Advanced Control System and Water Quality Prediction using ML for Aqua Farms

Anagha Patil, Shreyas Manoti, Dhruvil Mehta, Tushar Mittal

Anagha Patil is with Vidyavardhini's College of Engineering and Technology, Vasai, Maharashtra, India (e-mail: anagha.patil@vcet.edu.in).
Shreyas Manoti was with Vidyavardhini's College of Engineering and Technology, Vasai, Maharashtra, India (e-mail: manotishreyas@gmail.com).
Dhruvil Mehta was with Vidyavardhini's College of Engineering and Technology, Vasai, Maharashtra, India (e-mail: dhruvilmehta2606@gmail.com).
Tushar Mittal was with Vidyavardhini's College of Engineering and Technology, Vasai, Maharashtra, India (e-mail: tusharmittal1201@gmail.com).

ABSTRACT

India is an agricultural Country. It contributes to a significant portion of our GDP. Aquaculture in India contributes 5.15% of its Agricultural GDP. Aquaculture is one of the important sectors for India for livelihood and due to its export demands. Aquaculture is one of the most capital-intensive fields today, so it is crucial to have a good Return on Investment (RoI) for the farmers. In order to yield a healthy culture, various parameters have to be closely monitored and controlled. These parameters include Temperature, Salinity, Co2, PH Levels, Dissolved Oxygen, etc. In its current state Aqua Farms does not collect much data and thus cannot use this data for increasing the yield and its quality. This paper proposes a complete system to monitor, collect and use the data from the Aqua Farms to efficiently increase the productivity of the farms. We have also presented an implementation of a Wi-Fi-based Control System using IoT compatible with the above-mentioned monitoring system to control the pond aerators remotely using a Mobile Application. We have also proposed further automation of the entire pond aerators by integrating the Dissolved Oxygen Sensor readings with our system.

Keywords:

Agriculture, Motor Starter, Three Phase Motor, IoT, NodeMCU, Relay, Sensors, Data Science, Machine Learning, ML models

1. INTRODUCTION

India is an agricultural Country that provides the primary livelihood for most of its population. It also accounts for a significant portion of our GDP. Thus, Indian economy relies heavily on its agricultural yield. Aqua-Culture in India contributes about 5.15% of the Agricultural GDP. Thus, Aquaculture is one of the most important sectors for India for its livelihood as well as due to its export demands. In aquaculture, shrimp farming is continuously gaining a significant position because of its higher export potential. At the same time, its productivity is comparatively lower as compared to other countries.

In shrimp farming like any other aqua farming, the environment in the pond plays a crucial role in the growth of the yield. All the parameters have to be meticulously monitored in order to ensure a good yield. Parameters such as Temperature, Salinity, PH Levels, Co2, Dissolved Oxygen, etc. play a vital role in determining the quality of the yield. Thus, it is clear that in order to increase the yield and its quality, the monitoring and controlling of all these factors are extremely important. As of now, aqua farms in India does not collect data using various sensor or more importantly they do not have a system and methodology in place to properly collect all these data and further use this data to increase the quality of their yield. In its current state, all the readings are manually recorded by the farm workers. Temperature, Salinity, etc. Whereas PH levels and Dissolved Oxygen levels have to be tested manually every few hours. This in itself gives rise to numerous manual errors making it an inefficient system.

We have proposed an Optimized Control System to monitor and control all the operations in the pond. We have also implemented an IoT-based Cost-effective product using ESP32 Wi-Fi Nodes to control the pond aerators remotely and further automate it entirely based on the readings from the dissolved oxygen sensor. We believe that this would pave the way to a more Data oriented Farming System that would include monitoring all the parameters using different Sensors and integrating this data to control the ponds.

This data can further be also used for Data Analytics to reveal more information and can be used by authorities to study the farmlands and the changing variables in them.

2. Literature Survey

The ever-growing internet has opened up new horizons for development in various fields. The most basic component for automation is derived from simply controlling any device with the use of a relay and a microcontroller unit.

