ABSTRACT

Indian fisheries and aquaculture are an important sector of food production providing nutritional security, besides livelihood support and gainful employment to more than 14 million people and contributing to agricultural export. Aquatic organisms have specific tolerant range of various environmental parameters, thus fish farming of specific types of fish species requires certain conditions that must be met. Water quality is a critical factor while culturing aquatic organisms. In this project, we have built an application using ML for predicting water parameters and in turn monitor the fish farming ponds. The quality of water is predicted in an hourly manner to ensure growth and survival of aquatic life. A web application is built using Flask to alert the user in critical situations. The impact of water parameter change can be handled productively if the information is examined and water quality is anticipated in advance.

Keywords: Aquaculture, Machine Learning, LSTM, Root Mean Square Error, Flask.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

ANN	**Artificial Neural Network**
ARIMA	**Auto Regressive Integrated Moving Average**
DO	**Dissolved oxygen**
LSTM	**Long Short Term Memory**
ML	**Machine Learning**
MSE	**Mean Square Error**
pH	**Potential of Hydrogen**
RMSE	**Root Mean Square Error**
RNN	**Recurrent Neural Network**
SRU	**Simple Recurrent Unit**
SSVM	**Smooth Support Vector Machine**

CHAPTER 1

INTRODUCTION

1.1 Aquaculture

Research in aquaculture is an input to increase stabilized production. In last decade various scientists have made sustained efforts that resulted in development of modern production technologies that have revolutionized farm production. Fish farming have been used for more than three decades. Fish farming refers to farming variety of marine species such as shellfish, sport fish, bait fish, ornamental fish, crustaceans, mollusks, algae, sea vegetables, and fish eggs to breed, rear and harvest in different water environments such as ponds, rivers, lakes, and ocean. Fish are cold-blooded animals, regulating their body temperature directly by the water environment. Changes in water temperature affect the amount of dissolved oxygen in the water and fish oxygen consumption. Although the fish can withstand a broad water temperature range, any sudden, extreme changes in water temperature will have a considerable impact on fish physiology. A chilling injury will cause the fish to rush into, paralysis with a loss of balance, leading to death. The reason may be the respiratory center, or osmotic regulation is affected at high temperatures. As the water temperature increases the fish suffer respiratory arrest.

Fish World magazine found that the amount of dissolved oxygen in water increases or decreases with the seasons. When the water temperature rises, fish metabolic rate will be increased and results in less dissolved oxygen in the water. Low water temperature decreases fish metabolic rate and increases amount of dissolved oxygen in the water. If the amount of dissolved oxygen in water is reduced to below a certain limit fish growth will be hindered. When the amount of dissolved oxygen becomes lower than the fish survival conditions the fish will die. In general fish farming the acidity and alkaline of the water should be maintained between 6 to 8. Too acidic or alkaline will cause adverse effects, acid erosion of the gill tissue, tissue coagulation necrosis, increased mucus secretion, abdominal congestion and inflammation. If the PH value is less than 4.5, the fish will die. Table 1.1 shows the ranges of parameter tolerance of the species.

Table 1.1 Water Quality Tolerance by Species

Species	Temp °F	Dissolved Oxygen mg/L	pH	Alkalinity mg/L	Ammonia %	Nitrite mg/L
Baitfish	60-75	4-10	6-8	50-250	0-0.03	0-0.6
Catfish/Carp	65-80	3-10	6-8	50-250	0-0.03	0-0.6
Hybrid Striped Bass	70-85	4-10	6-8	50-250	0-0.03	0-0.6
Perch/Walleye	50-65	5-10	6-8	50-250	0-0.03	0-0.6
Salmon/Trout	45-68	5-12	6-8	50-250	0-0.03	0-0.6

Aquaculture is the fastest growing food production sector according to research, hence, to consolidate its efficiency and its sustainability. Water quality monitoring is the key to realize the success of Aquaculture. Aqua farmers are relying on manual testing for knowing the parameters of water, which is time consuming and inaccurate as parameters may alter with time.

To overcome this problem, modern technologies like IOT and ML should be brought to aquaculture. Thus, increases productivity and minimizes the losses by constant monitoring of water quality parameters. Water parameters that is needed to by monitored continuously are Temperature, Dissolved Oxygen, PH, turbidity, Salt, etc.

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people.

Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

Machine learning is a continuously developing field. Because of this, there are some considerations to keep in mind as you work with machine learning methodologies or analyze the impact of machine learning processes.

1.2 Problem Statement

The goal is to predict the water quality accurately and efficiently in an hourly manner, and to alert the user in advance. Regression analysis and prediction method through a large number of data samples for correlation analysis to obtain the correlation, and establish a regression equation, on the basis of considering the prediction error to determine the future water quality prediction value. Its model is complex, and the distribution of data and samples requires higher requirements.

In spite of the fact that ARIMA models are very adaptable to time series, their major impediment is that no non-linear can be captured by the ARIMA model. LSTM model is picked so as to build up a far-reaching approach for effective water quality prediction and analysis. LSTM can dynamically memorize and retain the historical water quality parameters information while learning new information.

1.3 Objectives

Our main objective is to increase the productivity of aquatic fishes and save its life.

- To increase the productivity of fishes, efficiently.
- To predict the water quality accurately, ML Algorithms are used.
- Predict the water parameters in an hourly manner. i.e., predict next hour values based on previous data.
- When predicted water quality seems to exceed the critical conditions alert is sent to user immediately to take the precautionary actions required.
- Build a web application using Flask to show alerts to the users as well as data visualization of the various water parameters considered.

