will serve as the rainfall rate.

4.1.3 Processing Module Setup

The processing module is divided into two different sections. The first part is found at the edge of the network. The sensor object contains a FRDM KL25Z to facilitate the collection of sensory data from the sensor with high resolution. Initial formatting of data is done in the sensor object with the help of the microcontroller before it is transmitted to the Raspberry Pi which acts as the fog node or the sensor network hub. The Raspberry Pi is the second part of the processing module. It is here that the data collected within a specific area is accumulated for further processing. Time sensitive prediction by the machine learning algorithm is carried out in the Raspberry Pi since as mentioned before, the python language contains many modules that allows for machine learning.

4.1.4 Communication Module Setup

For any form of data to be of use to any one, the information must be able to move from point A to point B. The communication module is responsible for this to occur. The communication between the sensor object and sensor network hub is different from the communication between the sensor network hub and the recipients of the information. Therefore, communication has been divided into two different parts using two different technologies.

Between sensor object and sensor network hub, LoRa SX1278 is used. The reason for this is because multiple sensor objects will be deployed within a community allowing for a wide coverage of the area. The LoRa will allow for each of these sensor objects to communicate with the central hub in that location without the issue of distance between them. For the communication between the sensor network hub and the recipients comprising of the database, disaster management authorities and individuals, a GSM module will be used. The GSM module possesses the ability to communicate with both the internet and any cellular device that has a signal. This ability will allow for the sensor network hub to send information disregarding the distance between itself and any of the recipient. A notification will be sent to both the disaster management authorities and individuals via SMS. Data accumulated in the sensor network hub will be sent to the data base via GPRS.

4.2 Software Implementation

4.2.1 Sensory Information Collection

The collection of sensory information is done at the edge of the network by the FRDM-KL25Z micro-controller which uses C as its programming language. The temperature and humidity data gathered from the sensors are averaged over a duration of thirty minutes and stored for transmission. The amount of rainfall measured by the tipping bucket rain gauge is used to obtain the rate of rainfall by dividing the amount of rainfall by a given duration which in this case is five minutes and extrapolate the amount of rainfall in the next two hours. The current amount of rainfall is also stored so as to keep amount of rainfall estimated in the next thirty minutes up to date. All these data are stored and then transmitted on a regular interval.

4.2.2 Information Processing

Information processing takes place in the fog node. The fog node is the central hub which all sensor objects transmit data to and is created with the help of a Raspberry Pi. Since the processing unit is a Raspberry Pi, the programming language used is Python. Information processing is split into three distinct portions. The first portion is where the data received is formatted correctly for processing, the second part is where the data received is passed through the machine learning algorithm in charge of performing the key part of the system, flood prediction, and the third portion is where the output is generated in order to transmit to the necessary recipient.

In order to get a good machine learning algorithm for the purpose of flood prediction, two different models were created and tested. The Python scikit-learn machine learning library was used in the creation of these models. Since the problem being solved is to determine if a flood would or would not occur at any given time the model created was a classification model. In the creation of the classification models, two different models were used, support vector machine (SVM) and multilayer perceptron (MLP) which is a feed forward artificial neural network.

4.2.3 Database

MySQL was used in the creation of the database responsible for the accumulation of data gathered by the various sensor objects that would be deployed. The Database is made up of three tables that are connected together. The first table is a table populated by the locations where the sensor objects have been deployed. The second table is a table of properties dealing with how close or far the location is to a water way. The third is a table that brings these two tables together as well as the amount of rainfall, the temperature, the humidity, the time and date, the population size of the location, and whether a flood has been predicted to occur at that location or not. The essence of the database is to be able to collect historical data on the location to help fine tune the system. It will also provide data to help drive further detailed analysis in order to combat the occurrence of flood at the various locations.

Chapter 5: Results and Analysis

The entire system hinges on the fact that the prediction algorithm created is able to predict the occurrence of a flood. Hence, it is the most important aspect of the system. In order to test this a test data was used. The test data is made up of rainfall, humidity, and temperature data from Albury, Australia. The data was acquired from [9] and the date when flood occurred in the location was acquired by mining data from [8]. Both these sites are government sponsored and hence the reliability of data acquired from them is high.

```
Index(['Year', 'Month', 'Day', 'Rainfall_mm', 'MaxTemp_C', 'Humidity',
       'Flood'],
      dtype='object')
       Rainfall_mm    MaxTemp_C     Humidity
count  3010.000000  3010.000000  3010.000000
mean      1.891096    22.826744    74.121262
std       6.195572     7.829814    17.440882
min       0.000000     6.800000    18.000000
25%       0.000000    15.900000    61.000000
50%       0.000000    22.200000    76.000000
75%       0.400000    29.100000    88.000000
max     104.200000    44.800000   100.000000
```