Prathmesh Shelke, et al have proposed a home automation system using NodeMCU [1]. The idea was to basically convert the analog switches into a wireless smart switch that involves operating it using a Mobile App. The entire process is then connected with the cloud to give access from anywhere in the world using the Internet as its backbone. With the use of ESP8266, they have managed to control the relay and thus automate the analog switches and a software interface for the users to interact with the hardware.

In the paper [2] Rajat Patinge et al have proposed and demonstrated the use of IoT in controlling Three Phase Motors that have industry-wide applications [2]. The primary goal was to remotely control or convert the traditional induction motor to be operated smartly with the use of IoT, MCU, and software. They have also inculcated the automation of the motor control systems in order to prevent system failures. In demonstrating this this paper presents the use of Arduino, NodeMCU ESP8266, Relay Module, and Power Board.

Sujit Thakare et al. [3] have proposed a smart irrigation system using Arduino, sensors, and ESP8266 WiFi Module. Herein, the authors have presented the use case and importance of integrating sensors with the smart control system to provide a framework for Monitoring and Controlling the Farm. The idea was to continuously monitor various parameters such as PH level, the moisture content in the soil, etc., and direct the NodeMCU to take appropriate actions dynamically. The authors have also implemented cloud features into this which means they can upload all the recorded parameters into the cloud for further data analysis as well as for record-keeping. Arduino collects the data from the sensors and then links those to the cloud. This brings in an important functionality of collecting live data from the farm.

Adithya Vadapalli et al. in their research [4] define the parameter required for having the remote control and a smart agriculture environment that would even monitor the weather, dampness in the soil, etc. with an aim to maximize data collection from the environment and using it to improve the farm. The authors have proposed a Wireless Sensor Network (WSN) for live monitoring and collection of data and making decisions based on the live data. They have used various sensors for improving irrigation like soil moisture sensors, temperature sensors, etc. The threshold for these sensors would be set according to the type of crop decided by the farmer. The potential also includes monitoring regular changes in environmental conditions such as climate, hydrology, temperature, rain, dampness, etc, physiology of plants, etc. This provides an outline of and importance of having a monitoring system integrated with the smart control system of any farm.

In the paper [5], authors propose the use of Solar Technology to support the wireless remote-control system and use of Cloud Technologies to collect and analyze real-time data from the farmlands. The authors have proposed the creation of a device they have named an IoT Stick, which would work as a Plug & Sense model. Using this device, the farmers can plug the device into any farm to get Live Data Feeds on various devices like smartphones, computers, etc. It also provides a functionality to share this sensor data with Agricultural experts or Agricultural Consultants anywhere remotely through the cloud. It also has provision for the analysis of sorts of different data using Big Data Analytics.

3. Proposed Methodology

The Proposed Methodology is being broken into three modules for ease of understanding and implementation.

3.1 IoT based Remote Control System

We Propose to create and implement an IoT-based Control System to reduce and improve the pond operations that are currently performed manually. Creating a wireless control system would enhance the operations of the farm and would make it easier to operate a farm with reduced manual labor. This system would be implemented for the operation of Pond Aerators.

This System consists of creating a Mesh Network using ESP32 Micro Controller Unit that would in turn control the Pond Aerators. ESP32 is a feature-rich MCU that has integrated Wi-Fi and Bluetooth connectivity that can be used for a wide range of applications. ESP32 is the successor to ESP8266 launched in 2016 by EspressIf Systems. ESP32 has a dual-core processor with more GPIOs than its predecessor. ESP32 has a higher clock frequency and is Bluetooth enabled, with higher security than ESP8266. It provides a self-forming and self-healing network capability. The network topology of ESP-WIFI-MESH can scale up to 1000 nodes in widespread large areas independent of any specific Wi-Fi infrastructure support. Another distinguishing characteristic of ESP32 lies in its cost. It is a low-cost MCU when compared to other Microcontrollers using technologies such as LoRA. It is a low-cost, low-power device that has advanced power management mechanisms. ESP32 offers five configurable power modes. Thus, using these various sleep modes, the ESP32 has one of the most advanced power management techniques.