1.4 Scope of the project

- The project focusses on the importance of water quality for aquaculture. It helps the aqua farmers to produce good quality of fishes, which in turn helps the economy of aquaculture sector.
- Machine Learning helps provide a better, accurate and faster prediction of quality of water based on the previous data collected.
- Predictive analysis can help to capture relationships among numerous variables that can help to assess risk with a particular set of conditions. LSTM NN is considered better and gives increasingly precise data to foresee and assess the water quality.
- In India, freshwater fishes like Rohu and Catla are predominantly grown through Aquaculture.

1.5 Methodology

To complete this project successfully agile methodology as shown in Table 1.2 is followed.

- Sprint 1: Literature survey is done to understand other related works in this domain and hence do the analysis of problem in the existing model.

- Sprint 2: Identifying the project goals, project requirements and formulating the problem statement. Collection of dataset and preprocessing.
- Sprint 3: Developing the models with various ML algorithms. These models are trained, tested and evaluated.
- Sprint 4: Choosing the best model and predicting the water parameters in hourly manner.
- Sprint 5: Developing web application for displaying the predicted values and alert.
- Sprint 6: Implementation of data visualization for better understanding of the parameter chances over the past hours.

Table 1.2 Software Agile Methodology

Story ID	Requirement description	User stories/Task	Description
1	Collection of water data with various parameters.	Data Collection	In .csv format.
2	Determining the relevant parameters needed for the water quality.	Cleaning the dataset.	Removal of noisy data.
3	Selection of Algorithm	Designing the model.	Identifying the best algorithm with the data available.
4	Training the datasets.	Training, testing and evaluating the model	Training the model to predict the water quality.
5	Prediction of water quality on hourly basis.	Predict water quality	Water quality is predicted for next hour based past 6 hours
6	Testing the results.	Accuracy calculation.	Calculating the accuracy of algorithm.
7	Alert system and data visualization	To display alert message and graph.	Web application showing the predicted values and appropriate message. Changes of the parameters are shown in the form of graphs.

The following Figure1.1 shows the stages followed in the project for the successful implementation.

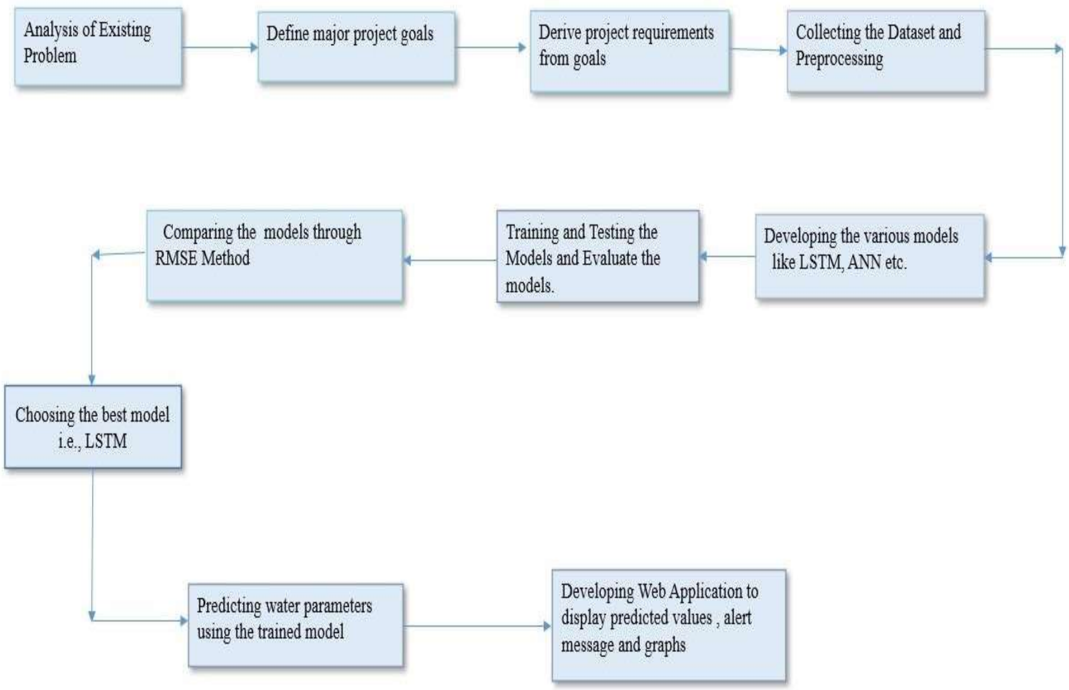

Figure 1.1 Methodology

1.6 Summary

In this chapter, we have discussed about what is aquaculture, its importance in the economy according to research. We have also discussed about the scope and the relevance of our project in the current trends. The main objectives of water quality testing in Aquaculture is discussed in this chapter.

CHAPTER 2

LITERATURE SURVEY

2.1 Related Work

Water quality prediction provides a significant reference for dynamic regulation of water quality and sudden events. After reading through the following research papers, we learnt about the importance of water quality in the aquaculture domain and its need to be monitored. The various parameters that will affect the water quality and their ideal ranges suitable for survival of fishes. The different methodologies used for water quality prediction.

In terms of water quality prediction, the main water quality prediction methods include time series method, Markov method, grey system theory method and support vector regression machine method. However, these methods have some problems, such as, the poor adaptive ability, low computational efficiency and inaccurate prediction results, which cannot meet the requirement in precision agriculture. In recent years, the methods based on artificial neural network have been proposed, which have the advantages of good robustness, high fault tolerance and sufficient fitting of complex nonlinear relations.

Juntao Liu et al.[1] focuses on the prediction of pH and water temperature parameters in key water quality parameters. Firstly, the water quality parameters are pre-processed by improved method. Then, the Pearson correlation coefficient method is used to find the correlation between the water quality parameters. Finally, the SRU (Simple Recurrent Unit) deep learning model is used to establish a prediction model for the key water quality parameters, so as to achieve accurate prediction. Meanwhile, we also evaluate the prediction effect of prediction model built by RNN (Recurrent Neural Network) deep learning network. This paper compares the working SRU and RNN and brings how SRU is better in terms of accuracy and how well SRU will fit the data compared to RNN. Figure 2.1 shows the flow of method followed for prediction.