Figure 5.1 Albury Climate Data Description

The figure above displays the description of the data set that was used in the testing the flood prediction algorithm. Since only one location was involved there was no need to include the location when training the algorithm.

```
[[746   2]
 [  3   2]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       748
           1       0.50      0.40      0.44         5

    accuracy                           0.99       753
   macro avg       0.75      0.70      0.72       753
weighted avg       0.99      0.99      0.99       753
```

Figure 5.2 Confusion matrix and classification report of MLP prediction model

After training and testing the MLP model, the above result was obtained. As can be seen from the classification report, accuracy is 99%. However, looking at the confusion matrix, it can be seen that majority of these appear to be from where the model predicted a true negative whereas the number of positive is 5 out of which only two were predicted to be true positive. This was obtained by using the train test and split method where the data is divided into 60% training and 40% for testing.

Using cross-validation is a better method of determining the reliability of a model. When cross-validation was applied to the MLP model the following accuracy scores were obtained. When the mean was calculated, the accuracy score obtained was 0.9940199335548172.

```
[0.99667774 0.99667774 0.99667774 0.99003322 0.99335548 0.99335548
 0.99335548 0.99335548 0.99335548 0.99335548]
```

Figure 5.3 MLP cross-validation accuracy scores

```
[[747   1]
 [  4   1]]
              precision    recall  f1-score   support

           0       0.99      1.00      1.00       748
           1       0.50      0.20      0.29         5

    accuracy                           0.99       753
   macro avg       0.75      0.60      0.64       753
weighted avg       0.99      0.99      0.99       753
```

Figure 5.4 Confusion matrix and classification report of SVM prediction model

When looking at the prediction made by the SVM prediction report, it can be seen from the classification report that the accuracy is also 99%. However, the difference appears when looking at the confusion matrix between MLP and SVM prediction model. Whiles the number of true positive recorded by MLP is 2 out 3, the number of true positive recorded by SVM model is only 1 out of 4. This difference when looked at on such a small scale does not look as much but when it is scaled up the amount of false positives predicted will affect the reliability of the model.

When cross-validation was applied to the SVM model, the accuracy figures that was reported was similar to that of MLP model with an accuracy of 0.9943521594684386.

```
[0.99667774 0.99667774 0.99667774 0.99335548 0.99335548 0.99335548
 0.99335548 0.99335548 0.99335548 0.99335548]
```

Figure 5.5 SVM cross-validation accuracy scores

Taking these two different prediction model used, the best between them would be MLP model. It must be noted that for any kind of machine learning algorithm, the kind of data that is fed to the model determines the reliability of model. Hence, even though the MLP model provided a better result to that of SVM model, just a single data set used in the testing of the model does not provide a comprehensive analysis of the models. The code for the two separate predictive models can be found in the appendix.

Chapter 6: Conclusion

6.1 Discussion

The main objective of the project was to create a flood prediction system which would be deployed in the urban areas of Ghana so as to first of all provide a warning to its inhabitants to the occurrence of floods in the area and second of all provide the relevant agencies the data needed to create plans and policy to prevent the occurrence of floods in these areas. A system which would be able to cover all necessary areas without requiring technologies that is not easily accessible.

Such a system was designed successfully, however a working prototype was not created due to inaccessibility of components as a result of the COVID-19 outbreak. Hence the concentration of energy towards the prediction algorithm which is the component on which all else relies on. In looking at the prediction model, the features used were limited to those discussed in the paper due to the fact that the other features that are related to the occurrence of floods in any area are hard to measure via sensors because of the nature of physical quantity to be measured. Examples of such features include, the topology of the area, and the nature of the soil found in the area.

6.2 Limitation

The system as it is, is limited by the kind of data that is used to train the model. This means that in areas where the needed data is not available the system will not be able to work. Also, the system will only predict outcome in relation to what it is trained with. The result of this means that if the initial data that used to train the system was flawed, the prediction made by the system will also be flawed. For the system to perform as intended, the accuracy of the data used to train the system must be high.

Another limitation of the system is that, it will only work in areas where cellular coverage is available. The only way the system can communicate with the database, relevant individuals and groups is via SMS and GPRS. Hence, the system cannot deployed without taking this into consideration.

6.3 Future works

The system as it at the moment can be improved significantly. One of the improvement that can be done to the system is the addition of a GPS system to the sensor objects to record the location of the sensor objects. When this is combined with a flood risk assessment map more data is obtained in determining the likelihood of a flood occurring at an area or not. Furthermore, a method of determining the status of each of the sensor objects will also go a long way in allowing for efficient maintenance of the system. The reason for this is because the number of sensor objects deployed when scaled over a country will be a lot and therefore the ability to pinpoint where maintenance is needed will cut down on the cost of hiring individuals to constantly check the system in person.

References

[1] S. Asumadu-Sarkodie, P. A. Owusu, and P. Rufangura, 'Impact analysis of flood in Accra, Ghana', p. 3625108 Bytes, 2016.