Thus, the proposed solution can be implemented by using a Mesh Network of ESP32 MCU which would Wirelessly Control the Relay connected to the Aerator Motor Starter.

3.2 Data Collection based on Sensors:

After the Initial Control System, the next big challenge faced by most farms is that of the data or lack thereof. Most farms today do not use the freely available data efficiently. We propose to implement different Sensors in the farm to measure different parameters crucial to the growth of the yield. Also, collection and analysis of these data would result in increased awareness about the happenings on the farm as well as would lead the farmers to a completely data-oriented decision-making process.

The parameters that affect the growth of the yield are:

1) **Dissolved Oxygen (DO):** Dissolved oxygen (DO) is an important factor that affects the life of aquatic organisms in aquaculture ponds. Aquatic organisms require oxygen for respiration, and the level of dissolved oxygen in the water can have a significant impact on their survival, growth, and reproduction.
2) **pH Level:** pH can have a significant impact on the growth and survival of aquatic beings in aquaculture ponds. Aquatic organisms have specific pH ranges within which they can thrive, and any deviation from their optimal pH range can cause stress, disease, and even death.
3) **Salinity:** Salinity levels in aquaculture ponds affect the growth and survival of aquatic beings. Most freshwater species cannot tolerate high salinity levels, while some marine species require a certain level of salinity. Deviation from the optimal salinity range can cause stress, disease, and death. Regular monitoring and adjustment of salinity levels are important to maintain optimal conditions for the specific species being raised, which can be done through freshwater or saltwater addition or water treatment techniques.
4) **TDS:** TDS (Total Dissolved Solids) is another important factor that can affect the growth and survival of aquatic beings in aquaculture ponds. TDS refers to the concentration of inorganic and organic materials, such as minerals and salts, that are dissolved in the water.
5) **Temperature:** Water temperature in aquaculture ponds affects the growth and survival of aquatic beings. Different species have specific temperature ranges within which they can thrive. Deviation from the optimal temperature range can cause stress, disease, and death. Regular monitoring and adjustments to maintain optimal temperature ranges are important, which can be done through temperature control devices or water management techniques.

These data can be collected with the help of the sensors which are available in the market. We can proceed to implement a completely automated control system process for the Pond Aerators with the help of these data. Generally, in aqua farms, Pond Aerators are used to increase the amount of Dissolved Oxygen in the water. Through the use of data from the DO Sensor we propose to implement a completely closed-loop automated Control System for pond aerators that would switch ON/OFF the motor starter based on the DO Sensor readings. This can help reduce the labor work and increase the ease of operating the farm to a large extent.

3.3 Data Science and Machine Learning:

Through the use of Data Science and Machine Learning, we propose to predict the health of the flora and fauna in the pond. With the use of installed sensors and different data, it would record and with the use of a machine learning model, we can predict the health and water quality in the pond.

With more data through sensors, we can predict the water quality. Additionally, we can also predict the number of days in which the yield would be ready based on all the parameters such as consistent DO levels, PH levels, temperature, salinity, etc.

4. DEPLOYMENT AND IMPLEMENTATION

We have divided the implementation and working of our project into three sections for ease of understanding and implementation.

4.1 IoT-based Remote Control System

The proposed methodology consists of using a number of ESP32 Development Boards. These ESP32 Nodes would form a Network that would enable long-distance communication within themselves.

Of these, One ESP32 Node would be connected to a Cloud Service via the Internet. This Board would act like a Server Node. Connected directly to this Server Node would be a Gateway Node through which the remaining ESP32 Nodes would be connected in a Painless Mesh Network. The control signals can thus be sent using the Internet to the Server Node which via the Gateway Node would broadcast the signal so that the intended Node can get the signal.