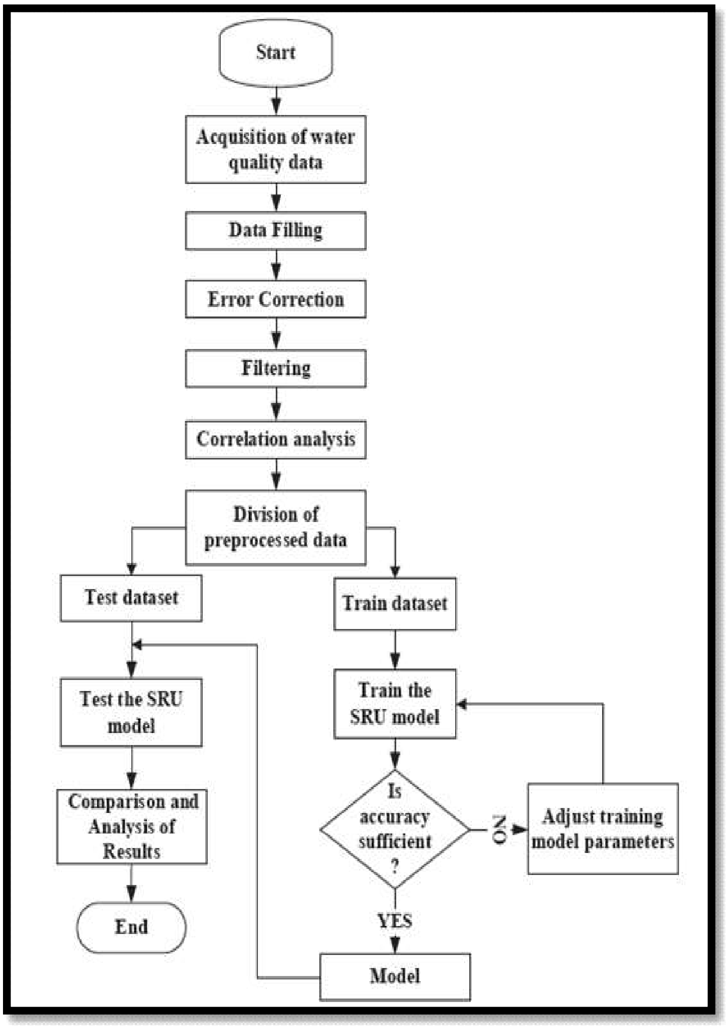

Figure 2.1 The Flowchart of SRU Model

Both models are close to the true value in prediction as seen in Figure 2.2, but it is obvious that the SRU performances a higher degree of fitting with the real value, so the SRU water quality parameter prediction model has better prediction result. The SRU adds a new unit states based on the structure of the RNN to control information in the network, and it is slightly redundant in the training time. In the pH prediction model, SRU consumed 16.929% more time than RNN. While in the water temperature prediction model, SRU takes 18.494% more time than RNN. It can be concluded that under the same conditions, SRU training is more time-expensive than RNN.

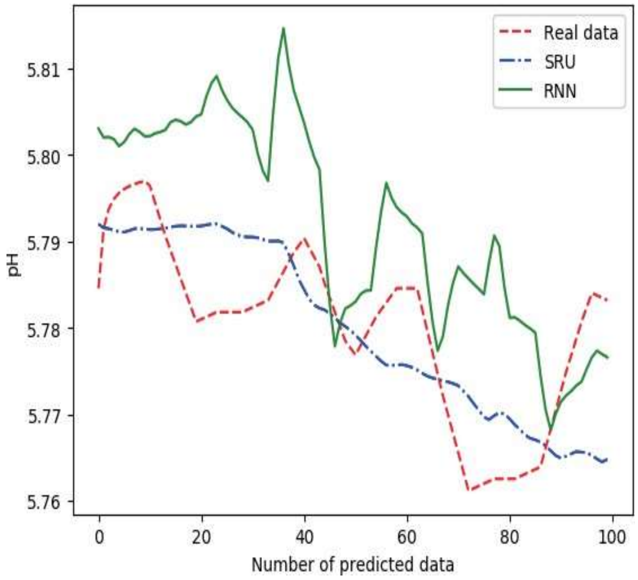

Figure 2.2 Comparation of predicted pH with real data

In modern intensive river aquaculture management, water quality prediction plays an important role. The water quality indicator series are nonlinear and non-stationer. Hence, the accuracy of the commonly used conventional methods, including regression analyses and neural networks, were limited. A prediction model based on Smooth Support Vector Machine (SSVM) is proposed by Wijayanti Nurul Khotimah[2] to predict the aquaculture water quality. SSVM is an algorithm that is used for solving no linear function estimation problems. The data used in this research are data of river in Surabaya collected for two years. The data have twenty variables that indicate water quality such as temperature, turbidity, color, SS, pH, alkalinity, free CO2, DO, Nitrite, Ammonia, Copper, phosphate, sulfide, iron, Hexavalent Chromium, Manganese, Zinc, Lead, COD, and Detergents. The Root Mean Square Error (RMSE) of the experiment is 0.0275. This value shows that SSVM proven to be an effective approach to predict aquaculture water quality.

Many researches to predict water quality has been done. Feifei Li et al. established short-term forecasting model to predict dissolved oxygen using backpropagation (BP) and autoregressive (AR) . Palani et al. developed a neural network model to forecast the amount of dissolved oxygen in seawater [6]. However, neural networks suffer from a few weaknesses, such as over-fitting. In this research, we propose SSVM to predict water quality (especially dissolve oxygen).