[2] EM-DAT: The Emergency Events Database - Université catholique de Louvain (UCL) - CRED, D. Guha-Sapir - www.emdat.be, Brussels, Belgium

[3] E. Basha and D. Rus, 'Design of early warning flood detection systems for developing countries', in *2007 International Conference on Information and Communication Technologies and Development*, Bangalore, India, 2007, pp. 1–10.

[4] Z. Kugler, 'The Global Flood Detection System', p. 45.

[5] J. L. F. Zamora *et al.*, 'Rain-Induced Disaster Alarm System Using Microwave and Acoustic Sensing', in *2011 IEEE Global Humanitarian Technology Conference*, Seattle, WA, USA, 2011, pp. 437–441.

[6] H. N. Do, M.-T. Vo, V.-S. Tran, P. V. Tan, and C. V. Trinh, 'An early flood detection system using mobile networks', in *2015 International Conference on Advanced Technologies for Communications (ATC)*, Ho Chi Minh, Vietnam, 2015, pp. 599–603.

[7] A. Mosavi, P. Ozturk, and K. Chau, 'Flood Prediction Using Machine Learning Models: Literature Review', *Water*, vol. 10, no. 11, p. 1536, Oct. 2018.

[8] "Disaster Assist," *Disaster Assist*. [Online]. Available:https://www.disasterassist.gov.au/. [Accessed: 11-May-2020].

[9] "Australian Government Bureau of Meteorology," *Australia's official weather forecasts & weather radar - Bureau of Meteorology*. [Online]. Available: http://www.bom.gov.au/. [Accessed: 11-May-2020].

Appendices

Appendix A: MLP Prediction Model Code Documentation

```
"""
Created on Fri Apr  3 13:30:51 2020

@author: Nunana E. Togo
"""
import pandas as pd
from sklearn.model_selection import train_test_split as tts
from sklearn.neural_network import MLPClassifier as mlp
from sklearn.metrics import mean_absolute_error as mae
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report,confusion_matrix
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score

FloodDataFilePath = "ClimateDataCSV.csv"
FloodData = pd.read_csv(FloodDataFilePath)
FloodData.fillna(0)
print(FloodData.columns)



Features = ['Rainfall_mm', 'MaxTemp_C', 'Humidity']

X = FloodData[Features]
print(X.describe())

y = FloodData.Flood
print(y.head)
```

```
# Using Train Test Split method of validation

train_X, val_X, train_y, val_y = tts(X,y,random_state=1)

print(train_X.describe())

scaler = StandardScaler()


scaler.fit(train_X)

train_X = scaler.transform(train_X)


Climate_model = mlp(hidden_layer_sizes = (1000,500),learning_rate_init
= 0.005,)

Climate_model.fit(train_X, train_y)


val_X = scaler.transform(val_X)

val_Prediction = Climate_model.predict(val_X)


print(confusion_matrix(val_y, val_Prediction))

print(classification_report(val_y,val_Prediction))

val_mae = mae(val_Prediction, val_y)

print(val_mae)


# Using K_fold Cross Validation method of validation

Kf = KFold(n_splits = 20, shuffle = True, random_state = 1)

KF_x = scaler.transform(X)

scores = cross_val_score(Climate_model, X, y, cv = 10, scoring =
'accuracy')

print(scores)

print(scores.mean())
```

Appendix B: SVM Prediction Model Code Documenation

```
"""
Created on Thu Apr 30 08:01:36 2020

@author: Nunana E. Togo
"""
import pandas as pd
from sklearn.model_selection import train_test_split as tts
from sklearn.svm import SVC
from sklearn.metrics import mean_absolute_error as mae
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report,confusion_matrix
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score

FloodDataFilePath = "ClimateDataCSV.csv"
FloodData = pd.read_csv(FloodDataFilePath)
FloodData.fillna(0)
print(FloodData.columns)



Features = ['Rainfall_mm', 'MaxTemp_C', 'Humidity']

X = FloodData[Features]
print(X.describe())

y = FloodData.Flood
print(y.head)

train_X, val_X, train_y, val_y = tts(X,y,random_state=1)
```

```
print(train_X.describe())
scaler = StandardScaler()

scaler.fit(train_X)
train_X = scaler.transform(train_X)

Climate_model = SVC()
Climate_model.fit(train_X, train_y)

val_X = scaler.transform(val_X)
val_Prediction = Climate_model.predict(val_X)

print(confusion_matrix(val_y, val_Prediction))
print(classification_report(val_y,val_Prediction))
val_mae = mae(val_Prediction, val_y)
print(val_mae)

# Using K_fold Cross Validation method of validation
Kf = KFold(n_splits = 20, shuffle = True, random_state = 1)
KF_x = scaler.transform(X)
scores = cross_val_score(Climate_model, X, y, cv = 10, scoring = 
'accuracy')
print(scores.mean())
print(scores)
```