There would be one ESP32 Node connected to each and every motor starter via an external relay which would enable the automation of the aerators.

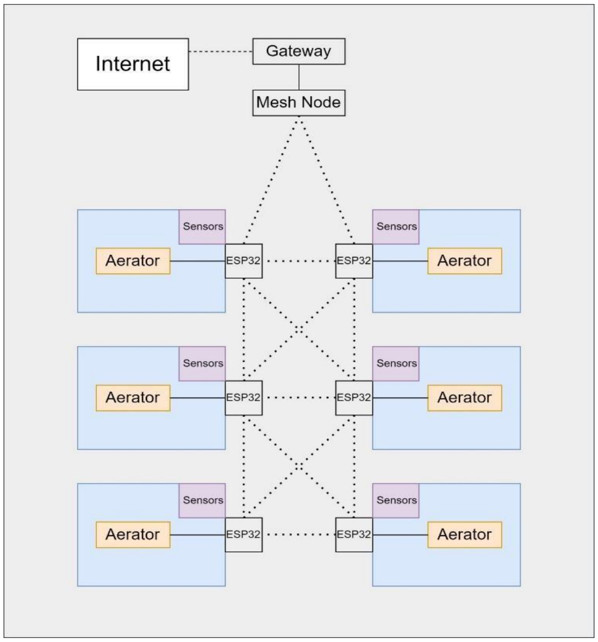

Fig 1:Pond Diagram

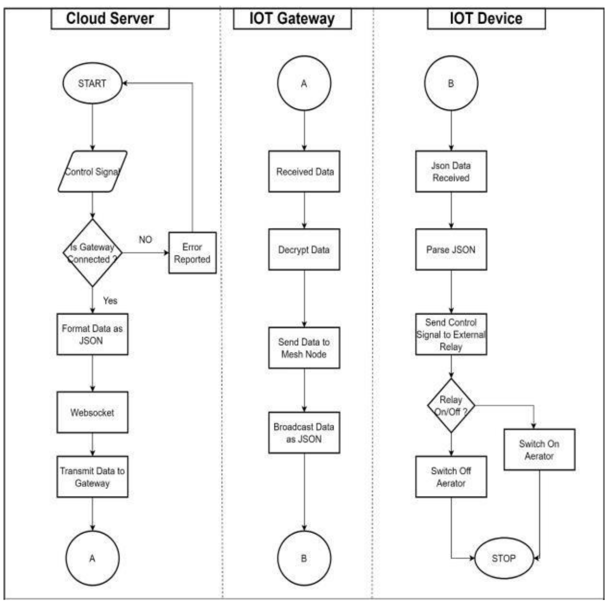

Fig 2: Flow Chart

painlessMesh is a library that provides us with true ad-hoc capabilities meaning that it doesn't require any kind of central controller or router to establish a network. Any system that consists of 1 or more nodes will self-organize into a fully functional mesh network. The maximum capacity of the mesh is limited by the amount of memory in the heap that is free to be allocated to the sub-connections buffer which makes its capacity quite large.

painlessMesh uses broadcasting services for communications between the connected nodes in the network. There are a couple of reasons for this. Firstly, it makes the code and the messages human-readable and easy to understand. It makes it literally painless to integrate painlessMesh with JavaScript front-ends, web applications, and other applications. painlessMesh was conceived and designed to be used with many MCUs like the Arduino, NodeMCU, etc but it does not use the Arduino Wi-Fi libraries explicitly. Instead, the networking is completely done by using the native ESP32 and ESP8266 SDK libraries, which are freely available. In almost all cases the networking libraries used don't matter to most users much, as one can just include painlessMesh.h, run the init() method, and then use the library via an API.