Nowadays, Support Vector Machine (SVM) becomes a method that is constantly evolving and increasingly popular in machine learning. This method has several advantages over other methods such as neural network. Those advantages are [10].

a. Assuring global optimum

b. Parameters that must be estimated relatively few

c. The model is stable

d. Relatively simple to use

e. Successfully applied in most real cases.

Many SVM developments proposed to improve its performance and efficiency. Lee and Mangasarian [11] have proposed a new formula of SVM with linear and non-linear kernel for classification using smoothing method. The method is called Smooth Support Vector Machine (SSVM). The basic concept of SSVM is changing primal SVM formula into non smooth optimization problem without constraint. Hence, the objective function of the optimization problem is not differentiable. Therefore, smoothing function was used in purpose to achieve differentiable objective function. This method was solved by Newton Armijo algorithm. By using the Newton-Armijo algorithm can be seen that the main difference between the smoothing approaches with SVM is that SSVM solve problem using linear equations, while SVM using quadratic programming problem.

The result of dissolved oxygen prediction using SSVM is shown in Figure 2.3. The figure shows the comparison between prediction value of this proposed method and actual value (original data). We used 5-fold for the experiment. The RMSE value of the prediction is 0.0275. This value shows that SSVM proven to be an effective approach to predict dissolved oxygen. The SSVM forecasting method to predict water quality can help avoid economic losses causes by water quality problems.

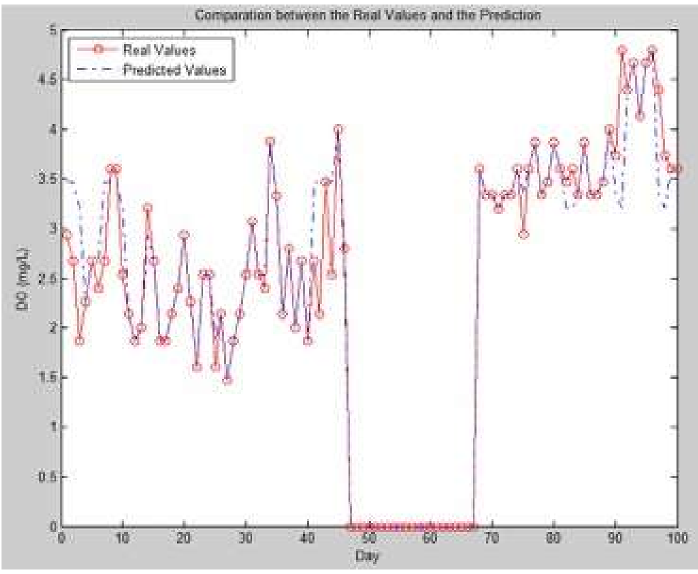

Figure 2.3 The Comparison between Predicted Value and Actual Value of DO

J. Wang et al.[3] investigates the characteristics of dynamic nonlinearity and correlation of water quality parameter information, as well as the gradient disappearance and gradient explosion caused by the training data of traditional RNN network model, etc. The long short-term memory network structure (LSTM) is introduced to optimize the structure of RNN network and the connection weight and threshold of hidden layer. A new water quality parameter prediction model of LSTM-RNN network based on improved RNN network structure is proposed by setting the number of storage units in the hidden layer of the network, the number of structural layers of the network model, and adjusting the time window size of the data training set. Combined with the water quality monitoring data of the River in Shanghai, the model is used to predict and verify the main pollutant index COD (potassium permanganate index) in the River. The simulation results show that compared with the traditional GM (grey model) and RNN network water quality prediction model, the sample approximation accuracy and generalization ability of the training prediction based on LSTM-RNN network model is higher and better than that of the traditional GM (grey model) and RNN network model. Good comprehensive prediction performance of river water quality is presented.

The commonly used methods are time series prediction, artificial neural network prediction method, regression analysis prediction method, grey system prediction method and so on. The time series prediction method belongs to the statistical prediction model with no cause variables. Its characteristic is that the mathematical theory foundation is perfect, the practice is more difficult, and the error is large. The prediction accuracy of the prediction model established by artificial neural network forecasting method can reach a satisfactory degree, but it is not clear from the prediction model that the specific reasons and internal relations of the water quality change trend can be clearly understood from the prediction model. Therefore, the prediction accuracy of the model is affected. Regression analysis and prediction methodthrough a large number of data samples for correlation analysis to obtain the correlation, and establish a regression equation, on the basis of considering the prediction error to determine the future water quality prediction value. Its model is complex, and the distribution of data and samples requires higher requirements. The grey system forecasting method can get regular time series by accumulating the raw data without regular rules once or more, and then establish a forecasting model to predict the water quality in the medium and long term. It is easy to be affected by data instability, and then will produce a large prediction error. Because water quality parameter is a dynamic time series, it is more suitable to use recurrent neural network (RNN). In addition, the prediction process of water quality parameters is gradual, that is to say, the current water quality parameters are correlated with the historical water quality parameters. Traditional water quality detection schemes are inefficient, which will greatly delay the timely management of fishery water. UAV has expanded our thinking of using it in the fishery field. It focus on the principle of pH, dissolved oxygen and ammonia nitrogen sensors and their application in fishery production practice. Figure 2.4 shows the framework of the fishery water quality assessment and prediction system.

Figure 2.4 Framework diagram of fishery water quality prediction system

Dong Yao et.al.,[17] discussed the Remote control has been used to park the drone on the planned route, use the computer terminal ground station to monitor the UAV's flight status, and the GPS satellite to locate the real-time area. After the drone is parked on the surface, three groups of sensors hidden in the EVA material began to monitor the water quality. During the floating sampling period, the data is transmitted to the database of the cloud server through GPRS module, and the data is processed at the back end. The evaluation and prediction results are finally displayed on the application side and PC side. In this system, each group of sensors uploads the average sampling value to the cloud server, and data are uploaded and updated once every 3 seconds interval. In the 40-minute evaluation operation, 10 minutes are used for sensor pre-processing, the remaining 30 minutes for data acquisition, each group of sensors will update 600 times of experimental data.

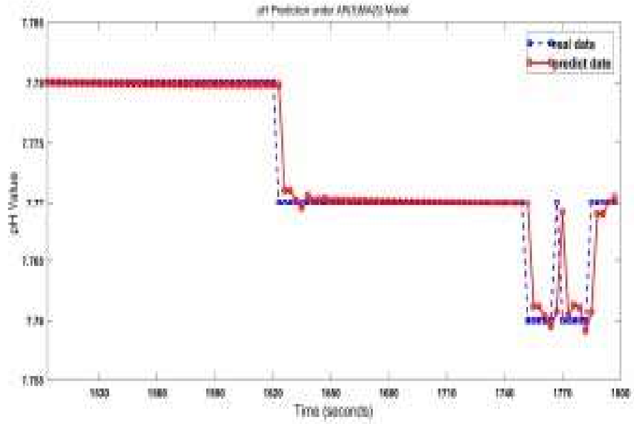

Figure 2.5 The pH Prediction under ARIMA Model

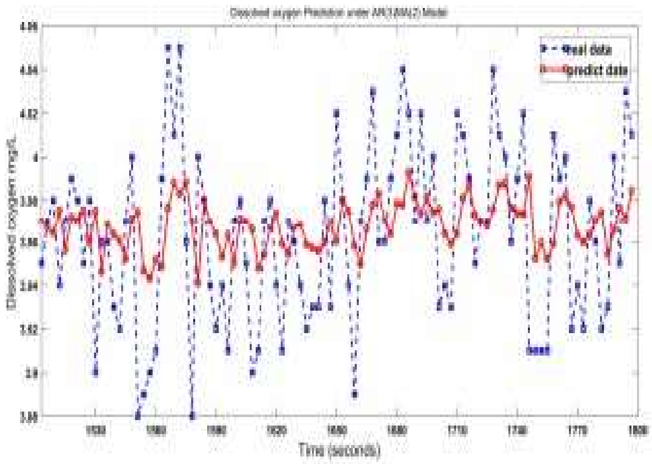

Figure 2.6 Dissolved oxygen (DO) Prediction under ARIMA Model

CHAPTER 3

REQUIREMENTS SPECIFICATION

3.1 Hardware Requirements

- System: Core i5 Processor
- Hard Disk :1 TB.
- Monitor: 15” LED
- RAM: 8GB

3.2 Software Requirements

- Operating system: Windows /UBUNTU.
- Programming Language: Python 3, HTML, CSS, JavaScript, JSON
- Software: Anaconda -Jupyter Notebook
- Web browser -Google chrome, Firefox, IE8+.

3.3 Summary

This chapter gives an insight into the functional an non-functional requirements that the system provides. It also describes the hardware and software requirements that are required for building the system.

CHAPTER 4

SYSTEM ANALYSIS AND DESIGN

4.1 System Architecture

The proposed system architecture (Figure 4.1) shows the complete working of the system starting from training the model using the collected dataset to showing the predicted result and appropriate message on the web application.