4.2 Data Collection based on Sensors:

We have gone through different types of sensors and implemented various sensors for recording various different parameters necessary for the optimal functioning of a farm. The sensors include DO, Temperature, Salinity, PH levels, etc. These sensors are installed in the farms so that every small change is accounted for and recorded so as to help us with a reliable data analysis process. We have connected each sensor with an individual ESP32 Node which is used to transmit the sensor data to the server. It Collects and transmits data every 10 minutes.

These data can be stored and collected over a long period of time to be used in Big Data Analytics and derive meaningful context about the surrounding and the farm from the data.

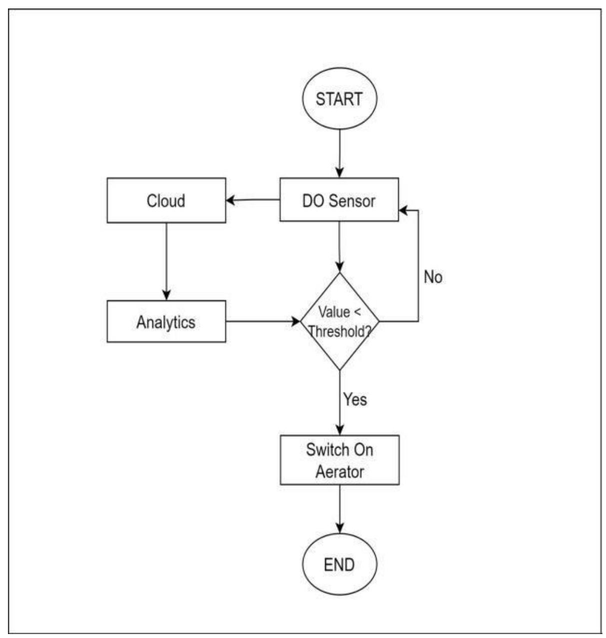

Fig 3: Data Collection and Analysis

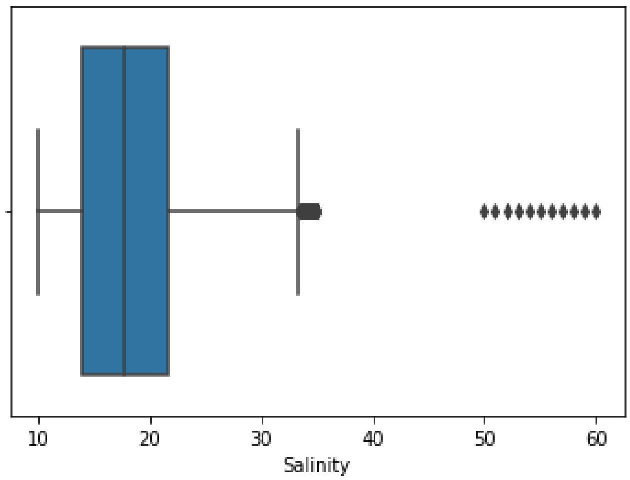

Fig 4: Salinity Outlier

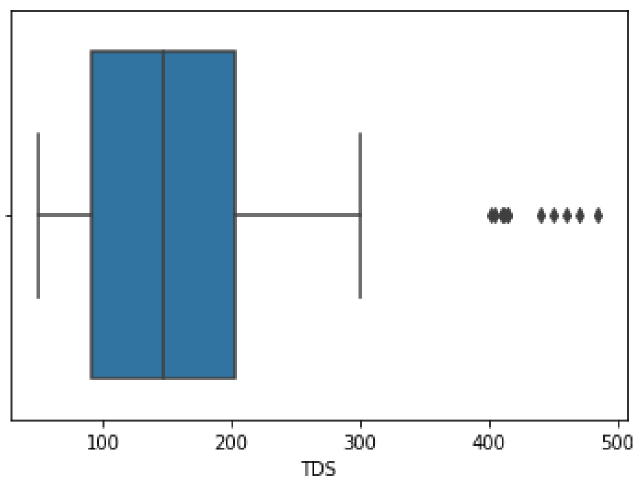

Fig 5: TDS Outlier

4.3 Water Health Prediction Using Machine Learning:

After collecting the data, now comes the part of Water Quality Prediction. The collected data were given the label of “1” (if water quality is good) or “0” (if water quality is bad). Our goal is to predict the label of Health which can be either “1” or “0”.