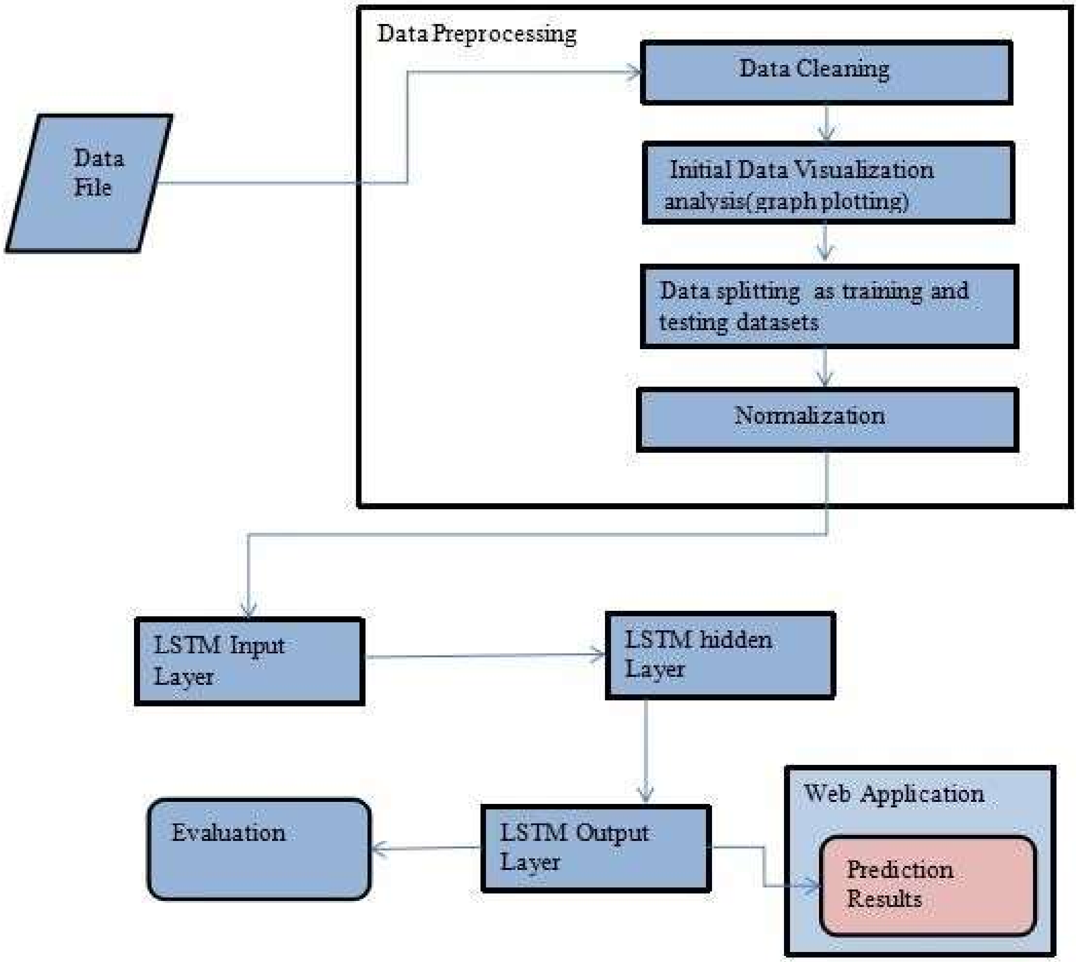

Figure 4.1 Proposed System Architecture

4.2 Flow Chart

The below flow chart (Figure 4.2) shows the step by step execution implemented at the backend and frontend of the system.

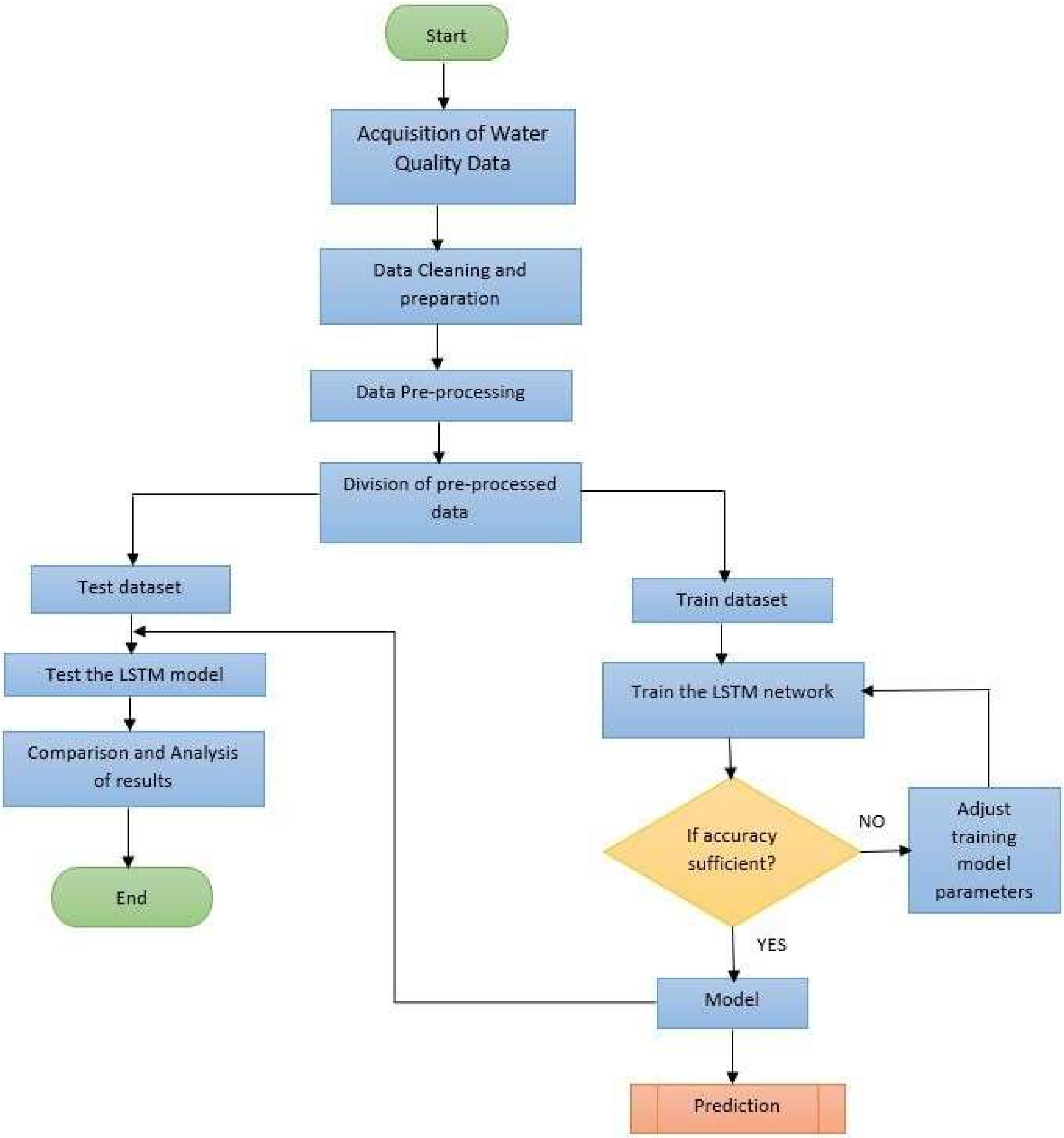

Figure 4.2 Flowchart of the Water Quality Prediction Model

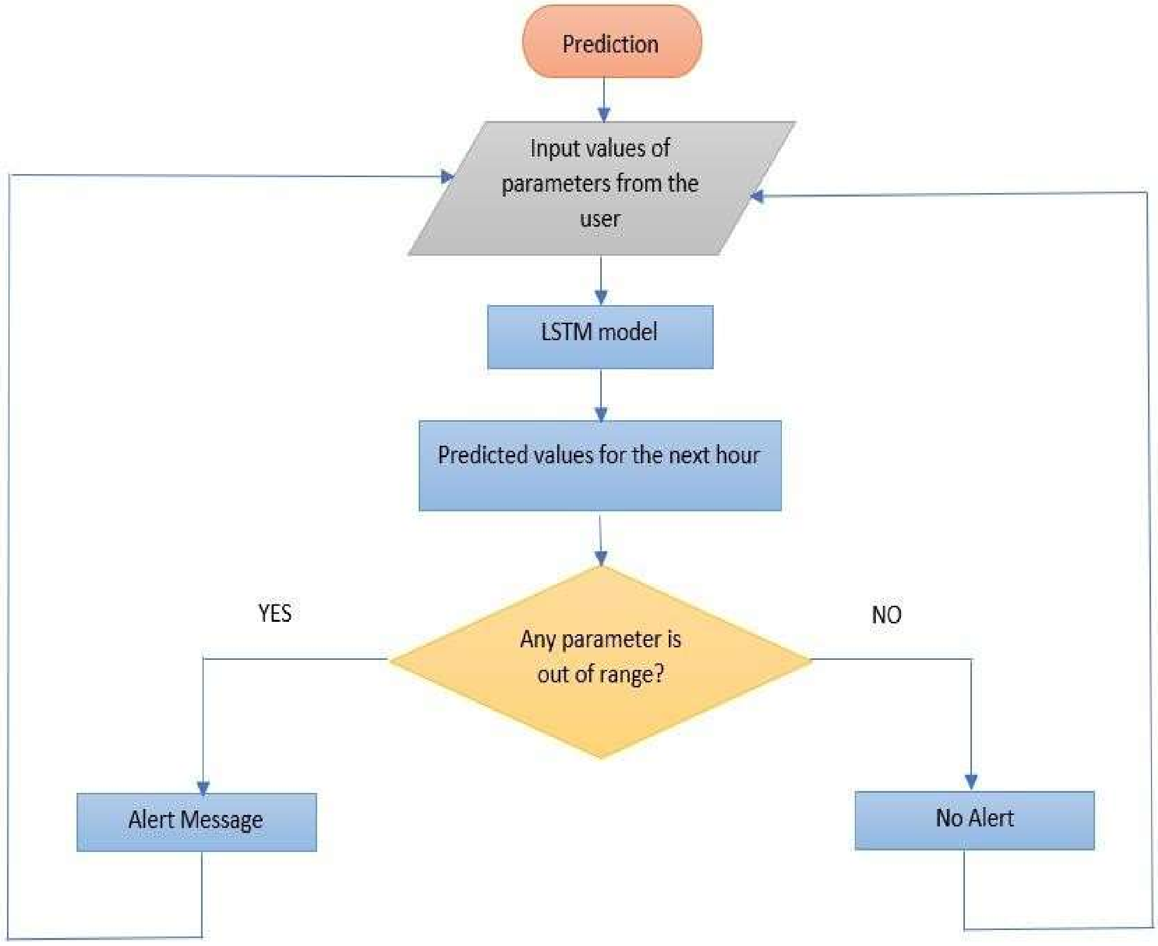

Figure 4.3 Flowchart of Prediction Subroutine

State Diagram

Figure 4.3 represents the transition between various states of the prediction system. It gives an idea about the various states and the events involved from data collection to generating an alert when any parameter is not in range.