We have used various classification algorithms to compare the accuracies and AUC-ROC curve.

Algorithms:

1) **Logistic Regression:** Logistic regression is a widely used and interpretable method for classification problems. It is a popular method for predicting the probability of a binary or multi-class outcome based on one or more predictor variables. The logistic regression model assumes that the relationship between the predictors and the outcome is linear on the logic scale, which is the natural logarithm of the odds of the outcome.

2) **KNeighbour Classification:** The k-nearest neighbors (k-NN) algorithm is a simple but powerful classification algorithm that can be used to predict the class of a new data point based on the class of its nearest neighbors in the feature space. The k-NN algorithm is based on the idea that points in the same class tend to be close to each other in the feature space.

3) **GuassianNB:** Gaussian Naive Bayes (GNB) is a classification algorithm that assumes that the features in each class are normally distributed and conditionally independent given the class label. The algorithm calculates the probability of each class label given the values of the features using Bayes' theorem and the normal distribution assumption, and determines the class label of the new data point as the one with the highest probability.

4) **Random-Forest Classifier:** Random Forest classification is a machine learning algorithm that builds an ensemble of decision trees to classify new data points. Each decision tree is trained on a random subset of the features and a random subset of the training data. The algorithm combines the predictions of the decision trees to determine the class label of a new data point. Random forest classification is robust to overfitting, can handle high-dimensional feature spaces, and is suitable for both binary and multi-class classification problems.

5) **Decision Tree Classifier:** Decision tree classifier is a machine learning algorithm that builds a tree-like model of decisions and their possible consequences to classify new data points. The algorithm works by recursively splitting the training data based on the values of the features that result in the largest information gain or the smallest impurity. The resulting tree is used to classify new data points by following the branches of the tree based on the values of their features until a leaf node is reached, which corresponds to a class label.

6) **AdaBoost Classifier:** AdaBoost (Adaptive Boosting) classifier is a machine learning algorithm that builds an ensemble of weak classifiers to classify new data points. The algorithm works by iteratively reweighting the training data and training a weak classifier on each iteration, such that the subsequent classifier focuses on the misclassified data points of the previous classifiers. The final classification decision is made by combining the predictions of all the weak classifiers, weighted by their classification accuracy.