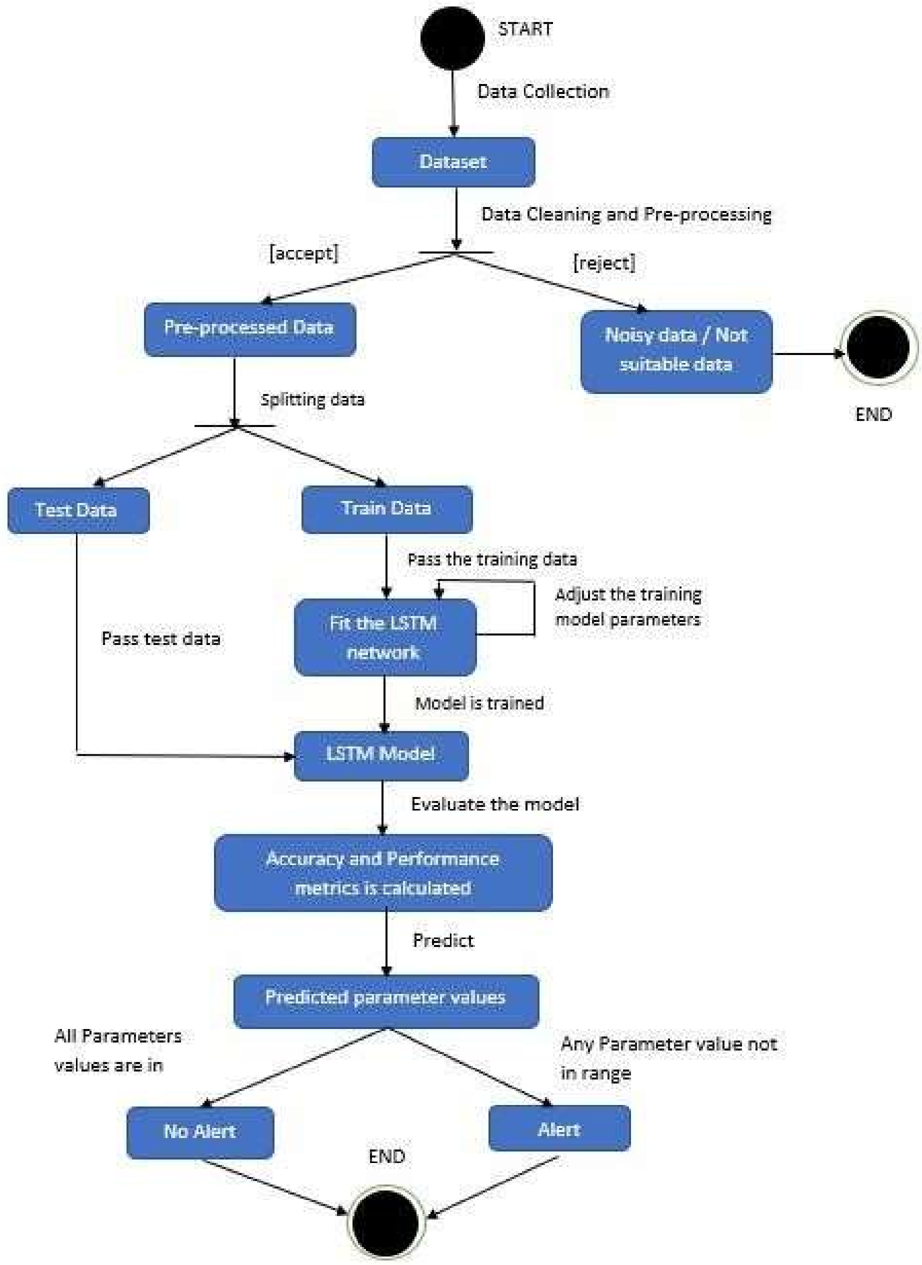

Figure 4.4 State Diagram of Water Quality Prediction System

Use Case Diagram

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved.Here in Figure 4.3 Supervisor is involved in the use cases View prediction result, View Visualization, View Alert and Managing the application. Whereas Aquafarmer is involved only in viewing the prediction result, visualization and alert.

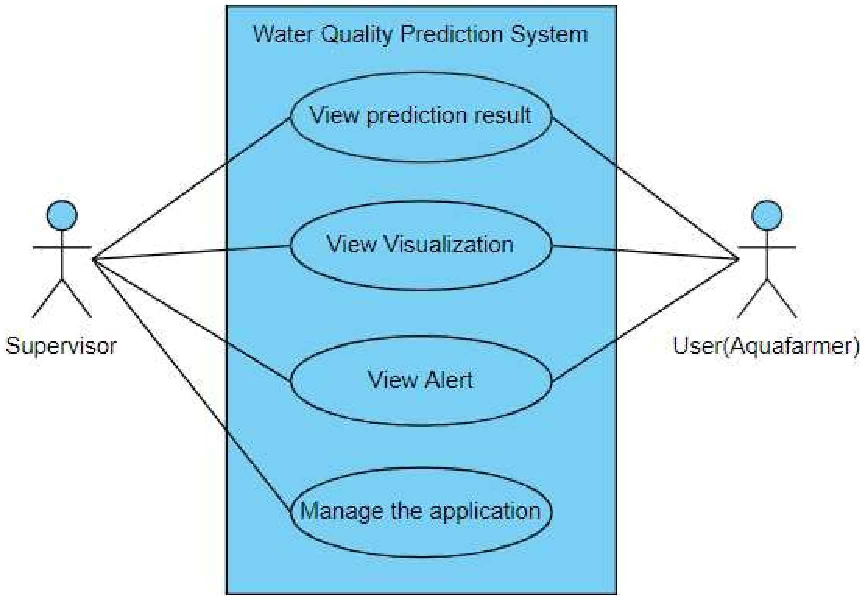

Figure 4.5 Use Case Diagram of Water Quality Prediction System

Sequence Diagram

Sequence Diagram (Figure 4.4) is interaction diagram that detail how operations are carried out.it shows how they interact over time and they are organized according to object (horizontally) and time(vertically). From the data set data is extracted and processed, respectively. Then ML model is trained, and input data is given by the user and future prediction is made. User can view the prediction results and visualization through the Web application.

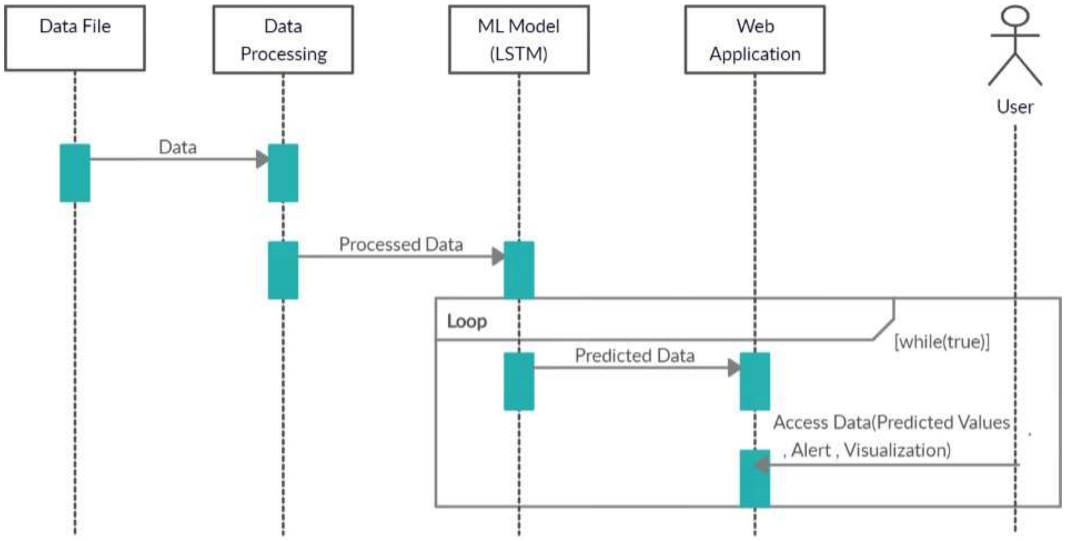

Figure 4.6 Sequence Diagram of Water Quality Prediction System