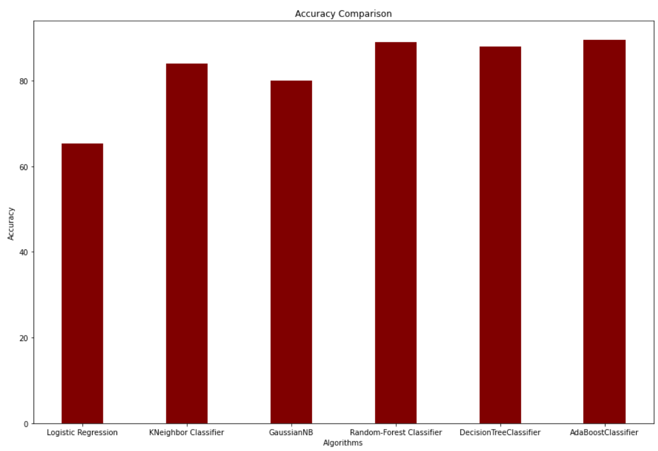

Fig 6: Accuracy Comparison of Algorithms

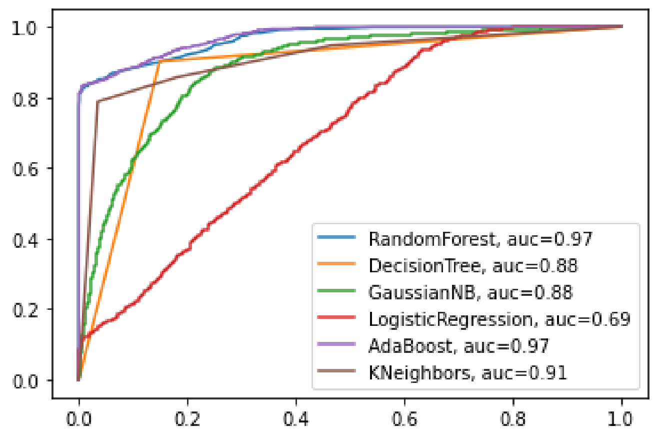

Fig 7: AUC-ROC Curve

5. RESULTS AND ANALYSIS

In practical testing, the range between two ESP32 Nodes went up to 30 meters. Multi-Hoping was successfully achieved using four nodes thereby demonstrating the ESP32 painlessMesh library. A delay of approximately 5 seconds was recorded between the action (Clicking of the button), and the reaction(Performing the operation).

Sensors sometimes misfunction and receive outliers, garbage, or Null values. Factors responsible for these redundant values may include the network, environment, or power issues. We have performed Data Cleaning on the dataset to further process and remove the redundancy from the dataset. We have found Outliers in the data by plotting Boxplot. These outliers were cleared from the dataset. Additionally, sometimes Sensors misfunction, and due to that, we receive some blank data in the dataset. These blank or Null data were replaced by either mean, median, or mode as required.

Upon applying various ML model, it is possible to ascertain certain water health quality suitable for the optimal growth of the culture in the pond.

From Fig: 6 and Fig: 7, AdaBoost Classifier has given the highest Accuracy Score and AUC and therefore we'll be using AdaBoost Classification for further operations in the project.

ACKNOWLEDGMENTS

This work was supported by Aqvashrimp Technology Solutions Private Limited. The authors gratefully acknowledge Aqvashrimp Technology Solutions technical and financial support throughout the entire work duration.

REFERENCES

.

1. P. Shelke, S. Kulkarni, S. Yelpale, O. Pawar, R. Singh and K. Deshpande, "A NodeMCU Based Home Automation System", IRJET, Vol. 05, no. 06, Jun. 2018.
2. R. Patinge , A. Waghade, P. Agarkar, P. Mohite and D. Bhoyar, "Three phase motor controlling using concept of IoT", IJARIIE, Vol. 05, no. 02, 2019.
3. S. Thakare and P.H Bhagat, "Arduino-Based Smart Irrigation Using Sensors and ESP8266 WiFi Module", ICICCS, 2018.
4. A. Vadapalli, S. Peravali and V. Dadi, "Smart Agriculture System using IoT Technology", IJARSE, Vol. 09, no. 09, Sept. 2020
5. A. Nayyar and E. V. Puri, "Smart Farming: IoT Based Smart Sensors Agriculture Stick for Live Temperature and Moisture Monitoring using Arduino, Cloud Computing & Solar Technology", ICCCS, Nov. 2016
6. S. G. Devi, "Random Forest Advice for Water Quality Prediction in the Regions of Kadapa District", IJITEE, Vol. 08, no. 6S4, April 2019.
7. T.U. Chengsheng, LIU Huacheng, and XU Bing, "Adaboost typical Algorithm and its application research", ICMITE, 2017.
8. A. Omambia, B. Maake, A. Wambua "Water Quality Monitoring Using IoT & Machine Learning", IST-Africa, 2022.
9. S. A. Vergina, S.Kayalvizhi, R.M. Bhavadharini, K.S. Devi, "A Real Time Water Quality Monitoring Using Machine Learning Algorithm", EJMCM, Vol. 07, no. 08, 2020.
10. M. Sun , X. Yang, Y. Xie, "Deep Learning in Aquaculture: A Review", Journal of Computers, Vol. 31, no. 01, 2020.